

Metody Matematyczne w Uczeniu Maszynowym

Soma Dutta

Uniwersytet Warmińsko-Mazurski w Olsztynie

Semestr Zimowy, 2025

Trzy podstawowe koncepcje uczenia maszynowego

1. **Dane** (Data): Uczenie maszynowe jest z natury oparte na danych.
2. **Model**: Model zapewnia formalny sposób opisywania danych, ich wzorców, ich relacji z powiązanymi koncepcjami itp.
3. **Uczenie się** (Learning): Uczenie się to proces ulepszania wydajności klasyfikowania niewidzianych danych na podstawie wzorców widzianych danych poprzez optymalizację parametrów modelu

Przykład : Z dużego korpusu dokumentów automatyczne wyszukiwanie odpowiednich tematów udostępnianych w dokumentach

Cztery filary uczenia maszynowego

1. Pierwszym krokiem do uczenia maszynowego jest reprezentacja i agregacja danych: **Algebra liniowa**
2. Drugim krokiem jest zaprojektowanie modelu, który może opisać naturę i zachowanie danych, które często są niepewne, oraz ocenić wydajność modelu w niewidzianych przypadkach testowych: **Prawdopodobieństwo i statystyka**
3. Trzecim krokiem jest nauczenie się, jak można poprawić wydajność, minimalizując błąd klasyfikacji danych: **Rachunek różniczkowy**
4. Ponadto niektóre **techniki optymalizacyjne** i wiedza z zakresu **matematyki dyskretnej** pomagają dostroić niektóre parametry wykorzystywane w algorytmach uczenia się i zwiększyć wydajność algorytmów.

Ogólny Plan

- ▶ Poznamy każdy z istotnych tematów matematycznych
- ▶ Dowiemy się, gdzie i jak taka wiedza matematyczna jest wykorzystywana w uczeniu maszynowym
- ▶ Podczas sesji ćwiczeniowych będą wykonywane ćwiczenia związane z rozwiązywaniem problemów (na papierze) z zakresu matematyki oraz ćwiczenia odnoszące się do praktyki (pisanie programów).

Algebra Liniowa

- ▶ Dane liczbowe są reprezentowane przez wektor, a tabela takich danych jest reprezentowana przez macierz.
- ▶ Każdy wektor przedstawia opis obiektu lub sytuacji w świecie rzeczywistym. Oczekuje się, że wektory o podobnej charakterystyce dadzą podobne wyniki, gdy stosuje się algorytmy uczenia maszynowego.
- ▶ Potrzebujemy więc operacji przedstawiających podobieństwo wektorów.
- ▶ Wektory są szczególnymi przypadkami macierzy.

Treści: Algebra Liniowa

1. Macierze i ich operacje
2. Układ równań liniowych
3. Przestrzeń wektorowa
4. Transformacja liniowa (z jednej przestrzeni wektorowej do innej przestrzeni wektorowej)

Macierze

Biorąc pod uwagę $m, n \in \mathbb{N}$, macierz o wartościach rzeczywistych A formatu $m \times n$ jest $m \cdot n$ -krotna i ma elementy a_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$, które są uporządkowane zgodnie z prostokątną tablicą składającą się z m wierszy i n kolumn:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

a_{kl} jest elementem wejściowym k -tego wiersza i l -tej kolumny.

Różne rodzaje macierzy

- ▶ **Macierz kwadratowa**: gdy liczba wierszy i kolumn jest taka sama, powiedźmy n , nazywamy to macierzą kwadratową stopnia n
- ▶ **Macierz diagonalna**: macierz kwadratowa, której jedynymi elementami niezerowymi są elementy na przekątnej
- ▶ **Macierz jednostka**: macierz kwadratowa posiadająca jedynki (1) na głównej przekątnej
- ▶ **Macierz trójkątna górna**: macierz kwadratowa posiadająca zera poniżej głównej przekątnej
- ▶ **Macierz trójkątna dolna** macierz kwadratowa posiadająca zera powyżej głównej przekątnej
- ▶ Macierze jednokolumnowe nazywamy **wektorami**

Operacje na macierzach

- **Transpozycja macierzy:** Mając daną macierz A formatu $m \times n$, A^t jest macierzą uzyskaną z A przez zamianę wierszy i kolumn macierzy oryginalnej.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

$$A^t = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

- Macierz kwadratowa A formatu n jest symetryczna, jeśli $A = A^t$.

Dodawanie macierzy

Suma dwóch macierzy A i B formatu $m \times n$ jest definiowana jako suma elementów.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}$$

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{pmatrix}$$

Właściwości dodawania macierzy

Przypuszczajmy $A = [a_{ij}]$, $B = [b_{ij}]$ i $C = [c_{ij}]$, a każda ma format $m \times n$.

1. (**Przemienność**) Kolejność dodawania macierzy nie ma znaczenia
$$A + B = B + A$$
2. (**Łączność**) Sposób grupowania macierzy przy dodawaniu nie wpływa na wynik
$$A + (B + C) = (A + B) + C$$
3. (**Element neutralny**): Istnieje unikalna macierz zerowa, która po dodaniu do dowolnej innej macierzy A nie zmienia jej wartości
$$A + 0 = A, \text{ gdzie } 0 \text{ to macierz zerowa formatu } m \times n$$
4. (**Macierz przeciwna**) Dla każdej macierzy A można znaleźć macierz przeciwną $(-A)$, która po dodaniu do macierzy A daje macierz zerową
$$A + (-A) = 0, \text{ gdzie } (-A) = [-a_{ij}]$$

Mnożenie macierzy

Niech A będzie macierzą formatu $m \times n$, a B macierzą formatu $n \times p$. Iloczyn A i B jest macierzą formatu $m \times p$ zdefiniowaną w następujący sposób.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{pmatrix}$$

$$AB = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mp} \end{pmatrix}$$

gdzie $c_{ij} = \sum_{l=1}^n a_{il} b_{lj}$, $i = 1, \dots, m$, $j = 1, \dots, p$

Właściwości mnożenia macierzy

1. **Nieprzemienność** (i) Przypuszczamy $A = [a_{ij}]$ z formatu 2×3 i $B = [b_{ij}]$ z formatu 3×4 . Wtedy AB istnieje, ale BA nie istnieje.
(ii) Nawet jeśli AB i BA istnieją, mogą nie być tym samym.
2. **Łączność** $A(BC) = (AB)C$
pod warunkiem $A = [a_{ij}]$ formatu $m \times n$, $B = [b_{jk}]$ formatu $n \times p$ i $C = [c_{kl}]$ formatu $p \times s$
3. **Element neutralny** Jeśli $A = [a_{ij}]$ z formatu $m \times n$, wtedy $AI_n = A = I_m A$ gdzie I_m i I_n są to macierze jednostkowe stopnia m i n odpowiednio.
4. Nie dla każdej macierzy A istnieje macierz odwrotna.
5. **Rozdzielność względem dodawania** Dla dwóch macierzy A i B formatu $m \times n$ i dwóch macierzy C i D formatu $n \times p$
 $(A + B)C = AC + BC$ i $A(C + D) = AC + AD$

Przykłady mnożenia macierzy w uczeniu maszynowym

- ▶ Załóżmy, że prosta sieć neuronowa ma wektor wejściowy $\vec{x} = (1, 0, 0, 5)$ i macierz wagową $W = \begin{pmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{pmatrix}$.

Do obliczenia wyniku warstwy ukrytej potrzebujemy

$$\vec{x} \cdot W = (1.0, 0.5) \cdot \begin{pmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{pmatrix}.$$

Wyznacznik macierzy kwadratowej

- ▶ Niektóre macierze kwadratowe posiadają macierze odwrotne
- ▶ Macierz kwadratową, A stopnia n dla której istnieje macierz odwrotną A^{-1} stopnia n , nazywamy odwracalną, albo nieosobliwą. Macierz, która nie posiada macierzy odwrotnej nazywamy osobliwą.
- ▶ **Warunek nieosobliwości macierzy**
 - ▶ Dla macierzy kwadratowej A rzędu n wyznacznik macierzy A , oznaczany jako $|A|$ lub $\det(A)$, jest liczbą (rzeczywistą) zdefiniowaną na podstawie różnych możliwych permutacji na $\{1, 2, \dots, n\}$.
 - ▶ **Przykład:** $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ Możliwe permutacje na $\{1, 2\}$ to $1 \rightarrow 1, 2 \rightarrow 2$ i $1 \rightarrow 2, 2 \rightarrow 1$.
 - ▶ Każda permutacja ma znak (dodatni lub ujemny) i określa, które elementy należy pomnożyć. Na przykład mnożymy $a_{11}a_{22}$ i $a_{12}a_{21}$. Następnie sumujemy iloczyn z ich znakami.
 - ▶ Szczególnie $|A| = a_{11}a_{22} - a_{12}a_{21}$.
 - ▶ Macierz A ma macierz odwrotną, jeśli $|A| \neq 0$, tzn., A nie jest osobliwa.

Permutacje na skończonym zbiorze liczb naturalnych i wyznacznika

- ▶ Niech A będzie macierzą kwadratową o rozmiarze n .
- ▶ Wyznacznik A , oznaczony jako $\det(A)$ lub $|A|$, jest zdefiniowany jako: $|A| = \sum_{\pi \in S_n} \text{Sign}(\pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)}$, gdzie S_n jest zbiorem wszystkich permutacji na $\{1, 2, 3, \dots, n\}$.

- ▶ **Przykład:** $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ Możliwe permutacje na $\{1, 2\}$:

$\pi_1 = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ i $\pi_2 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. π_1 nie ma pary odwrotnej, a π_2 ma jedną parę odwrotną. Więc $\text{Sign}(\pi_1) = (-1)^0$ (znak dodatni) i $\text{Sign}(\pi_2) = (-1)^1$ (znak ujemny). Wyznacznik A to $|A| = a_{11}a_{22} - a_{12}a_{21}$.

Odwrotność macierzy

- ▶ Dla macierzy kwadratowej A istnieje macierz odwrotna, oznaczana A^{-1} , jeśli $|A| \neq 0$.

- ▶ Niech $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$ i $|A| \neq 0$.

- ▶ $A^{-1} = \frac{1}{|A|} \text{Adj}(A)^t$ gdzie $\text{Adj}(A)$ (ang. Adjoint of A) to macierz dołączona macierzy A , zdefiniowana jako:

$$\text{Adj}(A) = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix} \text{ i } A_{ij} \text{ to kofaktor który}$$

odnosi się do wyznacznika podmacierzy otrzymanej przez usunięcie i -tego wiersza i j -tej kolumny z A i pomnożenie wartości przez znak $(-1)^{i+j}$.

Właściwości transpozycji i odwrotności

Niech A i B będą macierzami kwadratowymi o rozmiarze n .

- ▶ $AA^{-1} = I_n = A^{-1}A$.
- ▶ $(AB)^{-1} = B^{-1}A^{-1}$.
- ▶ $(A + B)^{-1} \neq A^{-1} + B^{-1}$.

Założmy, że A i B są macierzami o rozmiarze $m \times n$.

- ▶ $(A^t)^t = A$
- ▶ $(A + B)^t = A^t + B^t$.

Założmy, że A i B są macierzami kwadratowymi o rozmiarze $m \times n$ i $n \times p$ odpowiednio.

- ▶ $(AB)^t = B^t A^t$.

Operacja mnożenia skalarnego na macierzach

Niech $\mathcal{M}_{m \times n}$ będzie zbiorem wszystkich macierzy o rozmiarze $m \times n$ i $\lambda, \delta \in \mathbb{R}$ będą skalarami.

- ▶ $\lambda(\delta A) = \lambda\delta(A)$, $A \in \mathcal{M}_{m \times n}$.
- ▶ $\lambda(BC) = (\lambda B)C = B(\lambda C) = (BC)\lambda$, $B \in \mathcal{M}_{m \times n}$,
 $C \in \mathcal{M}_{n \times p}$.
- ▶ $\lambda(A + B) = \lambda A + \lambda B$, $A, B \in \mathcal{M}_{m \times n}$.
- ▶ $(\lambda + \delta)A = \lambda A + \delta A$, $A \in \mathcal{M}_{m \times n}$.

Taką operację mnożenia liczby rzeczywistej przez macierz można uważać za $\circ : \mathbb{R} \times \mathcal{M} \mapsto \mathcal{M}$ gdzie $\mathcal{M}$ jest zbiorem wszystkich macierzy i jest znana jako operacja zewnętrzna zwracająca macierz.

Układ równań liniowych i reprezentacja macierzowa

- Układ równań liniowych o liczbie równań m i liczbie zmiennych n :

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$\vdots$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

Liczby a_{ij} nazywamy i współczynnikami układu, a x_j to niewiadome lub zmienne ($i = 1, 2, \dots, m, j = 1, 2, \dots, n$).

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

$$AX = B$$

- Jeśli B to macierz zerowa $AX = B$ to układ jednorodny (ang. homogeneous) równań liniowych, a inaczej to układ niejednorodny (ang. non-homogeneous) równań liniowych.

Rozwiązywanie układu równań liniowych metodą odwracania macierzy

- ▶ W przypadku równań liniowych o tej samej liczbie zmiennych i równań (tzn. $m = n$)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

- ▶ Jeśli $|A| \neq 0$, znajdujemy A^{-1} .

- ▶ $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = A^{-1}B$

Elementarne działanie na macierzach

Kluczem do rozwiązania układu równań liniowych są elementarne przekształcenia, które zachowują ten sam zbiór rozwiązań, ale przekształcają układ równań do prostszej postaci

E1 Wymiana dwóch wierszy w macierzy

E2 Mnożenie wiersza w macierzy przez stałą (skalar) $\lambda \neq 0$

E3 Dodawanie dwóch wierszy macierzy

Wierszowa forma schodkowa macierzy (Row Echelon Form)

Niech A będzie macierz wymiaru $m \times n$ gdzie oznaczmy i -ty wiersz A jako $A^{(i,\cdot)}$, a j -ta kolumna A jako $A^{(\cdot,j)}$.

- ▶ **REF**: Macierz A jest w formie schodkowej (REF), jeśli A spełnia następujący warunek.
 - Albo $A^{(1,\cdot)}$ jest wierszem zerowym, albo pierwszy wpis różny od zera (od lewej do prawej) jest 1.
 - Dla dowolnego $i = 1, 2, \dots, m$, jeśli $A^{(i,\cdot)}$ jest wierszem zerowym, to wszystkie kolejne wiersze będą wierszami zerowymi.
 - Dla dowolnego $i = 1, 2, \dots, m$, jeśli $A^{(i,\cdot)}$ nie jest wierszem zerowym, to pierwszym niezerowym wpisem (od lewej do prawej) jest 1 i występuje on na prawo od początkowej 1 w $A^{(i-1,\cdot)}$.
- ▶ Początkowy element 1 w każdym wierszu, który nie jest wierszem zerowym, nazywany jest wiodącym elementem (ang. pivot element).
- ▶ **RREF**: Zredukowana forma schodkowa w macierzy: Każda kolumna $A^{(\cdot,j)}$ zawierająca element wiodący ma wszystkie pozostałe elementy równe zero.

Rząd macierzy

- **Rząd macierzy** to liczba niezerowych wierszy w jej postaci schodkowej (lub schodkowej zredukowanej)

- $A_1 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

A_1 jest w postaci REF i $\text{rzęd}(A_1) = 3$

- $A_2 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

A_2 jest w postaci REF i $\text{rzęd}(A_2) = 3$

Uzyskanie RREF macierzy

$$\begin{aligned} \blacktriangleright \begin{pmatrix} 2 & 5 & 3 \\ 1 & 2 & 3 \\ 1 & 0 & 8 \end{pmatrix} &\longrightarrow R^1 \leftrightarrow R^2 \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{pmatrix} \longrightarrow R^2 + (-2)R^1, R^3 + (-1)R^1 \\ \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -3 \\ 0 & -2 & 5 \end{pmatrix} &\longrightarrow R^3 + (2)R^2 \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & -1 \end{pmatrix} \longrightarrow (-1) \cdot R^3 \\ \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

Rozwiązywanie układu równań liniowych za pomocą elementarnych operacji na wierszach

Układ równań liniowych o liczbie równań m i liczbie zmiennych n :

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$\vdots$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

$$\rightarrow \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

- Przekształć macierz rozszerzoną $[A|B]$ do postaci schodkowej

wierszowej gdzie $[A|B] = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right)$

Przykład

$$2x_1 + 5x_2 + 3x_3 = 5$$

$$x_1 + 2x_2 + 3x_3 = 4$$

$$x_1 + 8x_3 = 9$$

► $[A|B] = \begin{pmatrix} 2 & 5 & 3|5 \\ 1 & 2 & 3|4 \\ 1 & 0 & 8|9 \end{pmatrix}$

- Konwertuj do formy schodkowej aby uzyskać rozwiązanie $x_1 = 1$, $x_2 = 0$ i $x_3 = 1$. item Proces ten polegający na przekształceniu układu równań liniowych do postaci schodkowej zredukowanej (RREF) i znalezieniu rozwiązań dla układu równań liniowych nazywa się algorytmem eliminacji Gaussa.

Rozwiązania jednorodnych i niejednorodnych układów równań liniowych

- ▶ Jednorodnych układ równań liniowych zawsze ma co najmniej jedno rozwiązanie w którym wszystkie zmienne mają wartość zero. Może mieć nieskończenie wiele rozwiązań, łącznie z rozwiązaniem zawierającym wszystkie wartości zerowe.
- ▶ Niejednorodnych układ równań liniowych może mieć jedno rozwiązanie, nieskończenie wiele rozwiązań lub nie mieć rozwiązań.

Warunki istnienia rozwiązań układu równań liniowych

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$\vdots$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

► $[A|B] = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right)$

- Znajdź rząd A i rząd $[A|B]$.
- Jeśli $\text{rząd}([A|B]) = \text{rząd}(A)$ wtedy układ równań liniowych ma rozwiązania. Dodatkowo, jeśli $m < n$, to rozwiązań jest nieskończenie wiele. W przeciwnym razie istnieje dokładnie jedno rozwiązanie.
 - Jeśli $\text{rząd}([A|B]) \neq \text{rząd}(A)$ wtedy układ równań liniowych jest niespójny i nie ma żadnego rozwiązania.

Przykłady

1. $x_1 - 2x_2 + x_3 - x_4 = 3$
 $2x_1 - 4x_2 + x_3 = x_4 = 2$
 $x_1 - 2x_2 - 2x_3 + 3x_4 = 1$

► $[A|B] \text{ (REF)} = \left(\begin{array}{cccc|c} 1 & -2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & -2 \end{array} \right)$

- Ma nieskończenie wiele rozwiązań, ponieważ x_1, x_3, x_4 są zmiennymi wiodącymi, a x_2 jest zmienną wolną.
- Przestrzeń rozwiązań = $\{(3 + 2x_2, x_2, -2, -2) : x_2 \in \mathbb{R}\}$.

Odwrotność macierzy poprzez zredukowaną formę schodkową wierszową

- ▶ Niech A będzie macierzą kwadratową rzędu n i szukamy X takiego, że $AX = I_n$.
- ▶ Zaczynamy od kroku: $[A|I_n]$

$$\left(\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1n} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 & 0 & \dots & 1 \end{array} \right)$$

- ▶ Konwertuj macierz rozszerzoną w formacie RREF (zredukowana forma wierszowa Echelon)
- ▶ $[I_n|A^{-1}]$

Przykład

► Niech $A = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$

► $\left(\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{array} \right)$

↓ RREF

► $\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -1 & 2 & -2 & 2 \\ 0 & 1 & 0 & 0 & 1 & -1 & 2 & -2 \\ 0 & 0 & 1 & 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 2 \end{array} \right)$

► $A = \begin{pmatrix} -1 & 2 & -2 & 2 \\ 1 & -1 & 2 & -2 \\ 1 & -1 & 1 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}$

Wektory: specjalne formy macierzy

- ▶ Wektor n -wymiarowy można uważać albo za macierz o rozmiarze $1 \times n$, albo o rozmiarze $n \times 1$. Pierwszy typ możemy nazwać wektorami wierszowymi, a drugi typem wektorami kolumnowymi.
- ▶ $\mathbb{R}^n = \{(a_1, a_2, \dots, a_n) : a_i \in \mathbb{R}, 1 \leq i \leq n\}$ zawiera takie n -wymiarowe (wierszowe) wektory i nazywamy je przestrzenią euklidesową o wymiarze n .
- ▶ Tak więc operacje takie jak dodawanie macierzy i mnożenie skalarne są możliwe w $\mathbb{R}^n$ - zwracając wektor z $\mathbb{R}^n$.
- ▶ Zanim zaczniemy uczyć się o przestrzeni wektorowej, zobaczmy, jaką strukturę algebraiczną tworzy taka przestrzeń.

Podstawowe struktury algebraiczne

- ▶ **Grupa:** Mając zbiór G i operację binarną $*$ na G , $(G, *)$ jest nazywana grupą, jeśli następujące warunki są spełnione.

- (i) $x * (y * z) = (x * y) * z$ dla dowolnych $x, y, z \in G$
- (ii) Istnieje $e \in G$ takie, że dla każdego $x \in G$, $x * e = x = e * x$.
- (iii) Dla każdego $x \in G$ istnieje $x^{-1} \in G$ takie, że $x * x^{-1} = e = x^{-1} * x$.

$(G, *)$ nazywana jest **grupą abelową (przemienną)**, jeśli dla dowolnych $x, y \in G$, $x * y = y * x$.

- ▶ **Przykłady:** $(\mathbb{Z}, +)$, $(\mathbb{R}, +)$, $(\mathbb{R} \setminus \{0\}, \cdot)$, $(\mathcal{M}_{m \times n}, +)$

Ciało: Ciałem nazywamy zbiór F z dwoma dwuargumentowymi działaniami wewnętrznymi, $\oplus$ i $\odot$, oznaczamy przez

$(F, \oplus, \odot)$, spełniające następujące warunki.

- (i) $(F, \oplus)$ to grupa przemienna
- (ii) $(F \setminus \{0\}, \odot)$ to grupa przemienna gdzie 0 to neutralny element w odniesieniu do operacji $\oplus$.
- (iii) $x \odot (y \oplus z) = (x \odot y) \oplus (x \odot z)$, dla dowolnych $x, y, z \in F$.

- ▶ **Przykład:** $(\mathbb{R}, +, \cdot)$

Przestrzeń liniowa nad ciałem liczb rzeczywistych

Pojęcie przestrzeni liniowej definiuje się za pomocą dwóch struktur algebraicznych – wektory tworzą grupę przemenną, a skalary pochodzą z ciała K . Tutaj skupimy się na $K = \mathbb{R}$.

Liniowa (wektorowa) przestrzeń $\mathcal{V}$ nad ciałem liczb rzeczywistych, oznaczana przez $(\mathcal{V}, +)$, spełnia następujące własności.

- (i) $(\mathcal{V}, +)$ jest grupą przemenną.
- (ii) Istnieje operacja zewnętrzna $\cdot : \mathbb{R} \times \mathcal{V} \mapsto \mathcal{V}$ taka, że spełnione są następujące warunki.
 - $\lambda(\overrightarrow{x + y}) = \lambda \overrightarrow{x} + \lambda \overrightarrow{y}$, $\overrightarrow{x}, \overrightarrow{y} \in \mathcal{V}$ i $\lambda \in \mathbb{R}$.
 - $(\lambda + \delta) \overrightarrow{x} = \lambda \overrightarrow{x} + \delta \overrightarrow{x}$, $\overrightarrow{x} \in \mathcal{V}$ i $\lambda, \delta \in \mathbb{R}$.
 - $\lambda(\delta \overrightarrow{x}) = (\lambda\delta) \overrightarrow{x}$, $\overrightarrow{x} \in \mathcal{V}$ i $\lambda, \delta \in \mathbb{R}$.
 - $1 \cdot \overrightarrow{x} = \overrightarrow{x}$, $\overrightarrow{x} \in \mathcal{V}$, $1 \in \mathbb{R}$.

Elementy $x \in \mathcal{V}$ nazywane są wektorami. Elementem neutralnym $(\mathcal{V}, +)$ jest wektor zerowy $\overrightarrow{0} = [0, \dots, 0]$, a działanie wewnętrzne $+$ nazywane jest dodawaniem wektorów. Elementy $\lambda \in \mathbb{R}$ nazywane są skalarami, a działanie zewnętrzne to mnożenie przez skalary.

Przykłady

- ▶ $(\mathcal{M}_{m \times n}, +)$ nad ciałem $\mathbb{R}$ jest przestrzenią liniową.
- ▶ $(\mathbb{R}^n, +) = (\mathcal{M}_{1 \times n}, +)$ nad ciałem $\mathbb{R}$ jest przestrzenią liniową.
Dla dowolnych $\vec{x}, \vec{y} \in \mathbb{R}^n$, $\vec{x} = [x_1, x_2, \dots, x_n]$, $\vec{y} = [y_1, y_2, \dots, y_n]$, mamy
 - (i) $\vec{x} + \vec{y} = [x_1 + y_1, x_2 + y_2, \dots, x_n + y_n]$ oraz
 - (ii) $\lambda \vec{x} = [\lambda x_1, \lambda x_2, \dots, \lambda x_n]$.

Przetwarzanie danych z wykorzystaniem operacji z przestrzeni liniowej

- ▶ Wyobraź sobie, że otwierasz zbiór danych: wiersze rekordów klientów, kolumny takie jak wiek, dochód, wydatków.
- ▶ Jeśli potraktujesz każdy wiersz jako wektor, od razu wkraczasz na teren algebry liniowej. A kiedy połączysz wszystkie wiersze (wiersze $\times$ kolumny), otrzymasz macierz.
- ▶ Aby każda jednostka wartości dla wszystkich wyżej wymienionych cech kolumn była standaryzowana, potrzebna jest operacja mnożenia skalarnego przestrzeni liniowej.
- ▶ Aby sprawdzić, czy dwa punkty danych reprezentujące informacje dwóch klientów są zależne czy nie, pomocne może okazać się mnożenie skalarne wraz z dodawaniem wektorów.

Podprzestrzenie liniowe

- **Definicja:** Niech $(\mathcal{V}, +)$ będzie przestrzenią liniową nad $\mathbb{R}$ i $\mathcal{U} \subseteq \mathcal{V}$. $(\mathcal{U}, +)$ nazywa się liniową podprzestrzenią $\mathcal{V}$, jeśli $(\mathcal{U}, +)$ jest przestrzenią liniową nad $\mathbb{R}$.
- Jeśli $\mathcal{U}$ jest podprzestrzenią przestrzeni liniowej $\mathcal{V}$ nad ciałem liczb rzeczywistych, możemy ją oznaczyć przez $\mathcal{U}_{|\mathbb{R}} \subseteq \mathcal{V}_{|\mathbb{R}}$.
- **Twierzenie:** Niech $(\mathcal{V}, +)$ będzie przestrzenią liniową nad $\mathbb{R}$ i $\mathcal{U} \subseteq \mathcal{V}$. $\mathcal{U}$ jest liniową podprzestrzenią $\mathcal{V}$, jeśli
 - (i) $\vec{0} \in \mathcal{U}$,
 - (ii) dla dowolnych $\vec{x}, \vec{y} \in \mathcal{U}$, $\overrightarrow{x+y} \in \mathcal{U}$, i
 - (iii) dla dowolnej $\lambda \in \mathbb{R}$, $\vec{x} \in \mathcal{U}$, $\lambda \vec{x} \in \mathcal{U}$.
- **Twierdzenie:** Część wspólna dowolnej rodziny podprzestrzeni przestrzeni liniowej $\mathcal{V}_{|\mathbb{R}}$ jest też podprzestrzenią $\mathcal{V}_{|\mathbb{R}}$.

Przykłady podprzestrzeni z macierzy

- ▶ Wiemy już, że $(\mathcal{M}_{m \times n}, +)$ jest przestrzenią liniową nad $\mathbb{R}$. Oznaczmy tę szczególną przestrzeń liniową jako $\mathcal{M}_{m \times n|\mathbb{R}}$.
- ▶ Rozważmy najpierw następujące podzbiory $\mathcal{M}_{m \times n}$.
 - Zbiór wszystkich macierzy trójkątnych górnych:
$$UTri^{m,n} = \{A \in \mathcal{M}_{m \times n} : \forall i > j, a_{ij} = 0\}$$
 - Zbiór wszystkich dolnych macierzy trójkątnych:
$$LTri^{m,n} = \{A \in \mathcal{M}_{m \times n} : \forall i < j, a_{ij} = 0\}$$
 - Zbiór wszystkich macierzy przekątnych:
$$Diag^{m,n} = \{A \in \mathcal{M}_{m \times n} : \forall i \neq j, a_{ij} = 0\}$$
- ▶ $UTri_{|\mathbb{R}|}^{m,n} \subseteq \mathcal{M}_{m \times n|\mathbb{R}}$
- ▶ $LTri_{|\mathbb{R}|}^{m,n} \subseteq \mathcal{M}_{m \times n|\mathbb{R}}$
- ▶ $Diag_{|\mathbb{R}|}^{m,n} \subseteq \mathcal{M}_{m \times n|\mathbb{R}}$

Rozpiętość przestrzeni liniowej

Niech $\mathcal{V}_{|\mathbb{R}}$ będzie przestrzenią liniową i $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} \subseteq \mathcal{V}$.

- ▶ **Kombinacja liniowa:** Wektor $\vec{v} \in \mathcal{V}$ nazywamy kombinacją liniową $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$, jeśli istnieją skalary $a_1, a_2, \dots, a_n \in \mathbb{R}$ takie, że $\vec{v} = \sum_{i=1}^n a_i \vec{v}_i$.
- ▶ **Rozpiętość zbioru wektorów** $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$, oznaczona jako $\text{Span}(\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\})$, to zbiór wektorów, które można wyrazić jako kombinację liniową $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$. Tzn.,
 $\text{Span}(\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}) = \{\vec{v} \in \mathcal{V} : \vec{v} = \sum_{i=1}^n a_i \vec{v}_i\}$
- ▶ **Przykład:** Rozważmy przestrzeń liniową $\mathbb{R}_{|\mathbb{R}}^2$ i wektory $(0, 1), (1, 1) \in \mathbb{R}_{|\mathbb{R}}^2$. Niech $(x, y) \in \mathbb{R}_{|\mathbb{R}}^2$.
 - Sprawdźmy, czy (x, y) jest kombinacją liniową $(0, 1)$ i $(1, 1)$. Tzn., szukamy takie $a_1, a_2 \in \mathbb{R}$, że $(x, y) = a_1(0, 1) + a_2(1, 1)$
 - Wtedy mamy układ równań liniowych $\begin{cases} a_2 = x \\ a_1 + a_2 = y \end{cases}$
 - Więc, $(x, y) = (y - x)(0, 1) + x(1, 1)$.
 - Ponieważ (x, y) jest dowolnym wektorem $\mathbb{R}_{|\mathbb{R}}^2$,

$$\text{Span}(\{(0, 1), (1, 1)\}) = \mathbb{R}_{|\mathbb{R}}^2.$$

Liniowa (nie)zależność

Niech $\mathcal{V}_{|\mathbb{R}}$ będzie przestrzenią liniową i $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} \subseteq \mathcal{V}$.

- ▶ $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ nazywa się **liniowo niezależnym**, jeśli rozwiązanie zerowe jest jedynym rozwiązaniem dla układu równań otrzymanych z $a_1 \vec{v}_1 + a_2 \vec{v}_2 + \dots + a_n \vec{v}_n = \vec{0}$. (tzn., jedynym rozwiązaniem jest $a_1 = a_2 = \dots = a_n = 0$.)
- ▶ Jeśli istnieje rozwiązanie niezerowe spełniające $a_1 \vec{v}_1 + a_2 \vec{v}_2 + \dots + a_n \vec{v}_n = \vec{0}$, wtedy $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ nazywa się **liniowo zależnym**.
- ▶ **Przykład:** Rozważmy przestrzeń liniową $\mathbb{R}_{|\mathbb{R}}^2$ i wektory $(0, 1), (1, 1) \in \mathbb{R}_{|\mathbb{R}}^2$.
 - Już wiemy, że $(x, y) = (y - x)(0, 1) + x(1, 1)$.
 - Aby przedstawić $(0, 0)$ jako (x, y) mamy tylko jedno rozwiązanie $x = y = 0$.
 - Więc, $\{(0, 1), (1, 1)\}$ jest liniowo niezależnym.

Baza i wymiar przestrzeni

Definicja: Niech $\mathcal{V}_{|\mathbb{R}}$ będzie przestrzenią liniową i $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} \subseteq \mathcal{V}$. Dla dowolnej podprzestrzeni $\mathcal{U}_{|\mathbb{R}} \subseteq \mathcal{V}_{|\mathbb{R}}$, $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ nazywamy **bazą** przestrzeni $\mathcal{U}_{|\mathbb{R}}$ jeśli

- (i) $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ jest liniowo niezależnym i
- (ii) $\text{Span}(\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}) = \mathcal{U}_{|\mathbb{R}}$.

Twierdzenie: Każda przestrzeń liniowa $\mathcal{U}_{|\mathbb{R}}$ ($\subseteq \mathcal{V}_{|\mathbb{R}}$) ma bazę. Ponadto, wszystkie bazy są równoliczne.

Definicja: Liczbę elementów bazy danej przestrzeni $\mathcal{U}_{|\mathbb{R}}$ nazywamy jej **wymiarem** i oznaczamy $\dim(\mathcal{U}_{|\mathbb{R}})$.

- ▶ **Przykład:** Już zobaczyliśmy $\{(0, 1), (1, 1)\}$ jest liniowo niezależnym i $\text{Span}(\{(0, 1), (1, 1)\}) = \mathbb{R}_{|\mathbb{R}}^2$. Więc, $\{(0, 1), (1, 1)\}$ jest bazą przestrzeni $\mathbb{R}_{|\mathbb{R}}^2$ i $\dim(\mathbb{R}_{|\mathbb{R}}^2) = 2$.
- ▶ Warto zauważyć, że przestrzeń liniowa może mieć różne bazy, ale wszystkie powinny mieć taką samą liczbę elementów.
- ▶ Dla przestrzeni $\mathbb{R}_{|\mathbb{R}}^2$, $\{(1, 0), (0, 1)\}$ jest **bazą standardową**.

Operacje na przestrzeniach liniowych

- ▶ Niech $\mathcal{U}_{|\mathbb{R}}$ i $\mathcal{W}_{|\mathbb{R}}$ będą podprzestrzeniami $\mathcal{V}_{|\mathbb{R}}$. Definiujemy iloczyn i sumę tych podprzestrzeni jako

- $\mathcal{U}_{|\mathbb{R}} \cap \mathcal{W}_{|\mathbb{R}} = \{ \vec{x} \in \mathcal{V}_{|\mathbb{R}} : \vec{x} \in \mathcal{U}_{|\mathbb{R}} \text{ i } \vec{x} \in \mathcal{W}_{|\mathbb{R}} \}$

-

$$\mathcal{U}_{|\mathbb{R}} + \mathcal{W}_{|\mathbb{R}} = \{ \vec{x} \in \mathcal{V}_{|\mathbb{R}} : \vec{x} = \vec{y} + \vec{z}, \vec{y} \in \mathcal{U}_{|\mathbb{R}} \text{ i } \vec{z} \in \mathcal{W}_{|\mathbb{R}} \}.$$

- ▶ **Twierdzenie:** Zarówno iloczyn $\mathcal{U}_{|\mathbb{R}} \cap \mathcal{W}_{|\mathbb{R}}$ jak i suma $\mathcal{U}_{|\mathbb{R}} + \mathcal{W}_{|\mathbb{R}}$ są podprzestrzeniami $\mathcal{V}_{|\mathbb{R}}$.
- ▶ **Twierdzenie:** Niech $\mathcal{V}_{|\mathbb{R}}$ będzie przestrzenią liniową i $v_1, v_2, \dots, v_n \in \mathcal{V}_{|\mathbb{R}}$. $\text{Span}(\{v_1, v_2, \dots, v_n\})$ jest podprzestrzenią liniową $\mathcal{V}_{|\mathbb{R}}$.
- ▶ Jeśli $\mathcal{V}_{|\mathbb{R}}$ jest skończenie wymiarową przestrzenią liniową, to, dla dowolnej podprzestrzeni $\mathcal{U}_{|\mathbb{R}}$ przestrzeni $\mathcal{V}_{|\mathbb{R}}$, $\dim(\mathcal{U}_{|\mathbb{R}}) \leq \dim(\mathcal{V}_{|\mathbb{R}})$.

Znajdowania bazy $U = \text{Span}(\{v_1, v_2, \dots, v_n\})$

1. Zapisz wektory rozpinające jako kolumny macierzy A .
2. Określ postać schodkową macierzy A .
3. Wektory rozpinające powiązane z wiodącymi kolumnami tworzą bazę U .

Przykład: Rozważmy przestrzeń liniową $\mathbb{R}_{|\mathbb{R}}^3$. Niech $U = \text{Span}(\{(1, 0, 0), (0, 1, 0), (1, 1, 0)\})$. Znajdź bazę i wymiar tej podprzestrzeni.

Znajdowania bazy $U = \text{Span}(\{v_1, v_2, \dots, v_n\})$

1. Zapisz wektory rozpinające jako kolumny macierzy A .
2. Określ postać schodkową macierzy A .
3. Wektory rozpinające powiązane z wiodącymi kolumnami tworzą bazę U .

Przykład: Rozważmy przestrzeń liniową $\mathbb{R}_{|\mathbb{R}}^3$. Niech $U = \text{Span}(\{(1, 0, 0), (0, 1, 0), (1, 1, 0)\})$. Znajdź bazę i wymiar tej podprzestrzeni.

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \hookrightarrow \text{Już jest w formie REF.}$$

Więc baza U to $\{(1, 0, 0), (0, 1, 0)\}$ i ma wymiar 2.

Znajdowania bazy $U = \text{Span}(\{v_1, v_2, \dots, v_n\})$

1. Zapisz wektory rozpinające jako kolumny macierzy A .
2. Określ postać schodkową macierzy A .
3. Wektory rozpinające powiązane z wiodącymi kolumnami tworzą bazę U .

Przykład: Rozważmy przestrzeń liniową $\mathbb{R}_{|\mathbb{R}}^3$. Niech $U = \text{Span}(\{(1, 0, 0), (0, 1, 0), (1, 1, 0)\})$. Znajdź bazę i wymiar tej podprzestrzeni.

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \hookrightarrow \text{Już jest w formie REF.}$$

Więc baza U to $\{(1, 0, 0), (0, 1, 0)\}$ i ma wymiar 2.

$\hookrightarrow$ **Warto zauważyć**, że liczba liniowo niezależnych kolumn macierzy A jest taka sama jak liczba liniowo niezależnych wierszy macierzy A i jest nazywana rzędem macierzy A .

Właściwości rzędu macierzy

- ▶ Dla dowolnej $A = [a_{ij}]_{m \times n}$, $\text{rank}(A) = \text{rank}(A^t)$.
- ▶ Kolumny macierzy A rozpinają podprzestrzeń U przestrzeni liniowej $\mathbb{R}^m$ i $\dim(U) = \text{rank}(A)$. Bazę U można znaleźć stosując eliminację Gaussa do A w celu identyfikacji kolumn wiodących.
- ▶ Wiersze macierzy A rozpinają podprzestrzeń W przestrzeni liniowej $\mathbb{R}^n$ i $\dim(W) = \text{rank}(A)$. Bazę W można znaleźć stosując eliminację Gaussa do A^t .
- ▶ Dla każdej macierzy kwadratowej A rzędu n , A jest regularna (odwracalna) wtedy i tylko wtedy, gdy $\text{rank}(A) = n$.

Przekształcenia liniowe

- ▶ **Definicja:** Niech $\mathcal{V}|_K$ i $\mathcal{W}|_K$ będą dwiema przestrzeniami liniowymi nad tym samym ciałem K . Przekształcenie $f : \mathcal{V} \mapsto \mathcal{W}$ nazywamy przekształceniem liniowym $\mathcal{V}$ w $\mathcal{W}$ jeśli
 - (i) dla dowolnych $\vec{v}_1, \vec{v}_2 \in \mathcal{V}$, $f(\vec{v}_1 + \vec{v}_2) = f(\vec{v}_1) + f(\vec{v}_2)$ i
 - (ii) $f(\lambda \vec{v}_1) = \lambda f(\vec{v}_1)$ dla dowolnego $\lambda \in K$.
- ▶ Warunki (i) i (ii) można również łączyć ze sobą jako
$$f(\lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2) = \lambda_1 f(\vec{v}_1) + \lambda_2 f(\vec{v}_2) \text{ dla dowolnych } \vec{v}_1, \vec{v}_2 \in \mathcal{V} \text{ i } \lambda_1, \lambda_2 \in K.$$
- ▶ Ponieważ tutaj skupiamy się tylko na przestrzeniach liniowych w ciele liczb rzeczywistych, K jest rozpatrywane jako $\mathbb{R}$.

Przykłady

- (1) Niech $T : \mathbb{R}_{|\mathbb{R}}^2 \mapsto \mathbb{R}_{|\mathbb{R}}$ będzie funkcją z dwuwymiarowej przestrzeni euklidesowej do jednowymiarowej przestrzeni euklidesowej zdefiniowanej jako $T(x, y) = 2x - y$. Sprawdź czy jest to przekształcenie liniowe.

Przykłady

- (1) Niech $T : \mathbb{R}_{|\mathbb{R}}^2 \mapsto \mathbb{R}_{|\mathbb{R}}$ będzie funkcją z dwuwymiarowej przestrzeni euklidesowej do jednowymiarowej przestrzeni euklidesowej zdefiniowanej jako $T(x, y) = 2x - y$. Sprawdź czy jest to przekształcenie liniowe.
- Dla dowolnych $(x_1, y_1), (x_2, y_2) \in \mathbb{R}_{|\mathbb{R}}^2$,
$$T(x_1 + x_2, y_1 + y_2) = 2(x_1 + x_2) - (y_1 + y_2)$$
$$= (2x_1 - y_1) + (2x_2 - y_2) = T(x_1, y_1) + T(x_2, y_2)$$
 - Dla dowolnego $(x_1, y_1) \in \mathbb{R}_{|\mathbb{R}}^2$ i $\lambda \in \mathbb{R}$,
$$T(\lambda x_1, \lambda y_1) = 2\lambda x_1 - \lambda y_1 = \lambda(2x_1 - y_1) = \lambda T(x_1, y_1)$$
- Więc, T jest przekształceniem liniowym.
- (2) Niech $T : \mathbb{R}_{|\mathbb{R}}^2 \mapsto \mathbb{R}_{|\mathbb{R}}^3$ będzie funkcją z dwuwymiarowej przestrzeni euklidesowej do jednowymiarowej przestrzeni euklidesowej zdefiniowanej jako $T(x, y) = (y, x + 2y, x + y)$. Sprawdź czy jest to przekształcenie liniowe.

Przekształcenia liniowe w uczeniu maszynowym

- ▶ **Dimensionality reduction** : Przekształcenia liniowe są wykorzystywane do redukcji wymiarowości
 - Techniki PCA (**Principal Component Analysis**) (Analiza Głównych Składowych) wykorzystują przekształcenia liniowe w celu ograniczenia liczby cech w zestawie danych, przy jednoczesnym zachowaniu najważniejszych informacji, co przyczynia się do zwiększenia wydajności modeli.
- ▶ **Feature Engineering** : Przekształcenia liniowe pozwalają tworzyć nowe cechy z istniejących. Na przykład łączenie lub manipulowanie cechami za pomocą kombinacji liniowych może zapewnić bardziej użyteczną reprezentację modelu.
- ▶ **Data preprocessing** : Przekształcenia liniowe są używane do przygotowania danych przez normalizację i standaryzację w celu skalowania cech, zapewniając ich zrównoważony wkład do modelu.
- ▶ **Artificial Neural Network** : Mnożenie wektora przez macierz wagową to przekształcenie liniowe, która pomaga sieci neuronowej uczyć się skomplikowanych wzorców.

Obraz, jądro i rząd przekształcenia liniowego

- ▶ **Definicja:** Niech $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{W}_{|\mathbb{R}}$ będzie przekształceniem liniowym.

(i) Definiujemy obraz przekształcenia liniowego T jako

$$Im(T) = \{ T(\vec{v}) : \vec{v} \in \mathcal{V}_{|\mathbb{R}} \}.$$

(ii) Definiujemy jądro przekształcenia liniowego T jako

$$Ker(T) = \{ \vec{v} \in \mathcal{V}_{|\mathbb{R}} : T(\vec{v}) = \vec{0} \}.$$

(iii) Definiujemy rząd przekształcenia liniowego T jako

$$rank(T) = dim(Im(T)).$$

- ▶ **Twierdzenie:** Dla dowolnego przekształcenia liniowego $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{W}_{|\mathbb{R}}$, $dim(\mathcal{V}_{|\mathbb{R}}) = dim(Ker(T)) + dim(Im(T))$.
- ▶ Jądro (ang. kernel) przekształcenia liniowego T ($Ker(T)$) często nazywane jest przestrzenią zerową (ang. null space) T , i oznaczone jako $Null(T)$.

Przykład

- ▶ Rozważmy przekształcenie liniowe $T : \mathbb{R}_{|\mathbb{R}}^2 \mapsto \mathbb{R}_{|\mathbb{R}}$ zdefiniowane jako $T(x, y) = 2x - y$. Znajdź $\text{Null}(T)$.
- ▶ $\text{Null}(T) = \{(x, y) : T(x, y) = 0\} = \{(x, y) : 2x - y = 0\}$
 $= \{(x, 2x) : x \in \mathbb{R}\}$
- ▶ $\text{Null}(T) = \text{Span}(\{(1, 2)\})$
- ▶ $\dim(\text{Null}(T)) = 1$ bo pojedynczy wektor różny od zera jest zawsze liniowo niezależny.

Reprezentacja macierzowa przekształceń liniowych

- ▶ Rozważmy przestrzeń liniową $\mathcal{V}_{\mathbb{R}}$. Przez **uporządkowaną bazę** $B = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$ tej przestrzeni $\mathcal{V}_{\mathbb{R}}$ rozumiemy bazę rozpatrywaną w określonym porządku. W przeciwieństwie do bazy, tutaj kolejność wektorów ma znaczenie.
- ▶ Rozważmy przestrzeń liniową $\mathcal{V}_{\mathbb{R}}$ i uporządkowaną bazę $B = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$. Dla dowolnego wektora $\vec{v} \in \mathcal{V}_{\mathbb{R}}$ otrzymujemy unikalną reprezentację (kombinację liniową)
$$\vec{v} = a_1 \vec{v}_1 + a_2 \vec{v}_2 + \dots + a_n \vec{v}_n.$$
- ▶ $a_1, a_2, \dots, a_n$ to współrzędne $\vec{v}$ względem B , a $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ jest uważane za **reprezentację współrzędnych** (ang. coordinate representation) $\vec{v}$ względem B .
- ▶ Baza skutecznie definiuje układ współrzędnych. Znamy dwuwymiarowy kartezjański układ współrzędnych, który jest rozpięty przez standardowe wektory bazowe $(1, 0)$, $(0, 1)$.

Reprezentacja współrzędnych zależy od wybranej bazy

- ▶ Jednakże każda baza $\mathbb{R}_{|\mathbb{R}}^2$ definiuje poprawny układ współrzędnych, a ten sam wektor (x, y) może mieć różną reprezentację współrzędnych w zależności od innej bazy.
- ▶ **Przykład:**
 $(3, 7) = 3(1, 0) + 7(0, 1)$ względem bazy $\{(1, 0), (0, 1)\}$
Ale $(3, 7) = 4(0, 1) + (3)(1, 1)$ względem $\{(0, 1), (1, 1)\}$.

Macierz reprezentująca przekształcenia liniową

- ▶ Niech $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{W}_{|\mathbb{R}}$ będzie przekształceniem liniowym, a $(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$ i $(\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m)$ będą odpowiednio uporządkowanymi bazami dla $\mathcal{V}_{|\mathbb{R}}$ i $\mathcal{W}_{|\mathbb{R}}$. Więc, dla każdego $\vec{v}_j$ ($1 \leq j \leq n$), $T(\vec{v}_j)$ ma unikalną reprezentację $T(\vec{v}_j) = a_{1j}\vec{w}_1 + a_{2j}\vec{w}_2 + \dots + a_{mj}\vec{w}_m$ względem $(\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m)$.
- ▶ Tworzy to macierz $M(T)$ wymiaru $m \times n$ dla przekształcenia liniowego T , gdzie (i, j) -ty element macierzy to a_{ij} .
- ▶ Tzn., z każdej unikalnej reprezentacji $T(\vec{v}_j) = a_{1j}\vec{w}_1 + a_{2j}\vec{w}_2 + \dots + a_{mj}\vec{w}_m$, $1 \leq j \leq n$, mamy

$$M(T) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Przykłady

- ▶ Rozważmy wcześniej omówiony przykład przekształcenia liniowego $T : \mathbb{R}_{|\mathbb{R}}^2 \mapsto \mathbb{R}_{|\mathbb{R}}^3$ zdefiniowanego jako
$$T(x, y) = (y, x + 2y, x + y).$$
- ▶ Rozważmy również bazy standardowe $\mathbb{R}_{|\mathbb{R}}^2$ i $\mathbb{R}_{|\mathbb{R}}^3$ jako odpowiednio uporządkowane bazy.
- ▶
$$\begin{aligned} T(1, 0) &= (0, 1, 1) = 0(1, 0, 0) + 1(0, 1, 0) + 1(0, 0, 1) \\ T(0, 1) &= (1, 2, 1) = 1(1, 0, 0) + 2(0, 1, 0) + 1(0, 0, 1) \end{aligned}$$
- ▶
$$M(T) = \begin{pmatrix} 0 & 1 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}$$
- ▶ Teraz, jeśli zmienimy którąkolwiek z uporządkowanych baz rozważanych dla $\mathbb{R}_{|\mathbb{R}}^2$ i $\mathbb{R}_{|\mathbb{R}}^2$, reprezentacja macierzowa ulegnie zmianie. Np. zmieńmy tylko bazę dla $\mathbb{R}_{|\mathbb{R}}^2$ i rozważmy $((0, 1), (1, 1))$ jako uporządkowaną nową bazą.
- ▶ Wtedy
$$M(T) = \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 2 \end{pmatrix}$$

Przekształcenie liniowe uzyskane z macierzy

- ▶ Niech $A = [a_{ij}]$ będzie macierz o wymiarze $m \times n$. Możemy zdefiniować odpowiednio przekształcenie liniowe $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{W}_{|\mathbb{R}}$, gdzie $\mathcal{V}_{|\mathbb{R}}$ i $\mathcal{W}_{|\mathbb{R}}$ są przestrzeniami liniowymi o wymiarach odpowiednio n i m , w następujący sposób.
- ▶ Najpierw, musimy rozważyć uporządkowane bazy dla $\mathcal{V}_{|\mathbb{R}}$ i $\mathcal{W}_{|\mathbb{R}}$, a niech będą to odpowiednio $(\vec{v}_1, \dots, \vec{v}_n)$ i $(\vec{w}_1, \dots, \vec{w}_m)$.
- ▶ Zdefiniujemy, $T(\vec{v}_j) = \sum_{i=1}^m a_{ij} \vec{w}_i$
- ▶ Dla dowolnego $\vec{v} \in \mathcal{V}_{|\mathbb{R}}$, istnieje unikalna reprezentacja, takie jak $\vec{v} = b_1 \vec{v}_1 + b_2 \vec{v}_2 + \dots + b_n \vec{v}_n$
$$\begin{aligned} T(\vec{v}) &= T(b_1 \vec{v}_1 + b_2 \vec{v}_2 + \dots + b_n \vec{v}_n) \\ &= b_1 T(\vec{v}_1) + b_2 T(\vec{v}_2) + \dots + b_n T(\vec{v}_n) \\ &= b_1 \sum_{i=1}^m a_{i1} \vec{w}_i + b_2 \sum_{i=1}^m a_{i2} \vec{w}_i + \dots + b_n \sum_{i=1}^m a_{in} \vec{w}_i \end{aligned}$$

Przykład

- ▶ Niech $A = \begin{pmatrix} 2 & 1 \\ -3 & -2 \end{pmatrix}$
- ▶ Zdefiniujemy przekształcenie liniowe $T : \mathbb{R}_{|\mathbb{R}}^2 \mapsto \mathbb{R}_{|\mathbb{R}}^2$ względem bazy $((1, 0), (0, 1))$.
- ▶ $T(1, 0) = 2(1, 0) + (-3)(0, 1) = (2, -3)$
 $T(0, 1) = 1(1, 0) + (-2)(0, 1) = (1, -2)$
- ▶ $T(x, y) = T(x(1, 0) + y(0, 1))$
 $= xT(1, 0) + yT(0, 1)$
 $= x(2, -3) + y(1, -2)$
 $= (2x + y, -3x - 2y)$

Podprzestrzeń niezmienna

- ▶ Przekształcenie liniowe T z przestrzeni liniowej $\mathcal{V}_{|\mathbb{R}}$ do samej siebie nazywa się **operatorem liniowym**.
- ▶ Niech $\mathcal{V}_{|\mathbb{R}}$ będzie skończenie wymiarową przestrzenią liniową o $\dim(\mathcal{V}_{|\mathbb{R}}) \geq 1$ i $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{V}_{|\mathbb{R}}$ będzie operatorem liniowym. Wówczas podprzestrzeń $\mathcal{U}_{|\mathbb{R}}$ przestrzeni $\mathcal{V}_{|\mathbb{R}}$ nazywana jest **podprzestrzenią niezmienną** względem T , jeśli $T(u) \in \mathcal{U}_{|\mathbb{R}}$ dla każdego $u \in \mathcal{U}_{|\mathbb{R}}$.
- ▶ **Przykład:** Rozważmy liniowy operator $T : \mathbb{R}_{|\mathbb{R}}^3 \mapsto \mathbb{R}_{|\mathbb{R}}^3$ zdefiniowanego jako $T(x, y, z) = (x + 2y, x + y, 2z)$ i podprzestrzeń $\mathcal{U}_{|\mathbb{R}} \subseteq \mathbb{R}_{|\mathbb{R}}^3$ gdzie $\mathcal{U}_{|\mathbb{R}} = \text{Span}(\{(0, 0, 1)\}) = \{(0, 0, c) : c \in \mathbb{R}\}$.
Dla dowolnego $(0, 0, c) \in \mathcal{U}_{|\mathbb{R}}$, $T(0, 0, c) = (0, 0, 2c) \in \mathcal{U}_{|\mathbb{R}}$.

Wartości własne i wektory własne

- ▶ Niech $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{V}_{|\mathbb{R}}$ będzie liniowym operatorem. Skalar $\lambda \in \mathbb{R}$ jest nazywane **wartością własną** (ang. Eigen¹ value) przekształcenia liniowego T jeśli istnieje nie zerowy wektor $\vec{v} \in \mathcal{V}_{|\mathbb{R}}$ takie, że $T(\vec{v}) = \lambda \vec{v}$. Wektor $\vec{v}$ jest nazywane **wektorem własnym** (ang. Eigen vector) T odpowiadającym λ .
- ▶ Innymi słowy, niech M będzie macierzą reprezentacją przekształcenia liniowego $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{V}_{|\mathbb{R}}$. Więc, M ma wymiar $n \times n$ gdzie $\dim(\mathcal{V}_{|\mathbb{R}}) = n$. Skalar $\lambda \in \mathbb{R}$ jest **wartością własną** macierzy M jeśli istnieje nie zerowy wektor $\vec{v} \in \mathcal{V}_{|\mathbb{R}}$, takie że $M\vec{v} = \lambda \vec{v}$.
- ▶ W szczególności, jeśli $\mathcal{V}_{|\mathbb{R}} = \mathbb{R}_{|\mathbb{R}}^n$, wówczas $M \in \mathbb{R}^n \times \mathbb{R}^n$ i $\vec{v} \in \mathbb{R}_{|\mathbb{R}}^n$.

¹Eigen to niemieckie słowo oznaczające 'charakterystyczny', 'własny' lub 'własny'.

Własności wartości własnych i wektorów własnych

► Poniższe stwierdzenia są równoważne:

- λ jest wartością własną $T : \mathbb{R}_{|\mathbb{R}}^n \mapsto \mathbb{R}_{|\mathbb{R}}^n$ (lub $M \in \mathbb{R}^n \times \mathbb{R}^n$).
- Istnieje $\vec{v} \in \mathbb{R}^n \setminus \{\vec{0}\}$ takie, że $T(\vec{v}) = \lambda \vec{v}$ (lub równoważnie $\det(M - \lambda I_n) \vec{v} = 0$ ma rozwiązanie niezerowe).
- $\text{Rząd}(M - \lambda I_n) < n$.
- $\det(M - \lambda I_n) = 0$.

► **Przykład:** Rozważmy wcześniej omówiony przykład przekształcenia liniowego $T : \mathbb{R}_{|\mathbb{R}}^3 \mapsto \mathbb{R}_{|\mathbb{R}}^3$ zdefiniowanego jako $T(x, y, z) = (x + 2y, x + y, 2z)$.

- Biorąc pod uwagę bazę standardową, mamy macierz $M = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

- Niech λ będzie wartością własną T . Wtedy, $\det(M - \lambda I_n) = 0$.

- $\begin{vmatrix} 1-\lambda & 2 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{vmatrix} = 0 \iff \lambda = 2, 1+\sqrt{2}, 1-\sqrt{2}.$

Przestrzeń własna odpowiadająca wartości własnej

► $T(x, y, z) = (x + 2y, x + y, 2z) = (\lambda x, \lambda y, \lambda z)$

$$(1 - \lambda)x + 2y = 0 \quad (i)$$

$$x + (1 - \lambda)y = 0 \quad (ii)$$

$$(2 - \lambda)z = 0 \quad (iii)$$

► Wektor własny dla $\lambda = 2$:

- Mamy rozkład równań liniowych: $-x + 2y = 0$ i $x - y = 0$
- Jedynym możliwym rozwiązaniem jest $x = y = 0$ i jest ono całkowicie niezależne od wartości z . Ponieważ jednak wektory własne są wektorami niezerowymi, wektor własny odpowiadający $\lambda = 2$ ma postać $(0, 0, c)$ gdzie $c \in \mathbb{R} \setminus \{0\}$.
- Przestrzeń własna dla $\lambda = 2$: $V_{\lambda=2} = \{(0, 0, c) : c \in \mathbb{R} \setminus \{0\}\}$.

► Wektor własny dla $\lambda = 1 + \sqrt{2}$:

- Mamy $x - \sqrt{2}y = 0$ i $(1 - \sqrt{2})z = 0$. Więc,
 $V_{\lambda=1+\sqrt{2}} = \{(\sqrt{2}c, c, 0) : c \in \mathbb{R} \setminus \{0\}\}$.

► Wektor własny dla $\lambda = 1 - \sqrt{2}$:

- $V_{\lambda=1-\sqrt{2}} = \{(-\sqrt{2}c, c, 0) : c \in \mathbb{R} \setminus \{0\}\}$.

Związek między wartościami własnymi, wektorami własnymi i bazą

- ▶ **Twierdzenie:** Dla dowolnego operatora liniowego $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{V}_{|\mathbb{R}}$ i wartości własnej λ , V_λ jest niezmienną podprzestrzenią $\mathcal{V}_{|\mathbb{R}}$ względem T .
- ▶ **Twierdzenie:** Dowolny operator liniowy $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{V}_{|\mathbb{R}}$ ma co najwyżej $n = \dim(\mathcal{V}_{|\mathbb{R}})$ liczbę wartości własnych.
- ▶ **Twierdzenie:** Niech $T : \mathcal{V}_{|\mathbb{R}} \mapsto \mathcal{V}_{|\mathbb{R}}$ będzie liniowym operatorem i $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$ będą m różnymi własnymi wartościami T . Wówczas odpowiadające im wektory własne $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ są liniowo niezależne.
- ▶ **Przykład:** Już zobaczyliśmy $T : \mathbb{R}_{|\mathbb{R}}^3 \mapsto \mathbb{R}_{|\mathbb{R}}^3$ zdefiniowanego jako $T(x, y, z) = (x + 2y, x + y, 2z)$ ma $3 = \dim(\mathbb{R}_{|\mathbb{R}}^3)$ wartości własne. Więc, odpowiedni zbiór wektorów własnych, takich jak $\{(\sqrt{2}, 1, 0), (-\sqrt{2}, 1, 0), (0, 0, 1)\}$ jest liniowym niezależnym.
- ▶ **Twierdzenie:** Wyznacznik macierzy kwadratowej $A = [a_{ij}]_{n \times n}$ jest iloczynem jej wartości własnych (możliwie powtarzalnych).

PageRank Google: Strony internetowe jako wektory własne

- ▶ Pomysł na algorytm **PageRank**, opracowany na Uniwersytecie Stanforda przez Larry'ego Page'a (Larry Page) i Siergieja Brina (Sergey Brin) w 1996 roku, opierał się na założeniu, że znaczenie dowolnej strony internetowej można oszacować na podstawie znaczenia stron, które do niej linkują.
- ▶ Wszystkie witryny internetowe są prezentowane w postaci ogromnego, skierowanego grafu, który pokazuje, która strona linkuje do której.
- ▶ PageRank oblicza wagę (ważność) $w_i > 0$ witryny internetowej p_i poprzez zliczanie liczby stron wskazujących na p_i i biorąc pod uwagę wagę witryn internetowych, które linkują do p_i .
- ▶ Na podstawie grafu tworzona jest macierz przejść A , która szacuje prawdopodobieństwo (kliknięcia), że ktoś rozpocznie przeglądanie jednej strony internetowej i przejdzie na inną stronę.
- ▶ PageRank reprezentuje wektor, powiedzmy $\vec{x}$ rankingu stron internetowych taki, że $A\vec{x} = \vec{x}$ - co odpowiada wartości własnej 1.

Rozkład własny i diagonalizacja (ang. Eigen decomposition and diagonalization)

- ▶ Przydatność macierzy diagonalnej polega na tym, że pozwala na szybkie obliczenie wyznacznika, potęg i macierzy odwrotnej. Np.,
 - Dla macierzy diagonalnej D wyznacznik jest iloczynem jej elementów diagonalnych.
 - Potęgę macierzy D^k uzyskuje się po prostu podnosząc każdy element przekątnej do potęgi k .
 - Odwrotność D^{-1} jest odwrotnością elementów przekątnej (tzn., $\frac{1}{d_{ii}}$ dla każdego d_{ii} macierzy D), jeśli wszystkie są różne od zera.
- ▶ Zatem przekształcenie danej macierzy A (reprezentującej dane) w macierz diagonalną jest jednym z ważniejszych kroków w uczeniu maszynowym.
- ▶ **Definicja:** Macierz kwadratowa $A = [a_{ij}]_{n \times n}$ jest **diagonalizowalna** (ang. diagonalizable), jeśli istnieje macierz odwracalna P taka, że $D = P^{-1}AP$.

Rola wartości i wektorów własnych w tworzeniu macierzy diagonalizowanej

- ▶ Niech A będzie macierzą o wymiaru $n \times n$ i $\lambda_1, \lambda_2, \dots, \lambda_n$ będą wartościami własnymi macierzy A .

- ▶ Skonstruujemy $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

- ▶ Stworzymy $P = (\vec{p}_1 \ \vec{p}_2 \ \dots \ \vec{p}_n)$ gdzie $\vec{p}_i \in \mathbb{R}^n$ jest wektorem własnym odpowiadającym λ_i .

- ▶ Istnieje twierdzenie, które zapewnia, że $AP = PD$

- ▶ Warto uważać, że macierz $AP = (\vec{A\vec{p}_1} \ \vec{A\vec{p}_2} \ \dots \ \vec{A\vec{p}_n})$ gdzie dla dowolnego i , $1 \leq i \leq n$, i -ta kolumna AP jest uzyskiwana przez pomnożenie $\vec{A\vec{p}_i}$. Podobnie,

$$PD = (\lambda_1 \vec{p}_1 \ \lambda_2 \vec{p}_2 \ \dots \ \lambda_n \vec{p}_n)$$

- ▶ Zatem $AP = PD$ oznacza $\vec{A\vec{p}_i} = \lambda_i \vec{p}_i$ dla każdej i , $1 \leq i \leq n$.

Przykład

- ▶ Znajdźmy rozkład własny $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$
- ▶ Uzyskanie wartości i wektorów własnych: $\begin{vmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{vmatrix} = 0$
 - $\lambda_1 = 1$ i $\lambda_2 = 3$
 - $A\vec{p}_1 = \vec{p}_1$ daje nam $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} p_{11} \\ p_{12} \end{pmatrix} = \begin{pmatrix} p_{11} \\ p_{12} \end{pmatrix}$.
Więc, $\vec{p}_1 = \begin{pmatrix} p_{11} \\ p_{12} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
 - $A\vec{p}_2 = \vec{p}_2$ daje nam $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} p_{21} \\ p_{22} \end{pmatrix} = \begin{pmatrix} p_{21} \\ p_{22} \end{pmatrix}$.
Więc, $\vec{p}_2 = \begin{pmatrix} p_{21} \\ p_{22} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
- ▶ Tworzenie P : $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$
- ▶ Tworzenie $D = P^{-1}AP = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$

Przydatność macierzy diagonalizowalnych

- ▶ Macierze diagonalne D można efektywnie podnieść do potęgi. Dlatego możemy znaleźć potęgę macierzy dla A poprzez rozkład wartości własnych (jeśli istnieje), tak aby $A^k = (PDP^{-1})^k = PD^kP^{-1}$.
Obliczanie D^k jest efektywne, ponieważ stosujemy tę operację indywidualnie do dowolnego elementu przekątnej.
- ▶ Jeśli istnieje rozkład własny $A = PDP^{-1}$, wtedy
$$\begin{aligned} \det(A) &= \det(PDP^{-1}) = \det(P)\det(D)\det(P^{-1}) \\ &= \det(D) = \prod_{i=1}^n d_{ii} \end{aligned}$$

Rozkład wartości osobliwych (ang. Singular Value Decomposition)

- ▶ Rozkład macierzy na wartości osobliwe (SVD) jest centralną metodą rozkładu macierzy w algebrze liniowej. Nazywa się ją „fundamentalnym twierdzeniem algebry liniowej” (Strang, 1993), ponieważ można ją zastosować do wszystkich macierzy, nie tylko do macierzy kwadratowych, i istnieje ona zawsze.
- ▶ **Twierdzenie:** Niech $A = [a_{ij}]_{m \times n}$ będzie macierzą prostokątną rzędu $r \in [0, \min(m, n)]$. **SVD macierzy A jest rozkładem postaci $A = U.S.V^t$** gdzie U jest macierzą ortogonalną wymiaru $m \times m$ z wektorami kolumnowymi u_i ($1 \leq i \leq m$), V jest macierzą ortogonalną wymiaru $n \times n$ z wektorami kolumnowymi v_j ($1 \leq j \leq n$), i S jest macierzą wymiaru $m \times n$, gdzie $s_{ii} = \sigma_i \geq 0$ i $s_{ij} = 0$ dla $i \neq j$.

SVD

- ▶ Pozycje diagonalne σ_i , $i = 1, \dots, r$ S nazywane są wartościami osobliwymi (singular values), u_i nazywane są wektorami lewostronnymi osobliwymi (left-singular vector), a v_i nazywane są wektorami prawostronnymi osobliwymi (right-singular vectors). Zgodnie z konwencją wartości osobliwe są uporządkowane, tj. $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$.
- ▶ Macierz wartości osobliwych S jest unikalna, ale wymaga pewnej uwagi. S ma taki sam rozmiar jak A . Oznacza to, że S ma diagonalną podmacierz, która zawiera wartości osobliwe i wymaga dodatkowego uzupełnienia zerami.

S z podmacierzą diagonalną

► Jeśli $m > n$,

$$\begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \\ 0 & & 0 \\ \vdots & & \vdots \\ 0 & & 0 \end{pmatrix}$$

► Jeśli $m < n$,

$$\begin{pmatrix} \sigma_1 & & 0 & 0 & \dots & 0 \\ & \ddots & & \vdots & & \vdots \\ 0 & & \sigma_m & 0 & \dots & 0 \end{pmatrix}$$

- SVD istnieje dla dowolnej macierzy A o wymiaru $m \times n$.

Drugi filar uczenia maszynowego: prawdopodobieństwa

- ▶ Aby określić niepewność danych, niepewność modelu uczenia maszynowego i niepewność przewidywań wykonywanych przez zaprojektowany model, często stosuje się narzędzia rachunku prawdopodobieństwa.
- ▶ Treści
 - Prawdopodobieństwo i istotne pojęcia
 - Zmienna losowa
 - Rozkład prawdopodobieństwa

Pojęcia prawdopodobieństwa według Laplace'a

- ▶ **Eksperyment** to procedura, która przeprowadzana w określonych warunkach prowadzi do uzyskania możliwego zbioru wyników.
- ▶ **Przykład**: Rzucanie kostką jest przykładem eksperymentu.
- ▶ Zbiór wszystkich możliwych wyników eksperymentu nazywa się przestrzenią **próby** (ang. sample or sample space).
- ▶ **Przykład**: Przestrzeń próby rzutu kostką to $\{1, 2, 3, 4, 5, 6\}$.
- ▶ **Zdarzenie** (ang. event) jest identyfikowane z podzbiorem przestrzeni próby eksperymentu. **Przestrzeń zdarzeń** jest przestrzenią potencjalnych wyników eksperymentu.
- ▶ **Przykład**: Uzyskanie nieparzystej liczby w eksperymencie z rzucaniem kostką odnosi się do zdarzenia $\{1, 3, 5\}$.
- ▶ Wyniki przestrzeni próby nazywa się **równie prawdopodobnymi**, jeśli w trakcie przeprowadzania eksperymentu, w każdym przypadku, wynikiem może być dowolny z elementów przestrzeni próby; to znaczy, że nie ma preferencji jednego wyniku nad innymi.

Prawdopodobieństwo zdarzenia na próbie skończonej

- ▶ **Prawdopodobieństwo zdarzenia** E , które jest podzbiorem skończonej przestrzeni prób S o jednakowo prawdopodobnych wynikach, wynosi $P(E) = \frac{|E|}{|S|}$.
- ▶ **Przykłady:**
 - Prawdopodobieństwo uzyskania nieparzystej liczby w rzucie kostką wynosi $\frac{3}{6} = \frac{1}{2}$.
 - Prawdopodobieństwo uzyskania siódemki w rzucie kostką wynosi 0.
 - Prawdopodobieństwo uzyskania parzystej lub nieparzystej oczki w rzucie kostką wynosi $\frac{6}{6} = 1$.
- ▶ Zdarzenie E o prawdopodobieństwie 0 nazywa się **zdarzeniem niemożliwym**. Zdarzenie polegające na uzyskaniu siódemki w rzucie kostką jest zdarzeniem niemożliwym.
- ▶ Zdarzenie E z prawdopodobieństwem 1 nazywa się **zdarzeniem pewnym**. Zdarzenie polegające na uzyskaniu parzystej lub nieparzystej oczki w rzucie kostką jest zdarzeniem pewnym.
- ▶ Dla dowolnego zdarzenia E przestrzeni prób S , $0 \leq P(E) \leq 1$.

Przykłady

1. Urna zawiera cztery niebieskie i pięć czerwonych kul. Jakie jest prawdopodobieństwo, że wybrana z urny kula będzie niebieska?
 - Istnieje 9 możliwych wyników, jeśli wylosujemy jedną kulę z urny. Spośród nich 4 są korzystne dla danego zdarzenia. Zatem prawdopodobieństwo wylosowania niebieskiej kuli wynosi $\frac{4}{9}$.
2. Jakie jest prawdopodobieństwo, że po rzucie dwiema kostkami suma oczek na obu kostkach wyniesie 7?
 - Jeśli rzucimy dwiema kostkami, liczba wyników w przestrzeni próby wyniesie $6^2 = 36$. Wśród nich wyniki korzystne dla zdarzenia „uzyskania 7 jako sumy oczek dwóch kostek” to (1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3). Zatem zdarzenie ma 6 możliwych przypadków. Stąd prawdopodobieństwo zdarzenia wynosi $\frac{6}{36} = \frac{1}{6}$.

Wybieranie rzeczy z możliwością zastąpienia i bez zastąpienia

- ▶ **Przykład:** Jakie jest prawdopodobieństwo, że liczby 11, 4, 17, 39, 23 zostaną wylosowane w tej kolejności z pojemnika zawierającego 50 kul oznaczonych liczbami od 1 do 50, jeśli (a) kula wybrana w każdej turze nie jest zwracana do pojemnika i (b) kula wybrana w każdej turze jest zwracana do pojemnika?
- (a) **Wybieranie rzeczy bez zastąpienia:** Ponieważ za każdym razem, gdy kula jest wybierana, w koszu jest o jedną kulę mniej, zgodnie z regułą iloczynu (liczenia) istnieje $50, 49, 48, 47, 46 = 254251200$ sposobów na wybranie pięciu kul z kosza. Spośród nich jest tylko jeden przypadek, który pasuje do wyboru 11, 4, 17, 39, 23. Zatem prawdopodobieństwo wylosowania tych pięciu liczb w powyższej kolejności wynosi $\frac{1}{254251200}$.

Wybieranie rzeczy z możliwością zastąpienia

- ▶ Ponieważ za każdym razem, gdy kulę po wybraniu wraca do kosza, zgodnie z regułą iloczynu istnieje $50^5 = 312500000$ możliwych sposobów wybrania 5 kul. Wśród nich jest tylko jeden przypadek, który pasuje do wyboru 11, 4, 17, 39, 23. Stąd prawdopodobieństwo wylosowania tych pięciu liczb w wyżej wymienionej kolejności wynosi $\frac{1}{312500000}$.

Prawdopodobieństwo kombinacji zdarzeń

- ▶ **Twierdzenie:** Niech E będzie zdarzeniem związanym z przestrzenią prób S . Prawdopodobieństwo zdarzenia E , będącego dopełnieniem E , jest dane wzorem: $P(\bar{E}) = 1 - P(E)$.
- ▶ **Przykład:** Sekwencja dziesięciu bitów jest generowana losowo. Jakie jest prawdopodobieństwo, że przynajmniej jeden z bitów będzie miał wartość 0?
 - Niech E = Bity o długości 10 posiadające co najmniej jedno 0. $\bar{E}$ = Bity o długości dziesięć, niezawierające żadnego 0. Tak więc istnieje tylko jeden możliwy przypadek na korzyść $\bar{E}$. Przestrzeń próby zawiera ciągi $2^{10} = 1024$ bitów. Więc $P(\bar{E}) = \frac{1}{1024}$. Tzn., $P(E) = 1 - \frac{1}{1024} = \frac{1023}{1024}$.
- ▶ **Twierdzenie:** Niech E_1, E_2 będą zdarzeniami przestrzeni prób S . Wtedy $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.
- ▶ **Przykład:** Jakie jest prawdopodobieństwo, że liczba całkowita dodatnia wybrana losowo z zestawu liczb całkowitych dodatnich nieprzekraczających 100 jest podzielna przez 2 lub 5?

Ograniczenia pojęcia prawdopodobieństwa Laplace'a

- ▶ Zakłada dwa warunki: (i) liczba wyników musi być skończona oraz (ii) każdy wynik jest równie prawdopodobny.
 - ▶ Jednak wiele eksperymentów może nie mieć równie prawdopodobnych wyników i/lub skończenie wielu wyników.
- (1) Przestrzeń próby odpowiadająca liczbie rzutów monetą do pierwszego orła to $\{1, 2, 3, \dots\}$
 - (2) Moneta może być tak obciążona, że orzeł wypada dwa razy częściej niż reszka.
 - (3) Rozważmy, że rzucamy dwiema monetami i wygrywamy, jeśli wypadnie przynajmniej jeden orzeł, w przeciwnym razie przegrywamy. Jakie jest prawdopodobieństwo wygranej?
 - **Jedno rozwiązanie:** Jeśli E jest zdarzeniem oznaczającym wygraną w grze, wtedy $E = \{HT, HH, TH\}$ gdzie przestrzeń próbek $S = \{HH, HT, TH, TT\}$. Więc, $P(E) = \frac{3}{4}$.
 - **Drugie rozwiązanie według D'Alemberta:** Jeśli pierwsza moneta wypadnie orłem, drugi rzut nie będzie miał miejsca, ponieważ gra jest już wygrana. Więc, $E = \{H, TH\}$, $S = \{H, TH, TT\}$ i stąd $P(E) = \frac{2}{3}$. (Istnieją kontrowersje dotyczące drugiego rozwiązania.)

Wyzwania uogólniania koncepcji prawdopodobieństwa

- ▶ W przypadku nieskończonej przestrzeni prób pojęcie kardynalności zbioru punktów sprzyjających danemu zdarzeniu E zmienia się na pojęcie **miary** (ang. **measure**).

Przykład miary dla zbioru nieskończonego: Praktycznym przykładem miary zbioru jest długość odcinka, gdzie zbiór jest podzbiorem osi liczbowej. Na przykład miarą zbioru punktów z przedziału $[a, b]$ jest jego długość, która wynosi $b - a$.

- ▶ Jeśli przestrzeń próby S jest nieprzeliczalnie nieskończona, to nie wszystkie podzbiory S (tj. wszystkie zdarzenia) są mierzalne (zgodnie z pojęciem σ -algebry).
- ▶ **Czym jest σ -algebra:** Składa się z mierzalnych zbiorów, dla których 'miarę' (sposób kwantyfikacji jego wielkości, taki jak długość, powierzchnia lub objętość) można zdefiniować w spójny sposób. Zbiory te są fundamentalne w teorii miary i są zamknięte na operacje takie jak dopełnienie i suma skończona, co oznacza, że jeśli zbiór jest mierzalny, jego dopełnienie również jest mierzalne, a suma skończonej liczby zbiorów mierzalnych również jest mierzalna.

Uogólnione pojęcie prawdopodobieństwa

- ▶ Kolmogorov (1933) wprowadził aksjomatyczną definicję miary prawdopodobieństwa dla σ -algebry $F \subseteq P(S)$ dla danej przestrzeni próbek S .
- ▶ Niech S będzie przestrzenią prób eksperymentu i $F \subseteq P(S)$ jest σ -algebrą. Prawdopodobieństwo jest funkcją $P : F \mapsto [0, 1]$ taką, że (i) $0 \leq P(E)$ dla dowolnego $E \in F$, (ii) $P(S) = 1$, i (iii) Dla dowolnego ciągu wzajemnie rozłącznych zdarzeń $E_1, E_2, \dots \in F$, $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$.
- ▶ Z powyższych warunków można udowodnić (iv) $P(\phi) = 0$ gdzie ϕ oznacza niemożliwe zdarzenie i (v) $P(E) \leq 1$ dla dowolnego $E \in F$.
- ▶ Dla przeliczalnej przestrzeni prób S , $P(S)$ to σ -algebra.
- ▶ Definicję prawdopodobieństwa Laplace'a można uzyskać jako szczególny przypadek tej definicji.

Kombinacje zdarzeń

- ▶ **Przykład:** Jakie prawdopodobieństwa powinniśmy przypisać wynikom H (orzeł) i T (reszka) przy rzucie uczciwą monetą? Jakie prawdopodobieństwa powinniśmy przypisać tym zdarzeniom, gdy moneta jest obciążona tak, że orzeł wypada dwa razy częściej niż reszka?
 - W przypadku uczciwej monety prawdopodobieństwo H i T jest jednakowe. To znaczy $P(H) = P(T) = \frac{1}{2}$.
 - Dla stronniczej monety $P(H) = 2P(T)$.
Więc, $P(H) + P(T) = 1$ daje nam $3P(T) = 1$.
Stąd, $P(T) = \frac{1}{3}$ i $P(H) = \frac{2}{3}$.
- ▶ Każdy wynik $s \in S$ otrzymuje zatem wartość $P(\{s\}) \in [0, 1]$, i możemy to oznaczyć jako $P(s)$.
- ▶ **Twierdzenie:** Dla dowolnych dwóch zdarzeń E_1, E_2 ,
 - (i) jeśli $E_1 \subseteq E_2$, wtedy $P(E_2 \setminus E_1) = P(E_2) - P(E_1)$
 - (ii) $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
 - (iii) $P(E_1 \cup E_2) \leq P(E_1) + P(E_2)$.

Prawdopodobieństwo warunkowe

- ▶ **Przykład:** Załóżmy, że rzucamy trzy razy uczciwą monetą i wiemy, że zdarzenie F oznacza, że pierwszy rzut wypadnie reszką. Biorąc pod uwagę te informacje, jakie jest prawdopodobieństwo zdarzenia E , że wypadnie nieparzysta liczba reszek?
- ▶ **Rozwiązanie:** Widzimy, że $F = \{TTT, TTH, THT, THH\}$ i $E = \{THH, HTH, HHT, TTT\}$. Skoro już określono, że zdarzenie F już wystąpiło, musimy poszukać zdarzenia F i E , tzn. $F \cap E = \{THH, TTT\}$. Ponieważ istnieją 4 możliwe sposoby wystąpienia zdarzenia F , a wśród nich 2 są takie, że zdarzenie E występuje razem z zdarzeniem F ,
prawdopodobieństwo wystąpienia zdarzenia E , zakładając, że zdarzenie F już wystąpiło, wynosi $\frac{2}{4} = \frac{1}{2}$.
- ▶ **Definicja:** Niech E i F będą dwoma zdarzeniami takimi, że $P(F) > 0$. Prawdopodobieństwo warunkowe E przy danym F , oznaczone jako $P(E|F) = \frac{P(E \cap F)}{P(F)}$.

Niezależne wydarzenia

- ▶ Dwa zdarzenia E i F w tej samej przestrzeni próbkowania S są nazywane niezależnymi, jeśli $P(E \cap F) = P(E).P(F)$
- ▶ **Warto uważać:** Zgodnie z tożsamością wynikającą z definicji prawdopodobieństwa warunkowego, jeśli wiemy, że zdarzenie F już wystąpiło, to $P(E \cap F) = P(E|F).P(F)$. Łącząc to teraz z definicją zdarzenia niezależnego, widzimy, że jeśli wystąpienie zdarzenia F nie ma żadnego wpływu na wystąpienie zdarzenia E , to $P(E \cap F) = P(E).P(F)$ - co oznacza $P(E|F) = P(E)$.
- ▶ **Przykład:** Niech E będzie zdarzeniem polegającym na tym, że losowo wygenerowany ciąg bitów o długości czterech zaczyna się od 1, a F zdarzeniem polegającym na tym, że ciąg bitów zawiera parzystą liczbę jedynek. Czy E i F są niezależne?
 - $E = \{1000, 1001, 1010, 1100, 1011, 1101, 1110, 1111\}$ i
 $F = \{0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111\}$.
Ponieważ w przestrzeni próbek znajduje się 16 ciągów bitowych $P(E) = P(F) = \frac{8}{16} = \frac{1}{2}$. $P(E \cap F) = \frac{4}{16} = \frac{1}{4} = P(E).P(F)$. Zatem E i F są zdarzeniami niezależnymi.

Twierdzenie Bayesa

- ▶ **Prawdopodobieństwo sumy skończonej liczby zdarzeń:**

$$P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) - \sum_{i \neq j} P(E_i \cap E_j) + \sum_{i \neq j \neq k} P(E_i \cap E_j \cap E_k) + \dots + (-1)^{n+1} P(E_1 \cap E_2 \cap \dots \cap E_n).$$

- ▶ Jeśli zdarzenia $E_1, E_2, \dots, E_n$ wykluczają się wzajemnie (tj. $E_i \cap E_j = \emptyset$ dla $i \neq j$), wtedy $P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i)$.
- ▶ Zdarzenia $E_1, E_2, \dots, E_n$ w przestrzeni próbek S nazywamy wyczerpującym, jeśli za każdym razem, gdy przeprowadzany jest eksperyment, musi nastąpić jedno z tych zdarzeń.
- ▶ Niech $E_1, E_2, \dots, E_n$ będą zbiorem wzajemnie wykluczających się i wyczerpujących zdarzeń w przestrzeni prób S , a B będzie dowolnym zdarzeniem w przestrzeni S . Wtedy $P(B) = \sum_{i=1}^n P(E_i \cap B) = \sum_{i=1}^n P(B|E_i)P(E_i)$. Nazywa się to **całkowitym prawdopodobieństwem** zdarzenia B .
- ▶ **Twierdzenie Bayesa:** Niech $E_1, E_2, \dots, E_n$ będą zbiorem wzajemnie wykluczających się i wyczerpujących zdarzeń w przestrzeni prób S , a B będzie dowolnym zdarzeniem w przestrzeni S . Wtedy, $P(E_i|B) = \frac{P(B|E_i)P(E_i)}{\sum_{i=1}^n P(B|E_i) \cdot P(E_i)}$.

Przykład

- ▶ Załóżmy, że test laboratoryjny mający na celu wykrycie pewnej choroby ma następujące statystyki. Niech A oznacza zdarzenie, że badana osoba jest chora, a B oznacza zdarzenie, że wynik testu jest pozytywny. Wiadomo, że $P(B|A) = 0.99$, $P(B|\bar{A}) = 0.005$, i 0,1 procenta populacji faktycznie cierpi na tę chorobę. Jakie jest prawdopodobieństwo, że dana osoba jest chora, jeśli wynik testu będzie pozytywny?

Rozwiązanie: $P(A) = 0.001$ i $P(\bar{A}) = 0.999$.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A).P(A)+P(B|\bar{A}).P(\bar{A})} = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.005 \times 0.999} = 0.165$$

Zmienna losowa

- ▶ Istnieje wiele problemów, w których musimy radzić sobie z liczbą wystąpień określonego wyniku eksperymentu. Na przykład, możemy potrzebować znać prawdopodobieństwo, że moneta wypadnie z 11 reszkami po rzucie 20 razy. W badaniu tego rodzaju problemów pomocne staje się użycie zmiennej losowej.
- ▶ **Przykład:** Załóżmy, że moneta jest rzucana trzy razy. Niech $X(s)$ reprezentuje liczbę orłów dla każdego $s \in S$, gdzie S jest przestrzenią prób dla eksperymentu polegającego na trzykrotnym rzucie monetą. Wtedy, $X(HHH) = 3$, $X(HHT) = X(HTH) = X(THH) = 2$, $X(TTH) = X(THT) = X(HTT) = 1$, i $X(TTT) = 0$.
- ▶ **Definicja:** Zmienna losowa to funkcja z przestrzeni prób eksperymentu do zbioru liczb rzeczywistych. Oznacza to, że zmienna losowa przypisuje liczbę rzeczywistą każdemu możliwemu wynikowi eksperymentu.

Rozkład prawdopodobieństwa zmiennej losowej

- ▶ **Przykład:** Nawiązując do poprzedniego przykładu, widzimy $P(X = 3) = \frac{1}{8}$, $P(X = 2) = \frac{3}{8} = P(X = 1)$, $P(X = 0) = \frac{1}{8}$, i $P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0) = 1$
- ▶ **Definicja:** Rozkład prawdopodobieństwa zmiennej losowej X w przestrzeni próby S to specyfikacja prawdopodobieństw $P(X = r)$ dla wszystkich możliwych wartości r takich, że $X(s) = r$ dla $s \in S$.
- ▶ **Przykład funkcji rozkładu prawdopodobieństwa:**

X	$P(X)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

- ▶ **Właściwości funkcji rozkładu prawdopodobieństwa:**
Dla zmiennej losowej X w przestrzeni próby S ,
(i) $0 \leq P(X = r) \leq 1$ dla dowolnej r takiej, że $X(s) = r$ dla $s \in S$, i (ii) $\sum_{X=r} P(X = r) = 1$.

Funkcja dystrybucji skumulowanej

- **Definicja:** Funkcja dystrybucji skumulowanej (ang. cumulative distribution function) (cdf) dla zmiennej losowej X jest zdefiniowany jako $F_X(x) = P(X \leq x)$, gdzie $-\infty < x < \infty$

- **Przykład:** $F_X(0) = P(X \leq 0) = \frac{1}{8}$

$$F_X(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}.$$

$$F_X(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\ = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}.$$

$$F_X(3) = P(X \leq 2) =$$

$$P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1.$$

$$F_X(x) = \begin{cases} 0 & \text{if } -\infty < x < 0 \\ \frac{1}{8} & \text{if } 0 \leq x < 1 \\ \frac{1}{2} & \text{if } 1 \leq x < 2 \\ \frac{7}{8} & \text{if } 2 \leq x < 3 \\ 1 & \text{if } 3 \leq x < \infty \end{cases}$$

Dyskretna zmienna losowa i rozkład prawdopodobieństwa

► Właściwości F_X :

- (i) $0 \leq F_X(x) \leq 1$
- (ii) Jeśli $x_1 \leq x_2$, $F_X(x_1) \leq F_X(x_2)$
- (iii) $\lim_{x \rightarrow \infty} F_X(x) = 1$
- (iv) $\lim_{x \rightarrow -\infty} F_X(x) = 0$

► Znajdowanie prawdopodobieństw z F_X :

- $P(a < X \leq b) = F_X(b) - F_X(a)$
- $P(X > a) = 1 - P(X \leq a) = 1 - F_X(a)$
- $P(X < b) = \lim_{x \rightarrow b^-} F_X(x)$

► Zmienna losowa jest dyskretna, jeśli przyjmuje wartości z dyskretnego zbioru. Dla zmiennej losowej dyskretnej $F_X(x)$ reprezentuje funkcję schodkową.

Funkcja masy prawdopodobieństwa

- ▶ Dla dyskretnej zmiennej losowej prawdopodobieństwo w dowolnym przedziale oblicza się zasadniczo na podstawie prawdopodobieństw w pewnych dyskretnych punktach leżących w tym przedziale. Zatem dla dyskretnej zmiennej losowej X prawdopodobieństwo w dowolnym punkcie x nazywa się **funkcją masy prawdopodobieństwa** (ang. **probability mass function** lub **pmf**) X w x .
- ▶ $P(X = x) = p_X(x)$ gdzie p reprezentuje pmf zmiennej X .
- ▶ Dla dyskretnej zmiennej przyjmującej wartości $x_1, x_2, \dots$,
 $p_X(x_i) = P(X = x_i) = F_X(x_i) - F_X(x_{i-1})$
- ▶ **Właściwości pmf:**
 - (i) $0 \leq p_X(x_i) \leq 1$ dla $i = 1, 2, \dots$
 - (ii) $p_X(x) = 0$ jeśli $x \neq x_i$ dla dowolnej $i = 1, 2, \dots$
 - (iii) $\sum_i p_X(x_i) = 1$
 - (iv) $F_X(x_i) = \sum_{j=1}^i p_X(x_j)$.

Ciągła zmienna losowa i rozkład prawdopodobieństwa

- ▶ Ciągła zmienna losowa X przyjmuje wartości ze zbioru ciągłego.
- ▶ W przypadku zbioru ciągłego, takiego jak przedział $[a, b]$, zmienna losowa X ma oczywiście nieprzeliczalnie wiele wartości do przyjęcia, a zatem rozkład prawdopodobieństwa X jest rozumiany jako skumulowany rozkład prawdopodobieństwa w całym przedziale. Zatem w przypadku zbioru ciągłego rozkład prawdopodobieństwa jest znany jako **funkcja gęstości prawdopodobieństwa (ang. probability density function lub pdf)**.
- ▶ Funkcja gęstości prawdopodobieństwa dla zmiennej losowej X w zakresie wartości x to $f_X(x) = \frac{d}{dx} F_X(x)$
- ▶ Funkcja gęstości prawdopodobieństwa X w $[a, b]$: $\int_a^b f_X(x) dx$
- ▶ **Właściwości pdf**: (i) $f_X(x) \geq 0$ (ii) $\int_{-\infty}^{\infty} f_X(x) dx = 1$
(iii) $P(a \leq X \leq b) = \int_a^b f_X(x) dx$
(iv) $P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = F_X(b) - F_X(a)$.

Niektóre rozkłady dyskretne

1. Załóżmy, że X jest zmienną losową nad S . Dyskretny rozkład prawdopodobieństwa nazywa się **rozkładem jednostajnym** (ang. uniform distribution), jeśli $P(X = r)$ jest takie samo dla każdej wartości r , tak że $X(s) = r$ dla $s \in S$.
 - ▶ **Przykład**: Rozważmy eksperyment z rzutem kostką. Zatem $S = \{1, 2, 3, 4, 5, 6\}$. Niech X będzie zmienną losową na S taką, że $X(s) = s$. Wówczas $P(X = s) = \frac{1}{6}$ dla każdego $s \in S$.
 - ▶ Jeśli eksperyment jest taki, że każda próba może mieć tylko dwa możliwe wyniki, a prawdopodobieństwo każdego wyniku pozostaje takie samo we wszystkich próbach, każda próba takiego eksperymentu nazywana jest **próbą Bernoulliego**. Dwa możliwe wyniki próby Bernoulliego są zazwyczaj nazywane sukcesem i porażką.
2. **Rozkład dwumianowy**: Prawdopodobieństwo dokładnie k sukcesów w n niezależnych próbach Bernoulliego z prawdopodobieństwem sukcesu p i prawdopodobieństwem porażki $q = 1 - p$, wynosi $C(n, k)p^k q^{n-k}$.

Przykład

- ▶ Jakie jest prawdopodobieństwo uzyskania dwóch orłów w 5 rzutach monetą?

Rozwiązanie: Po próbach Bernoulliego wykonano tutaj 5 prób rzutu monetą i w każdej próbie prawdopodobieństwo sukcesu (czyli wyrzucenia orła) wynosi $\frac{1}{2}$. Więc, prawdopodobieństwo uzyskania 2 orłów 5 rzutach wynosi $C(5, 2)\left(\frac{1}{2}\right)^5 = \frac{10}{32}$.

- ▶ **Zmienna losowa Bernoulliego:** Zmienna losowa X ma dwie wartości: 'sukces' i 'porażkę', które można odpowiednio uznać za 1 i 0. Niech $P(X = 1) = p$ i $P(X = 0) = 1 - p$. **pmf jest zdefiniowany jako:** $p(X = i) = p^i(1 - p)^{1-i}$ gdzie $i = 0, 1$.

-

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Rozkład Poissona

- ▶ To rozkład dwumianowy o dużej liczbie prób. Oznacza to, że $n \rightarrow \infty$ i w rezultacie p prawdopodobieństwo sukcesu staje się bardzo małe ($p \rightarrow 0$), a relacja między n i p jest taka, że np ($= \lambda$) jest stała.
- ▶ **Przykład:** Liczba połączeń telefonicznych przychodzących w ustalonym odstępie czasu. Niech λ = długość odstępu czasu, n liczba przychodzących połączeń, a prawdopodobieństwo konkretnego połączenia będzie równe p . Wtedy, $p = \frac{\lambda}{n}$. Teraz, jeśli n wzrasta, p maleje, ponieważ λ jest stała.
- ▶ X jest zmienną losową Poissona z parametrem λ (> 0), jeśli jej pmf $p_X(i) = \frac{e^{-\lambda} \lambda^i}{i!}$, $i = 0, 1, 2, \dots$
- ▶ **Warto wiedzieć:** Prawdopodobieństwo otrzymania i liczby połączeń to $P(X = i) = \frac{n!}{i!(n-i)!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$. Gdy $n \rightarrow \infty$ $P(X = i) \rightarrow \frac{e^{-\lambda} \lambda^i}{i!}$.

Dziękuję za uwagę