1 Analiza Matematyczna

1.1 Pochodna funkcji złożonej

$$h = f \circ g \quad h'(x) = f'(g(x)) \cdot g'(x)$$

Przykład 1.1

Pochodną funkcji $f(x) = (x^9 + x)^9$ jest:

$$f'(x) = 9(x^9 + x)^8 \cdot (9x^8 + 1)$$

1.2 Jakobian

$$f : \mathbb{R}^n \to \mathbb{R}^m \quad f(x) = (f_1(x), \ldots, f_m(x))$$

Dla powyższej funkcji, macierz Jakobiego to:

$$J_{ij} = \frac{\partial f_i}{\partial x_j} \quad J_f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Jakobian, to wyznacznik macierzy Jakobiego.

Przykład 1.2

$$f(x, y, z) = (x, y, z) \quad f_1(x, y, z) = x \quad f_2(x, y, z) = y \quad f_3(x, y, z) = z$$

$$J_f = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\det(J_f) = 1$$

2 Rachunek Prawdopodobieństwa

Gęstością pewnego zdarzenia nazywamy funkcję $f : \Omega \to [0, 1]$, która dla każdego $x \in \Omega$ zwraca prawdopodobieństwo zdarzenia x. Wymogiem formalnym gęstości, jest to, że się sumuje do 1:

$$\sum_{x \in \Omega} f(x) = 1 \quad \int_\Omega f(x)dx = 1$$

Przykład 2.1

Na przykład, dla zmiennej losowej X, określonej przez wynik losowego rzutu kością, prawdopodobieństwo zdarzenia $X = k$ wynosi $\frac{1}{6}$, gdzie k to wynik rzutu. Stąd też, $f(x) = \frac{1}{6}$ dla $x = k$.

Dystrybuanta, to funkcja $F : \Omega \to [0, 1]$, ściśle związana z gęstością:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

W pewnym sensie, reprezentuje prawdopodobieństwo, że zmienna losowa X ma wartość mniejszą lub równą x.

Przykład 2.2

Wracając do naszego przykładu z kością, dystrybuanta $F(x)$ dla $x = k$ wynosi

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{k} \frac{1}{6}dt = \frac{k}{6}$$

W tym przypadku, $F(x)$ jest funkcją skokową, która przyjmuje wartość 0 dla $x < 1$, $1/6$ dla $x = 1$, $2/6$ dla $x = 2$, itd.

Dystrybuanta nie musi się sumować do 1, wręcz przeciwnie, ale musi być niemalejąca, prawostronnie ciągła (lub lewostronnie, ale nie może mieć dziur).

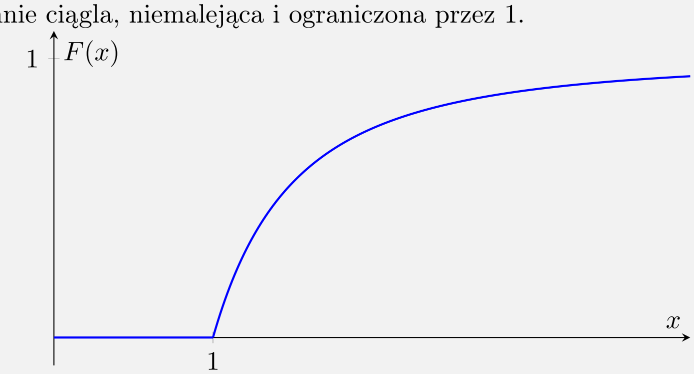
$$\lim_{x\to\infty} F(x) = 1 \qquad \lim_{x\to-\infty} F(x) = 0$$

Przykład 2.3

Poniższa funkcja $F(x)$ jest dystrybuantą:

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 - \frac{1}{x^2} & x \geq 1 \end{cases}$$

Ponieważ jest prawostronnie ciągla, niemalejąca i ograniczona przez 1.



3 Algebra Liniowa

Wektor to element przestrzeni wektorowej V, czyli $x \in V$. Wektory są ortogonalne, jeśli ich iloczyn skalarny jest równy 0, czyli $x \cdot y = 0$ dla dowolnych $x, y \in V$.

Przykład 3.1

Wektory $(1, 1, 1)$ i $(-2, 1, 1)$ są ortogonalne, ponieważ ich iloczyn wynosi:

$$\begin{bmatrix}1 & 1 & 1\end{bmatrix} \begin{bmatrix}-2 \\ 1 \\ 1\end{bmatrix} = 0$$

Normą indukowaną w $\mathbb{R}^n$ jest funkcja $||x|| = \sqrt{x^T \cdot x} = \sqrt{\langle x|x \rangle}$.

3.1 Wyznacznik

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{n-1} a_{nn} \det A_{nn}$$

gdzie A_{ij} to macierz A z usuniętym wierszem i i kolumną j.

Przykład 3.2

Dla macierzy A rozmiaru 3×3:

$$\det A = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Dla macierzy A rozmiaru 4×4:

$$\det A = a_{11} \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} - a_{14} \begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix}$$

3.2 Odwzorowania liniowe

Odwzorowanie liniowe $T : V \to U$, to macierzowa transformacja

$$T(x) = Ax$$

gdzie A to macierz $\dim V \times \dim U$, a x to wektor $\dim V$-elementowy.

$$\ker T = \{x : T(x) = 0\} \quad \operatorname{Im} T = \{T(x) : x \in V\}$$

Zbiór wektorów jest liniowo niezależny, gdy nie istnieje żadna kombinacja liniowa tych wektorów, która daje wynik równy 0.

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$$

Ranga macierzy A to maksymalna liczba liniowo niezależnych wierszy (lub kolumn) macierzy A, czyli $rank(A) = \dim \operatorname{Im} T$. Formalnie; wymiar przestrzeni rozpiętej przez jej wiersze. Jeśli ranga jest równa zero, to macierz wysyła wektory do zera, czyli $\ker T = V$.

Operator liniowy jest różnowartościowy wtedy i tylko wtedy, gdy $\ker T = \{0\}$. Macierz przekształcenia musi wtedy też mieć rangę n, gdzie n to liczba wierszy macierzy A.

$$\dim \mathbb{R}^n = n \quad \dim V = \dim \operatorname{Im} T + \dim \ker T$$

Przykład 3.3

Jeśli wiemy, że $T : \mathbb{R}^4 \to \mathbb{R}^3$, to wiemy, że macierz T ma rozmiar 3×4.

$$T(x) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 \end{bmatrix}$$

T, może mieć nietrywialne $\ker T \neq \{0\}$, może mieć rangę w $[0, 4]$ bo ma 4 kolumny, lecz nie może być różnowartościowe, ponieważ:

$$\dim V = \dim \operatorname{Im} T + \dim \ker T \Rightarrow 4 = 3 + \dim \ker T \Rightarrow \dim \ker T = 1$$

Więc $\ker T \neq \{0\}$, więc T nie jest różnowartościowy.

Aby macierz była odwracalna, to musi być kwadratowa, czyli $n \times n$, gdzie n to liczba wierszy, jej ranga musi być maksylana $\dim \operatorname{Im} T = n$, więc $\dim \ker T = 0$, czyli $\ker T = \{0\}$, i jej wyznacznik musi być różny od zera, czyli $\det T \neq 0$.

3.3 Wartości własne

Dla równania:

$$A \left|\psi\right\rangle = \lambda \left|\psi\right\rangle$$

$|\psi\rangle$ jest wektorem własnym macierzy A, a λ jest jej wartością własną - odpowiadającą wektorowi własnemu $|\psi\rangle$.

$$\det(A - \lambda I) = 0 \quad (A - \lambda I)\,|\psi\rangle = 0$$

Wielomian charakterystyczny macierzy to:

$$p_A(t) = (t - \lambda_1)^{k_1}(t - \lambda_2)^{k_2} \cdots (t - \lambda_n)^{k_n} = 0 \quad p_A(A) = 0$$

jest to zazwyczaj pochodna poszukiwania wartości i wektorów własnych.

3.4 Rozkłady

Każdą rzeczywistą, symetryczną macierz dodatnio określonej formy kwadratowej można zapisać jako iloczyn macierzy dolnotrójkątnej i jej transpozycji. Nazywamy to rozkładem Choleskiego:

$$A = LL^T$$

4 Statystyka

Estymator, to funkcja $\hat{\theta}(X)$, która na podstawie próbki $X \subset U$, gdzie U to cała populacja, estymuje wartość parametru $\theta(U)$.

Estymator jest nieobiążony, jeśli:

$$\mathbb{E}[\hat{\theta}(X)] = \theta(U)$$

czyli wartość oczekiwana estymatora jest równa wartości parametru. Mówimy wtedy, że $\hat{\theta}(X)$ jest estymatorem największej wiarygodności (ENW).

Przykład 4.1

Dla próbki $X \subset U$, gdzie U to cała populacja, estymator średniej $\hat{\mu}(X) = \overline{X}$, jest nieobciążonym estymatorem średniej $\mu(U)$, ponieważ:

$$\mathbb{E}[\hat{\mu}(X)] = \mu(U)$$

Intuicyjnie, estymator jest nieobciążony, ponieważ dla odpowiednio dużej próbki X, średnia z próbki będzie bliska wartości parametru $\mu(U)$.

4.1 Charakterystyka próbek losowych

Dla zmiennej losowej X,

- $\mathbb{E}[X] = \mu$ - wartość oczekiwana to średnia
- $\mathrm{Var}[X] = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Cov}[X, X] = \sigma^2$
- $\sigma = \sqrt{\mathrm{Var}[X]}$ - odchylenie standardowe

4.2 Rozkład Normalny

Rozkład normalny, lub rozkład Gaussa, jest rozkładem prawdopodobieństwa zdefiniowanym na rzeczywistych liczbach rzeczywistych. Dla zmiennej losowej X o rozkładzie normalnym:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

gdzie μ to średnia, a σ^2 to wariancja.

Typowy estymator wariancji z próby:

$$\hat{\sigma}^2(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \mathbb{E}[\hat{\sigma}^2(X)] = \frac{n}{n-1}\mathrm{Var}[X]$$

jak widać, jest obciążony, albowiem $\mathbb{E}[\hat{\sigma}^2(X)] \neq \mathrm{Var}[X]$. Estymator nieobciążony wariancji to:

$$S^2(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \mathbb{E}[S^2(X)] = \mathrm{Var}[X]$$

4.3 Metoda najmniejszych kwadratów

Jest to metoda estymowania parametrów modelu. Model ma postać $f(x,\theta)$, gdzie $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$. Celem jest znalezienie wartości θ dla której minimalizowane jest:

$$r_i = y_i - f(x_i, \theta) \quad S = \sum_{i=1}^{n} r_i^2$$

Rozwiązaniem problemu jest znalezienie takich m, dla których gradient S jest równy zero:

$$\nabla_\theta S = 0$$

4.4 Metoda największej wiarygodności

W tej metodzie mamy funkcję wiarygodności:

$$L(\theta) = \prod_{i=1}^{n} f(y_i, \theta)$$

gdzie $f(y_i, \theta)$ to prawdopodobieństwo wystąpienia wartości y_i dla parametrów θ. Celem jest znalezienie takich θ, dla których $L(\theta)$ jest maksymalne.