

Machine Learning Math Study Sheet

1. Linear Algebra

Examples

1. Dot product:

$$\mathbf{a} \cdot \mathbf{b} = [1, 2] \cdot [3, 4] = 1 \times 3 + 2 \times 4 = 11$$

2. Matrix-vector multiplication:

$$\begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 8 \\ 11 \end{bmatrix}$$

3. Transpose and product:

$$\mathbf{A}^T \mathbf{A}, \quad \text{for } \mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Exercises

1. Compute the dot product of $[3, -2, 5]$ and $[1, 4, -1]$.
2. Multiply the matrix $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ with the vector $\begin{bmatrix} 5 \\ 6 \end{bmatrix}$.
3. Find the transpose of $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.
4. Compute the determinant of $\begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix}$.
5. Verify if the vectors $[1, 2]$ and $[2, -1]$ are orthogonal.

2. Calculus (Multivariable)

Examples

1. Gradient:

$$\nabla f(x, y) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right], \quad f(x, y) = x^2 + xy$$

2. Derivative of sigmoid:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad \sigma'(x) = \sigma(x)(1 - \sigma(x))$$

3. Partial derivative:

$$\frac{\partial}{\partial x}(x^2y + y^3) = 2xy$$

Exercises

1. Find the gradient of $f(x, y) = 3x^2 + 2y^2 + xy$.
2. Compute $\frac{d}{dx}[\tanh(x)]$.
3. Evaluate $\frac{\partial}{\partial y}(x^2y + y^2)$.
4. Derive the derivative of the ReLU function: $\text{ReLU}(x) = \max(0, x)$.
5. Show that the derivative of the sigmoid function is $\sigma(x)(1 - \sigma(x))$.

3. Probability & Statistics

Examples

1. Conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2. Expectation:

$$E[X] = \sum xP(x)$$

3. Variance:

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

Exercises

1. A coin is tossed 3 times. What is the probability of getting exactly 2 heads?
2. Given $P(A) = 0.6$, $P(B) = 0.5$, $P(A \cap B) = 0.3$, find $P(A|B)$.
3. Compute $E[X]$ for $X = \{1, 2, 3\}$ with $P(X) = \{0.2, 0.5, 0.3\}$.
4. Find the variance of a die roll.
5. A biased coin shows heads with $P(H) = 0.7$. What is the expected number of heads in 10 tosses?

4. Optimization

Examples

1. Gradient descent update:

$$\theta := \theta - \alpha \nabla J(\theta)$$

2. Minimizing a quadratic:

$$f(x) = (x - 3)^2 \Rightarrow \min \text{ at } x = 3$$

3. L2 regularization:

$$J(\theta) = \text{MSE} + \lambda \|\theta\|^2$$

Exercises

1. Perform 1 step of gradient descent on $f(x) = x^2 + 2x$ with learning rate 0.1, starting at $x = 2$.
2. Find the minimum of $f(x) = x^2 + 4x + 4$.
3. Differentiate $J(\theta) = \theta^2 + 3\theta + 2$.
4. Compute gradient of $f(x, y) = x^2 + y^2$.
5. Given a loss function $L(w) = (wx - y)^2$, compute $\frac{dL}{dw}$.