Exercices

Luxformel

Mathematics for Machine Learning

Exercise 1 : Basic vector operations

Let

$$\mathbf{u} = (1, 2, -1)^\top, \quad \mathbf{v} = (0, 1, 3)^\top.$$

Compute:

1. $\mathbf{u}, \mathbf{v}$
2. $\|\mathbf{u}\|_2$ and $\|\mathbf{v}\|_2$
3. Projection of $\mathbf{u}$ onto $\mathbf{v}$.

Exercise 2 : Linear system

Solve:

$$A = \begin{pmatrix} 2 & -1 & 0 \\ 1 & 1 & 1 \\ 0 & 2 & -1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}, \quad A\mathbf{x} = \mathbf{b}.$$

Exercise 3 : Data interpretation

Given points (1,2), (2,3), (3,5), form the design matrix $\mathbf{X}$ for

$$y = w_0 + w_1 x$$

and write the **normal equations** for least squares.

Exercise 4 : Partial derivatives

For

$$f(x, y) = 3x^2y - 4xy + \sin y,$$

compute $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

Exercise 5 : Gradient and Hessian

Let

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x},$$

compute $\nabla f(\mathbf{x})$, $H(f)$, and state when f is convex.

Exercise 6 : Chain rule

Let

$$g(\mathbf{x}) = \log(1 + e^{\mathbf{a}^\top \mathbf{x}}),$$

compute $\nabla g(\mathbf{x})$.

Exercise 7 : One-dimensional GD

For

$$f(x) = x^2 - 4x + 5,$$

starting from $x_0 = 0$, apply gradient descent with $\eta = 0.1$ for three iterations.

Exercise 8 : Quadratic form

For

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x},$$

find the largest η guaranteeing convergence.

Exercise 9 : Least squares

For

$$J(w) = \frac{1}{2m}\|Xw - y\|_2^2,$$

derive $\nabla_w J(w)$ and write one gradient descent update.

Exercise 10 : Ridge closed form

Show that

$$w^\star = (X^\top X + \lambda I)^{-1} X^\top y.$$

Why does λI improve conditioning?

Exercise 11 : L1 vs L2 regularization

Explain why **L1** produces sparse models while **L2** does not.

Exercise 12 : Regularized gradient

For

$$J_\lambda(w) = \frac{1}{2m}\|Xw - y\|_2^2 + \frac{\lambda}{2}\|w\|_2^2,$$

compute $\nabla_w J_\lambda(w)$.

Exercise 13 : Eigenpairs & PCA intuition

Given

$$C = \frac{1}{m} X^\top X,$$

explain why eigenvectors with largest eigenvalues are **principal components**, and find the projection of a point $\mathbf{x}$.

Exercise 14 : Gradient descent rate

For

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x},$$

show that gradient descent with

$$0 < \eta < \frac{2}{\lambda_{\max}}$$

converges linearly with rate involving

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}.$$