Principles of Machine Learning Course - Instructor: S. A. Emami
Dept. of Aerospace Engineering - Sharif University of Technology - Spring 2025
Due: 15 March 2025, Before midnight.

HW#1

Submission: Submit your assignment as a single zip file containing your code and report (in PDF). Name the file as `HW1_LastName.zip` and upload it to the CW portal.

Problem 1: Linear Independence and Orthogonality

Given the vector $v_1 = \begin{bmatrix} 2 & -5 & 3 \end{bmatrix}^T$,

a) Express vector v_1 as a linear combination of the vectors $u_1 = \begin{bmatrix} 1 & -3 & 2 \end{bmatrix}^T$, $u_2 = \begin{bmatrix} 2 & -4 & -1 \end{bmatrix}^T$ and $u_3 = \begin{bmatrix} 1 & -5 & 7 \end{bmatrix}^T$.

b) Determine the value of k such that the vectors $u = \begin{bmatrix} 3 & 3k & -4 \end{bmatrix}^T$ and v are orthogonal.

Problem 2: Eigenvalues and Eigenvectors

Consider the matrix A:

$$A = \begin{bmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{bmatrix}$$

a) Find the eigenvalues and eigenvectors of A.

b) Is the matrix A diagonalizable? In other words, does there exist an invertible matrix P such that, with the following relation, the resulting matrix D is diagonal:

$$D = P^{-1}AP$$

c) Determine the rank of the matrix.

d) Show that the sum of the eigenvalues is equal to the trace of the matrix (the sum of its diagonal elements).

e) Show that the product of the eigenvalues is equal to the determinant of the matrix.

Problem 3: Linear Transformations

Each point $(x, y) \in \mathbb{R}^2$ can be identified with the point $(x, y, 1)$ on the plane in $\mathbb{R}^3$ that lies one unit above the xy-plane. We say that (x, y) has *homogeneous coordinates* $(x, y, 1)$.

a) Based on the homogeneous coordinates, find the corresponding 3×3 transformation matrix that translates a 2D point (x, y) to a shifted point $(x + h, y + k)$.

b) Any linear transformation on $\mathbb{R}^2$ is represented with respect to homogeneous coordinates by a partitioned matrix of the form

$$\begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}$$

where A is a 2×2 matrix. At first, find the corresponding 3×3 transformation matrix that scales x by s and y by t. Afterward, find the 3×3 transformation matrix that result in a counterclockwise rotation of ϕ about the origin.

c) Find the 3×3 matrix that corresponds to the composite transformation of a scaling by .3, a rotation of 90 degrees about the origin, and finally, a translation that adds $(-0.5, 2)$ to the point (x, y).

d) (Bonus) Write a Python program that creates the original figure and applies the given transformations in part c to obtain the final figure.

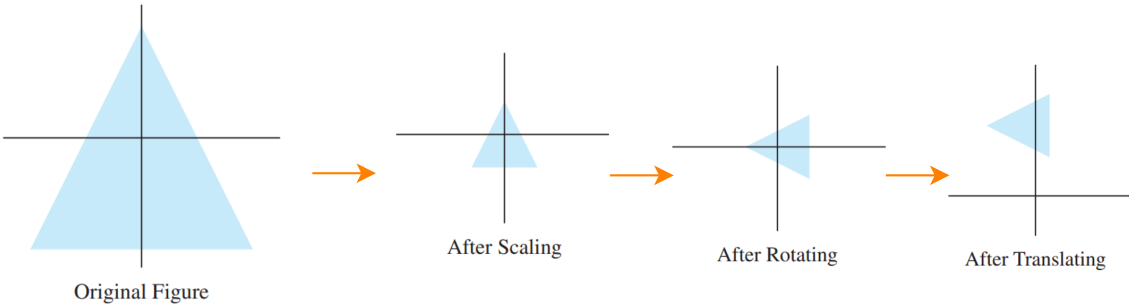

Figura 1: Visualization of the original figure and the sequential transformations

Problem 4: Calculus

Compute the derivatives df/dx of the following functions by using the chain rule. Provide the dimensions of every single partial derivative. Describe your steps in detail.

a) $f(z) = log(1+z), \quad z = x^T x, \quad x \in \mathbb{R}^N$

b) $f = tanh(z), \quad z = Ax + b, \quad x \in \mathbb{R}^N, \quad A \in \mathbb{R}^{M \times N}, \quad b \in \mathbb{R}^M, \quad f \in \mathbb{R}^M$

Problem 5: Linear Algebra, Least Squares and Calculus

Consider the optimization problem of least squares with $\ell 2$-regularization:

$$w^* = \text{argmin}_w f(w)$$
$$f(w) = \frac{1}{2n} \left\| X^T w - y \right\|_2^2 + \frac{\lambda}{2} \|w\|_2^2$$

To solve the above problem, we use the gradient descent method step by step. In each step, we move in the opposite direction of the gradient to reach a local minimum for the optimization problem. The initial value is chosen randomly.

$$w_{k+1} = w_k - \alpha \nabla f(w_k)$$

The value of α is an arbitrary number and is considered as:

$$\alpha = \frac{1}{\sigma_{max}(A)}$$

where $\sigma_{max}(A)$ is the largest singular value of matrix A, which is defined as follows:

$$A = \frac{1}{n} X X^T + \lambda I$$

a) Prove that:

$$\nabla f(w) = Aw - \frac{1}{n} Xy = A(w - w^*)$$

b) Prove that matrix A is positive semi-definite.

c) Prove that:

$$\|w_{k+1} - w^*\| \leq \left(1 - \frac{\sigma_{\min}(A)}{\sigma_{\max}(A)}\right) \|w_k - w^*\|$$

Problem 6: Least Square Method for Curve Fitting

Find the least-squares line $y = \beta_0 + \beta_1 x$ that best fits the data $(-2,3), (-1,5), (0,5), (1,4)$ and $(2,3)$. Suppose the errors in measuring the y-values of the last two data points are greater than for the other points. Weight these data half as much as the rest of the data. (Hint: Formulate this as a weighted least squares problem in the form $WAx = Wy$, and find the solution x^* by applying the least squares method.)

Problem 7: Optimization

A manufacturing company produces two products, x and y. The profit function is given by:

$$P(x, y) = 100x + 150y - 0.1x^2 - 0.2y^2 - 0.05xy$$

The company has the constraint of that the total resources used cannot exceed 1000 units:

$$2x + 3y \leq 1000$$

The goal is to maximize the profit. Solve the problem and obtain the optimal production levels and the corresponding profit in the following cases:

a) Unconstrained optimization

b) Constrained optimization (using Lagrangian multipliers)

Problem 8: Probability and Statistics

Consider normal random variables X and Y with means and variances $\mu_X = 0$, $\sigma_X^2 = 1$, $\mu_Y = -1$, $\sigma_Y^2 = 4$ and correlation coefficient $\rho = -1/2$.

(a) Find $P(X + Y > 0)$.

(b) Find a given that $X + 2Y$ and $aX + Y$ are independent.

(c) Find the correlation between $X + Y$ and $2X - Y$.

Problem 9: Image Noise Reduction using SVD

In this exercise, you should reduce the noise in the noisy image included in the homework files on CW using singular value decomposition (SVD) and then reconstruct the image. To this end, you can read the image using `matplotlib`, extract the RGB matrices, and apply SVD using the `np.linalg` library to compute the U, S, and V arrays. Then, construct the diagonal S matrix, where the singular values are placed along the diagonal:

$$S = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \sigma_k & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{bmatrix}$$

Smaller singular values (σ_k and beyond) have a significant influence on noise and are much lower in value compared to the first few singular values (e.g., σ_1, σ_2, etc.). To reduce noise, you should ignore these smaller singular values by replacing them with zero and reconstruct the image using only the larger singular values, which contain the main features of the image. Determine the appropriate value of k through trial and error to balance the trade-off between noise reduction and preserving the main features of the image. If k is too large, the noise may remain unchanged, whereas if k is too small, the image will be overly compressed, potentially losing resolution and quality. Save the final image as a PNG file using `matplotlib`. A sample output of the processed image is shown below:

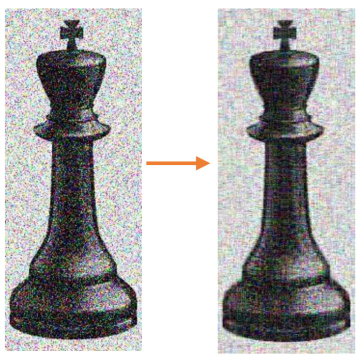

Figura 2: Noisy image (left) and processed image (right)

Good luck!