# CX4240 Computing for Data Analysis - Homework 1

Name:
GTID:

Deadline: 11:59 pm EST, Feb 04

- Submit your answers as one single PDF file on Gradescope.
- You will be allowed 2 total late days (48 hours) without penalty for the entire semester. Once those days are used, you will be penalized according to the following policy:
  - Homework is worth full credit before the due time.
  - It is worth 75% credit for the next 24 hours.
  - It is worth 50% credit for the second 24 hours.
  - It is worth zero credit after that.
- You are required to use Latex, or word processing software, to generate your solutions to the written questions. Handwritten solutions WILL NOT BE ACCEPTED.

## 1 Linear Algebra [35pts]

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. We say that $A$ is **positive semidefinite (PSD)** if

$$x^T A x \geq 0 \quad \text{for all } x \in \mathbb{R}^n.$$

(a) [**10 points**] Prove that if $A$ is PSD, then $cA$ is PSD for every scalar $c \geq 0$.

(b) [**10 points**] Suppose $A$ has an orthonormal eigenbasis $\{v_1, \ldots, v_n\}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. Show that for any $x \in \mathbb{R}^n$,

$$x^T A x = \sum_{i=1}^{n} \lambda_i (v_i^T x)^2.$$

(c) [**15 points**] Use part (b) to prove the following eigenvalue criterion:

$$A \text{ is PSD} \quad \Longleftrightarrow \quad \lambda_i \geq 0 \ \text{ for all } i = 1, \ldots, n.$$

(*Hint: You may use without proof that symmetric matrices have an orthonormal eigenbasis.*)

**Solution:**

## 2 Probability and Statistics [35pts]

1. **[15 points]** A discrete random variable $X$ takes values in $\{0, 1, 2, 3\}$ with

$$\mathbb{P}(X = 0) = 0.1, \quad \mathbb{P}(X = 1) = 0.2, \quad \mathbb{P}(X = 2) = 0.3, \quad \mathbb{P}(X = 3) = 0.4.$$

Let $X_1, \ldots, X_5$ be i.i.d. copies of $X$, and define the sample mean

$$\overline{X} = \frac{1}{5} \sum_{i=1}^{5} X_i.$$

Compute:

(a) $\mathbb{E}[X]$ and $\mathrm{Var}(X)$;

(b) $\mathbb{E}[\overline{X}]$ and $\mathrm{Var}(\overline{X})$.

(c) Define $Y = 2X - 1$. Compute $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$.

**Solution:**

2. **[20 points]**

Let $X \sim \mathcal{N}(\mu, \Sigma)$ and consider the affine transformation

$$Y = AX + b,$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

(a) Derive the distribution of $Y$ by computing its mean and covariance matrix.

(b) Now let's consider the one-dimensional case and let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X' \sim \mathcal{N}(\mu_2, \sigma_2^2)$, where $X, X' \in \mathbb{R}$. The Kullback-Leibler (KL) divergence between two distributions with densities $p(x)$ and $q(x)$ is defined as

$$D_{\mathrm{KL}}(p \,\|\, q) = \int_{\mathbb{R}^n} p(x) \log \frac{p(x)}{q(x)} \, dx.$$

Derive the closed-form expression for

$$D_{\mathrm{KL}}(\mathcal{N}(\mu_1, \sigma_1^2) \,\|\, \mathcal{N}(\mu_2, \sigma_2^2)).$$

**Solution:**

# 3 Optimization [30pts]

Suppose we want to minimize the function:

$$F(x, y) = \frac{10x^2 + y^2}{2}.$$

The actual minimum is $F = 0$ at $(x^*, y^*) = (0, 0)$. Solve the following questions in *vector* notation.

1. **[10 points]**

   Give the expression of the gradient vector $\nabla F$ at point $(x, y)$.

**Solution:**

2. **[10 points]** Let the initial point be $(x_0, y_0) = (1, 1)$. Perform gradient descent with step size $s = 0.5$ for two iterations. For each iteration, explicitly show:

   - the gradient computation, and
   - the updated solution.

   Based on your results, discuss whether the resulting solution sequence will converge to the optimal solution.

**Solution:**

3. **[10 points]** Again, let $(x_0, y_0) = (1, 1)$. Perform gradient descent with step size $s = 0.1$ for two iterations. As before, clearly demonstrate:

   - the gradient computation, and
   - the updated solution at each step.

   Determine whether the resulting sequence is convergent.

**Solution:**

## References

**Solution:**
Please mention any AI tools, people, post or blog etc. you used.