ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Foundations of Data Science — Assignment date: Saturday, January 31, 2026, 9:15
Fall 2025 — Due date: Saturday, January 31, 2026, 12:15

Final Exam – CE 1 4

This exam is open book. No electronic devices of any kind are allowed. There are 4 problems. Choose the ones you find easiest and collect as many points as possible. Good luck!

Name: ______________________________

Problem 1	/ 10
Problem 2	/ 10
Problem 3	/ 10
Problem 4	/ 10
Total	/40

Problem 1 (A perspective on PCA – 10 pts)**.** Let $\mathbf{X} \in \mathbb{R}^d$ be a zero-mean random vector with covariance matrix

$$\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{X}\mathbf{X}^\top].$$

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$ denote the eigenvalues of $\boldsymbol{\Sigma}$.

(i) (2 pts) Let $\mathbf{w} \in \mathbb{R}^d$ with $\|\mathbf{w}\|_2 = 1$. Show that

$$\mathrm{Var}(\mathbf{w}^\top \mathbf{X}) = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}.$$

(ii) (4 pts) Show that the unit vector $\mathbf{w}$ maximizing $\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$ is the eigenvector corresponding to the largest eigenvalue λ_1.

(iii) (3 pts) Let $\mathbf{u}_1$ be the eigenvector associated with λ_1, and define the rank-1 reconstruction

$$\hat{\mathbf{X}} = \mathbf{u}_1 \mathbf{u}_1^\top \mathbf{X}.$$

Compute the rank-1 reconstruction error

$$\mathbb{E}\left[\|\mathbf{X} - \hat{\mathbf{X}}\|_2^2\right].$$

(iv) (1 pt) Define the explained variance ratio of the first principal component as

$$\mathrm{EVR}(1) = \frac{\lambda_1}{\sum_{i=1}^d \lambda_i}.$$

Discuss and interpret EVR(1) in words. Then, give a practical criterion for deciding whether a rank-1 PCA approximation is sufficient.

(space for problem 1)

(space for problem 1)

(space for problem 1)

Problem 2 (The Log-Likelihood Ratio – 10 pts)**.** Given samples $X_1, X_2, \ldots, X_n$, we studied the problem of deciding whether the data consists of i.i.d. samples from distribution P_0 or from distribution P_1. We saw that the key is the log-likelihood ratio

$$\Lambda_n(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{P_1(X_i)}{P_0(X_i)}. \tag{1}$$

Now, let us assume that indeed, $X_1, X_2, \ldots, X_n$, are i.i.d. samples from P_1. In that case, we showed in class that $\mathbb{E}_{P_1}[\Lambda_n(X_1, \ldots, X_n)] = D(P_1 \| P_0)$. Give a good *upper* bound on the probability

$$\mathbb{P}_{P_1}\left\{ |\Lambda_n(X_1, \ldots, X_n) - D(P_1 \| P_0)| \geq \eta \right\}. \tag{2}$$

The better your bound, the more points. As always, full step-by-step justifications are required for full credit.

Hint: You may start by assuming that X_i are binary. In that case, P_1 is a Bernoulli(α_1) distribution and P_0 is a Bernoulli(α_0) distribution.

(space for problem 2)

(space for problem 2)

Problem 3 (Projection Theorem – 10 pts). Consider a Hilbert space H spanned by the orthonormal basis $\{\mathbf{z}_i\}_{i\in\mathbb{Z}_+}$. Let $G \subseteq H$ be a (Hilbert) subspace of H, spanned by the first N basis vectors, that is, $G = \text{span}\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$.

(i) (3 pts) For a fixed vector $\mathbf{d} \in H$, give an expression for $\min_{\hat{\mathbf{d}}\in G} \|\mathbf{d} - \hat{\mathbf{d}}\|^2$ in terms of $\mathbf{d}$ and $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$:

$$\min_{\hat{\mathbf{d}}\in G} \|\mathbf{d} - \hat{\mathbf{d}}\|^2 = \ldots \tag{3}$$

The simpler your expression, the more points you get. *Hint: Check your lecture notes. Recall that we are given an orthonormal basis of G!*

(ii) (1 pt) Now let H be the space of all zero-mean finite variance (real-valued) random variables. Here, we prefer to denote the abstract vectors $\mathbf{d}$ and $\mathbf{z}_i$ more explicitly as random variables D and Z_i. Take as the inner product $\langle \mathbf{d}, \mathbf{z}\rangle = \mathbb{E}[DZ]$. This is known to be a Hilbert space. As above, let $G \subseteq H$ be a subspace spanned by N *orthonormal basis vectors,* that is, $G = \text{span}\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$. Explicitly write what this means for the random variables $Z_1, Z_2, \ldots, Z_N$.

(iii) (3 pts) Take your general result from Part (i) and write it now for the special Hilbert space from Part (ii), that is, in terms of standard random variable notation:

$$\min_{\hat{D}\in\text{span}\{\ldots\}} \ldots = \ldots \tag{4}$$

(iv) (3 pts) In class, we have carefully studied estimation problems with respect to the mean-squared error (MSE) criterion, identifying the MMSE estimator and the LMMSE estimator as well as their respective performance. See Section 6.2 of the Lecture Notes. Precisely explain the connection between your result from Part (iii) above to MSE estimation! Feel free to refer to formulas from the lecture notes simply by their equation number.

(space for problem 3)

(space for problem 3)

Problem 4 (The Maximum Likelihood Estimator and I-projections – 10 pts)**.** Let X^n be i.i.d. random variables taking values in a finite set $\mathcal{X}$ of size k. Recall that when restricted to distributions from a set $\mathcal{P}$, the maximum likelihood estimator $\hat{p}_{\mathrm{MLE}} : \mathcal{X}^n \to \mathcal{P}$ is given as

$$P^*_{\mathrm{MLE}} = \hat{p}_{\mathrm{MLE}}(x^n) = \arg\max_{P \in \mathcal{P}} P\left(X^n = x^n\right).$$

Note that the two parts of this problem are independent of each other.

(i) (4 pts) Show that the maximum likelihood estimate P^*_{MLE} is equal to the I-projection

$$P^* := \arg\min_{P \in \mathcal{P}} D\left(\hat{P} \middle\| P\right)$$

where $\hat{P}$ denotes the empirical distribution of x^n.

(ii) (6 pts) Let P_0 be a distribution over $\mathcal{X}$ and let $\mathcal{P}$ be defined as the minimal exponential family

$$\mathcal{P} := \left\{P : P(x) = P_0(x) e^{-\langle \theta, \phi(x) \rangle - A(\theta)}, \theta \in \Theta\right\}$$

for some open parameter space $\Theta \subseteq \mathbb{R}^d$, sufficient statistic ϕ, and normalization function A. Then show that the maximum likelihood estimate P^*_{MLE} is the I-projection P^* of P_0 onto the linear family

$$\mathcal{L}(x^n) := \left\{P : E_P\left[\phi(X)\right] = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i)\right\}.$$

In other words, show that

$$\arg\min_{P \in \mathcal{L}(x^n)} D\left(P_0 \| P\right) = \arg\max_{Q \in \mathcal{P}} Q(X^n = x^n).$$

(space for problem 4)

(space for problem 4)

(space for problem 4)