Jean Monnet University

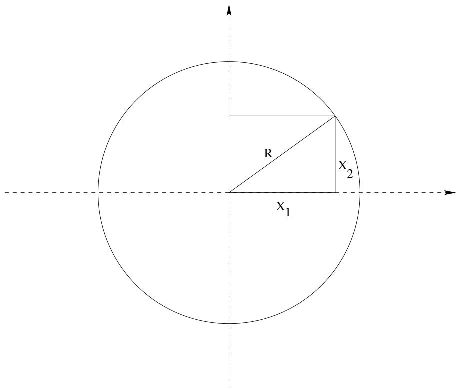
Exam of Data Analysis

December 2019 - **(3 hours)**

Pocket calculator and 2 two-sided A4 papers of handwritten notes are allowed.

Part I: Convex Optimization (2.5 points)

We are interested in minimizing the radius R of a circle of equation $R^2 = X_1^2 + X_2^2$ such that X_1 and X_2 describe a rectangle of perimeter equal to 6 (see the figure below for an illustration of the problem).



1. Write the formulation of the corresponding convex optimization problem.
2. Using the method of Lagrange multipliers, find the optimal values X_1 and X_2. Deduce R.

Part II: Random Variables and Probabilities (3 points)

Let X be a continuous random variable distributed according to the following density on $[0, 1]$:

$$f(x) = \frac{3}{2}x^2 + x + a, \qquad \forall x \in [0, 1] \text{ and } 0 \text{ otherwise.}$$

1. Find the value of the constant a such that $f(x)$ is an actual density function.
2. Compute the expected value $E(X)$ and the variance $V(X)$.
3. Compute $P(X > 0.5)$.

Part III: Likelihood Maximization (2.5 points)

Let $x_1, x_2, ..., x_n$ be the observations of n random variables $X_1, X_2, ..., X_n$ i.i.d. according to the following distribution f_θ:

$$f_\theta(x) = \frac{\sqrt{\theta}}{\sqrt{2\pi}} e^{-\frac{\theta x^2}{2}}$$

Using the likelihood maximization, find the empirical estimate $\hat{\theta}$ of the parameter θ. For the sake of simplicity, you can make use of the log-likelihood.

Part IV: Principal Component Analysis (6 points)

Let $X = \{A, B, C\}$ be a set of 3 examples lying in a 3-dimensional feature space where $A = (1, 2, 4)$, $B = (0, 1, 2)$ and $C = (2, 3, 6)$.

1. Compute the 3×3 covariance matrix Σ from the **zero mean values** of A, B and C.
2. Compute the determinant of Σ. What do you conclude?
3. Find the eigenvalues of Σ. Deduce the rank of Σ.
4. Compute the unit-eigenvector $\vec{u}$ corresponding to the largest eigenvalue.
5. Find the new coordinates of the 3 points in $\mathbb{R}$ according to $\vec{u}$.

Part V: Linear Algebra (2 points)

Let us consider the following matrix $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 3 \\ 2 & 3 & 2 \end{pmatrix}$.

1. Compute the inverse matrix A^{-1}.

Part VI: Multiple Choice Questions (4 points)

Guidelines:
- Circle the letter corresponding to the correct answer (**only one is correct**).
- You can leave questions unanswered. Each correct answer **adds 0.5**
- Each incorrect answer **subtracts 0.25**

1. You run gradient descent for 15 iterations with a learning rate $\alpha = 0.3$ and compute $J(\theta)$ after each iteration. You find that the value of $J(\theta)$ decreases quickly and then levels off. Based on this, which of the following conclusions seems most plausible?

   a. Rather than using the current value of α, use a larger value (say $\alpha = 1.0$).

   b. Rather than using the current value of α, use a smaller value (say $\alpha = 0.1$).

   c. $\alpha = 0.3$ is an effective choice of learning rate.

2. Which of the following is a reasonable way to select the number of principal components in PCA?

   a. Choose the smallest value so that at least 90% of the variance is retained.

   b. Choose this number to be 90% of the number of samples.

   c. Choose the largest value so that at least 90% of the variance is retained.

3. How can you prevent k-means algorithm from getting stuck in bad local optima?

   a. Set the same seed value for each run.

   b. Use multiple random initializations.

   c. Both a and b.

4. Suppose you have trained a logistic regression classifier and it outputs a new example x with a prediction $h(x) = 0.2$. This means

   a. The estimate for $P(y = 1|x) = 0.8$

   b. The estimate for $P(y = 0|x) = 0.2$

   c. The estimate for $P(y = 1|x) = 0.2$

5. In linear regression, the mean-Square error (MSE) is used as follows:

$$\min_{\theta} J(\theta) = \min_{\theta} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2 .$$

   We have a half term in the front because,

   a. Scaling MSE by half makes gradient descent converge faster.

   b. Scaling MSE by half makes gradient descent get better results.

   c. Scaling MSE by half simplifies the close-form solution.

6. What is the rank of the following matrix A?

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

   a. 1.

   b. 2.

   c. 3.

7. Complete the following sentence. *Logistic regression is a....regression technique that is used from data having a ...outcome*

   a. linear,continuous.

   b. nonlinear, continuous.

   c. nonlinear, binary.

8. What is the correct statement?

   a. PCA and t-SNE are both linear.

   b. PCA is linear while t-SNE is not.

   c. The solution of both PCA and t-SNE can be expressed in closed-form.

Part IV: Principal Component Analysis (6 points)

Let $X = \{A, B, C\}$ be a set of 3 examples lying in a 3-dimensional feature space where $A = (1, 2, 4)$, $B = (0, 1, 2)$ and $C = (2, 3, 6)$.

1. Compute the 3×3 covariance matrix Σ from the **zero mean values** of A, B and C.
2. Compute the determinant of Σ. What do you conclude?
3. Find the eigenvalues of Σ. Deduce the rank of Σ.
4. Compute the unit-eigenvector $\vec{u}$ corresponding to the largest eigenvalue.
5. Find the new coordinates of the 3 points in $\mathbb{R}$ according to $\vec{u}$.

Part V: Linear Algebra (2 points)

Let us consider the following matrix $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 3 \\ 2 & 3 & 2 \end{pmatrix}$.

1. Compute the inverse matrix A^{-1}.

Part VI: Multiple Choice Questions (4 points)

Guidelines:
- Circle the letter corresponding to the correct answer (**only one is correct**).
- You can leave questions unanswered. Each correct answer **adds 0.5**
- Each incorrect answer **subtracts 0.25**

1. You run gradient descent for 15 iterations with a learning rate $\alpha = 0.3$ and compute $J(\theta)$ after each iteration. You find that the value of $J(\theta)$ decreases quickly and then levels off. Based on this, which of the following conclusions seems most plausible?

   a. Rather than using the current value of α, use a larger value (say $\alpha = 1.0$).
   b. Rather than using the current value of α, use a smaller value (say $\alpha = 0.1$).
   c. $\alpha = 0.3$ is an effective choice of learning rate.

2. Which of the following is a reasonable way to select the number of principal components in PCA?

   a. Choose the smallest value so that at least 90% of the variance is retained.
   b. Choose this number to be 90% of the number of samples.
   c. Choose the largest value so that at least 90% of the variance is retained.

3. How can you prevent k-means algorithm from getting stuck in bad local optima?

   a. Set the same seed value for each run.
   b. Use multiple random initializations.
   c. Both a and b.