EE 212 – Mathematical Foundations for Machine Learning and Data Science

**Final-Examination (Fall 2021)**

---

**INSTRUCTIONS:**

- We require you to solve the exam without any external or electronic assistance.
- We require you to solve the exam on A4 paper, use new sheet for each question and write sheet number on every sheet.
- Clearly outline all your steps in order to obtain any partial credit.
- The exam is closed book and notes. You are allowed to have two A4 sheets with you with hand-written notes on both sides. Calculators can be used.
- For the sake of completeness, we require you to write the following statement on your first page of submission: *I commit myself to uphold the highest standards of (academic) integrity.*
- If you are ready, please proceed to the next page.

EE 212 – Mathematical Foundations for Machine Learning and Data Science

**Final-Examination (Fall 2021)**

**Total marks:** 60 points **Duration:** 1 hour and 45 minutes

---

**Problem 1.** (**10 pts**)

Consider a random variable $X$ with pdf given by

$$f_X(x) = \begin{cases} A(x^2 - b) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

and expected value $E[X] = \frac{7}{12}$. Find values of $A$ and $b$.

**Problem 2.** (**10 pts**)

The problem of finding model parameters $x \in \mathbf{R}^m$ in linear regression can be formulated as an optimization problem where we minimize the following least-squares function

$$f(x) = \|y - Ax\|_2,$$

for $y \in \mathbf{R}^n$ and $A \in \mathbf{R}^{n \times m}$. Show that the function $f(x)$ is convex in $x$.

**Hint:** Show that the Hessian of the function is positive semi-definite.

**Problem 3.** (**10 pts**)

In a rare collaboration, computer scientists, doctors and engineers developed tests for detecting Examophobia disease among the students. Research suggests that

- the probability of positive outcome for the student who is suffering from this disease is 0.8.
- the probability of positive outcome for the student who is not suffering from this disease is 0.01.
- 35% of the students suffer from Examophobia.

What is the probability that the student tested positive is suffering from Examophobia?

**Problem 4.** (**10 pts**)

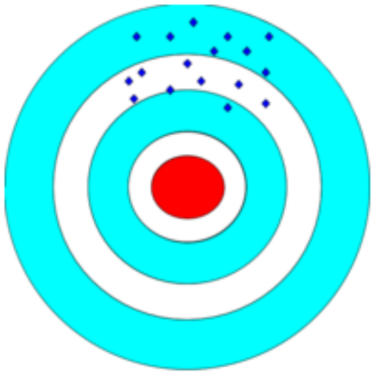
Emirates Airlines have developed 2 different classifiers (A and B) for the prediction whether a flight originating from Dubai will arrive at its final destination on time or not. True or Positive here is 'On time' and it refers to the case when the flight that is no more than 5 minutes late as per schedule. The classifiers were tested on a data-set of 500 flights, and the results are as follows:

| | Actual | |
|---|---|---|
| | On time | Late |
| Classifier A, predicted on time | 131 | 155 |
| Classifier A, predicted late | 19 | 195 |
| Classifier B, predicted on time | 82 | 72 |
| Classifier B, predicted late | 68 | 278 |

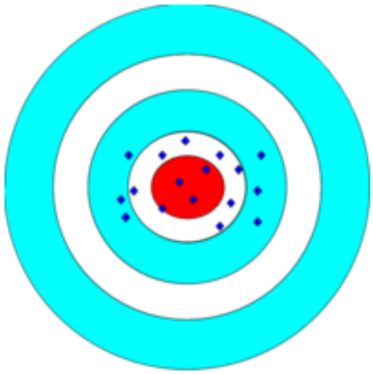
Which is the preferable classifier in terms of $F_1$ score?

**Problem 5.** (**6 pts**)

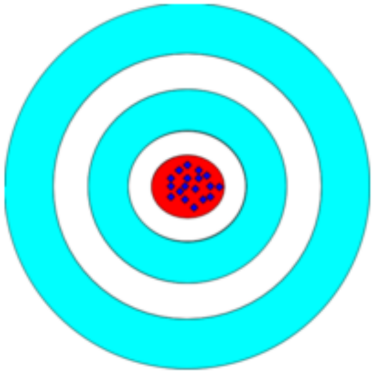
We can visualize **bias** and **variance** in machine learning using a classic example of a dart game. The actual (true) output can be considered as the bull's-eye on a target and the estimated values for each sample as arrows. Given the following boards for 4 players (models), which of the following statement(s) is/are correct. Provide a brief explanation to support your answer.



Board of player 1



Board of player 2



Board of player 3

Board of player 4

(a) Player 4 has high variance compared to player 1

(b) Player 4 has low variance compared to player 1

(c) Player 2 exhibited more bias compared to player 3.

**Problem 6.** (**6 pts**)

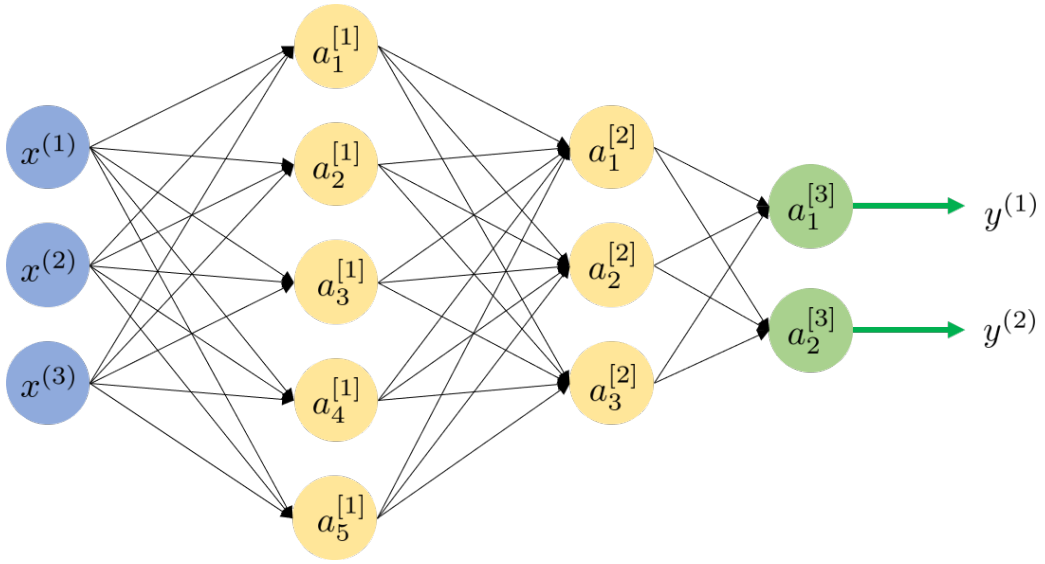
For kNN **regression** defined as an average of k nearest neighbors, you are given the following five training instances (2 inputs and one output)

- Data Point 1: $x_1 = 2; x_2 = 1; y = 4$
- Data Point 2: $x_1 = 6; x_2 = 3; y = 2$
- Data Point 3: $x_1 = 2; x_2 = 5; y = 2$
- Data Point 4: $x_1 = 6; x_2 = 7; y = 3$
- Data Point 5: $x_1 = 10; x_2 = 7; y = 3$

Predict the output value using kNN for $k = 2$.

**Problem 7.** (**8 pts**)

Consider a neural network shown in the figure below. We have three inputs and two outputs.




(a) (**2 pts**) Define the weight matrices and bias vectors for each layer using the appropriate notation and specify their sizes.

(b) (**4 pts**) Formulate a set of equations for Forward pass.

(c) (**2 pts**) Calculate the total number of trainable parameters of the neural network.