There are four problems below. Please choose three to solve. If you choose to solve all four, only the lowest three scores will count. Show all your work and write in a readable way.

Question 1

Overfitting, underfitting and cross-validation.

(a) Please give pictorial examples in regression for overfitting and underfitting respectively, and explain the relation between over/underfitting and model class flexibility.

(b) We would like to perform k-fold cross-validation to select models. What should k be? Discuss the pros and cons of large or small values of k (in terms of bias, variance and computation).

Question 2

Bayes classifier

(a) Write down the Bayes classifier $f : X \to Y$ (the classifier that minimizes the expected loss $E(L(Y, f(X))))$ for binary classification $Y \in \{-1, +1\}$ with non 0-1 loss (a is the loss for falsely predicting negative and b is the loss for falsely predicting positive). Simplify the classification rule as much as you can.

(b) If $P(X|Y = y)$ is a multivariate Gaussian and assuming the 0/1 loss, write the Bayes classifier as $f(X) = \text{sign}(h(X))$ and simplify h as much as possible. What is the geometric shape of the decision boundary?

(c) Repeat (b) when the two Gaussians have identical covariance matrices. What is the geometric shape of the decision boundary?

(d) Repeat (b) when the two Gaussians have covariance matrix that equals the identity matrix. Describe the geometric shape of the decision boundary as much as possible.

Question 3

Multiclass classification.

Multiclass classification tries to assign one of several class labels (rather than binary labels) to an object. Can you give **THREE** ways to solve multiclass classification problem? What are the pros and cons of these different methods, e.g., in terms of computional complexity or the applicability of the method?

Question 4

Hidden Markov Models.

Suppose we are in a casino. The casino player has two biased dice. In each toss, the player switches back and forth between the two dice. Given N die tosses denoted by $x^1, x^2, ..., x^i, ..., x^N$, $x^i \in \{1, 2, ..., 6\}$ indicates the outcome of a specific toss. Let $z^1, z^2, ..., z^i, ..., z^N$ represent the identity of the die tossed each time, where $z^i = 0$ means the first die and $z^i = 1$ means the second die for the i^{th} toss. The transition probabilities between any two states are represented by $p(z^t = j|z^{t-1} = i) = A_{ij}, i, j \in \{0, 1\}$. At the beginning, with probability π_0 the player chooses the first dice, and with probability $\pi_1 = 1 - \pi_0$ the player chooses the second dice. Given a specific dice, the probability associated with each outcome is given by $p(x^t = k|z^t = i) = E_{ik}, i \in \{0, 1\}, k \in \{1, ..., 6\}$, which is a multinomial distribution. As the observer, we do not know $A_{ij}, i, j \in \{0, 1\}$, π_i, and $E_{ik}, i \in \{0, 1\}, k \in \{1, ..., 6\}$ in advance.

(a) Suppose we can observe all the variables x^i and z^i (possibly multiple sequences), how do we estimate the parameters A_{ij}, π_i and E_{ik} using maximum likelihood estimation? Derive the results.

(b) Suppose we only observe the x^i variables (possibly multiple sequences), how do we estimate the parameters A_{ij}, π_i and E_{ik} using maximum likelihood estimation? Derive an algorithm.