Tutorial 12 - Preparatory Questions for the Exam

Question1

[marks]

What are eigenvalues and eigenvectors of matrix:

Eiganvalue = [4, 2]
Eiganvector = [1, 0

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

Question2

[marks]

Alice and Bob both need to buy a bicycle. The bike store has a stock of 4 green, 3 yellow and 2 red bikes. Alice randomly picks one of the bikes and buy it. Immediately after, Bob does the same. The sale price of the green, yellow and red bikes are \$300, \$200 and \$100, respectively.
Let A be the event that Alice bought a green bike, and B be the event that Bob bought a green bike.

1. What is $\mathbf{P}(A)$? What is $\mathbf{P}(A|B)$? Are A and B independent events? Justify your answer.

   The events are not independents, the event probability of Bob picking green is dependent on what Alice picks first.

   P(A) = 4/9 P(A|B) = P(B|A)*P(A)/P(B), P(B|A) = 3/8, P(A|B) = 3/8 * 4/9 / 4/9 = 3/8

2. What is the probability that Alice and Bob bought bicycles of different colors?

   Probability of buying same colour = P(G,G) +P(Y, Y) + P(R, R) = 4/9 * 3/8 + 3/9 * 2/8 + 2/9 + 1/8 Probability of buying different colours = 1- ans = ans

3. What is the probability that at least one of them bought a green bike?

   P(A u B) = P(A) + P(B) - P(A n B) = P(A) + P(B) - P(A|B)P(B) = 4/9 + 4/9 - 3/8 * 4/9 = 0.7222

4. Given that Bob bought a green bike, what is the expected value of the amount of money spent by Alice?

   3/8 * 300 + 3/8 * 200 + 2/8 * 100 =

Question3

[marks]

In Linear Regression given the following cost function

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \left(y^{(i)} - h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) \right)^2 \tag{1}$$

where $h_{\boldsymbol{\theta}}\left(\mathbf{x}^{(i)}\right) = \boldsymbol{\theta}^T \mathbf{x}^{(i)}$, feature vector $\mathbf{x}^{(i)} \in R^d$ of the i-th sample, and there are n data samples. We usually use the gradient descent to learn the minimum value of the cost function: $\theta := \theta - \alpha \nabla J(\theta)$.

1. What is the name of the cost function above? Mean Squared Error

2. Show step-by-step the gradient descent update for this cost function.

Question4

[8 marks]

Suppose we are using a linear SVM (i.e., no kernel), with some large C value, and are given the following data set.

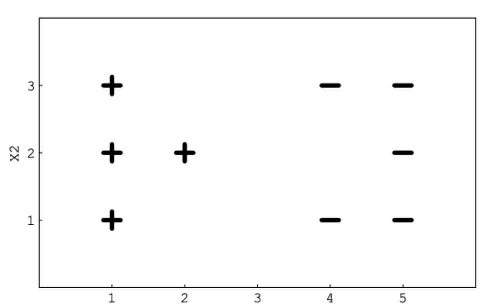

Figure 1: A tiny dataset for SVM

1. Draw and write the equation corresponding to the decision boundary of linear SVM. Give a brief explanation.
2. Circle the points such that by removing any one of them from the training set and retraining SVM, we would get a different decision boundary than training on the full data sets (give one or two sentence to explain)

Question5 [marks]

Explain why sometimes maximising the likelihood of logistic regression directly will lead to over-fitting. What can be done to fix the problem?

Question6 [marks]

Consider the random variables X, Y, Z which have the following joint distribution:

$$p(X, Y, Z) = p(X)p(Y|X)p(Z|Y). \quad (2)$$

Show that X and Z are conditionally independent given Y

Question7 [marks]

Consider the training set with ten data points shown in Fig Q2a. Use the k-NN algorithm with three nearest neighbours, i.e. 3-NN, to determine the class of an unknown data point $(x = 4, y = 4)$. Clearly indicate how you determined the nearest neighbours.

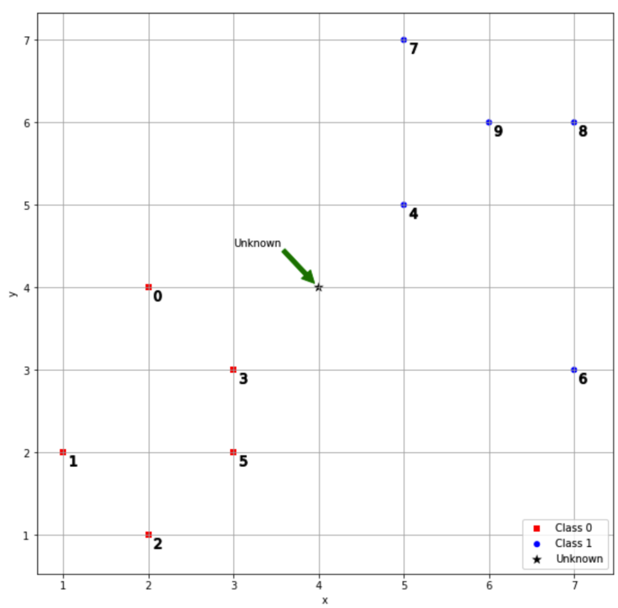

Figure 2: Fig. Q2a