Question Paper

Exam Date & Time: 29-May-2023 (10:00 AM - 01:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal
Second Semester Master of Engineering - ME (Artificial Intelligence and Machine Learning) Degree Examination - May 2023

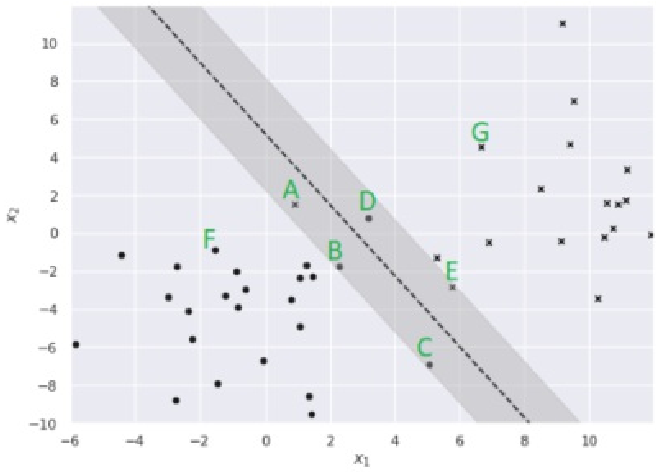
Machine Learning Principles and Applications [AML 5203]

Marks: 100 **Duration: 180 mins.**

Monday, May 29, 2023

Answer all the questions.

1) 1. [10 points] [TLO 1.1, CO 1] Answer the questions based on the SVM linear decision boundary shown below (10)



(a) *True/false*: The data is linearly separable.

(b) There are ____ support vectors.

(c) For each sample $A - G$, choose one of the following for the slack ξ with a brief explanation as to why:

$$\xi = 0, \; 0 < \xi < 1, \; \xi \geq 1.$$

(d) Calculate the full-margin width if $w = \begin{bmatrix} -0.7 & -0.4 \end{bmatrix}^{\mathrm{T}}$.

2) (10)

[10 points] [TLO 1.1, CO 1] Choose the correct option with a brief explanation in each of the following: a *large* value of the SVC hyperparameter C results in

(a) more/less misclassifications;

(b) more/less regularization;

(c) more/less overfitting;

(d) more/less number of support vectors;

(e) a decision boundary that is close to linear/nonlinear.

3) [10 points] [TLO 1.2, CO 1] Consider a data matrix for linear regression as follows: (10)

$$X = \begin{bmatrix} x^{(1)\mathrm{T}} & 1 \\ x^{(2)\mathrm{T}} & 1 \end{bmatrix}$$

$$X = \begin{bmatrix} \vdots & \vdots \\ x^{(n)\mathrm{T}} & 1 \end{bmatrix},$$

where $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ are p-vectors.

(a) What does the "1" in the data matrix represent?

(b) Consider the LMS loss function:

$$L = \frac{1}{2}\|X\theta - y\|^2.$$

How many components does the vector θ has?

(c) Fill in the missing entries below in an expanded representation of L:

$$L = \frac{1}{2}\sum_{i=1}^{?}(?\times?-?)^{?}.$$

(d) Use the fact that $\nabla_{\theta_j} L = \left(x^{(i)\mathrm{T}}\theta - y^{(i)}\right) x_j^{(i)}$ to write down the gradient descent update rule for θ_j.

(e) Using the result from the previous part, briefly explain when a parameter would have large and small changes.

4) [10 points] [TLO 2.2, CO 1] Suppose we have a dataset with 10^3 samples. Our model has 2 hyperparameters such that there are 4 values for each hyperparameter that we want to investigate. We want to do a nested-cross validation comprising a 5-fold outer-cross validation and a 3-fold inner-cross validation loop. (10)

(a) In one or two lines explain the purposes of the outer- and inner-cross validation loops?

(b) How many samples do we have in the (a) training part (b) test part (c) sub-train part (d) validation part?

(c) How many times in the nested-cross validation procedure will a model be trained?

5) [10 points] [TLO 3.1, CO 3] For a particular binary classification task, your friend uses a neural network with the following loss function: (10)

$$L = \boldsymbol{\alpha}\left[-y\log\left(a^{[1]}\right)\right] - \boldsymbol{\beta}\left[(1-y)\log\left(1-a^{[1]}\right)\right],$$

(a) The network has __ input and __ output layer.

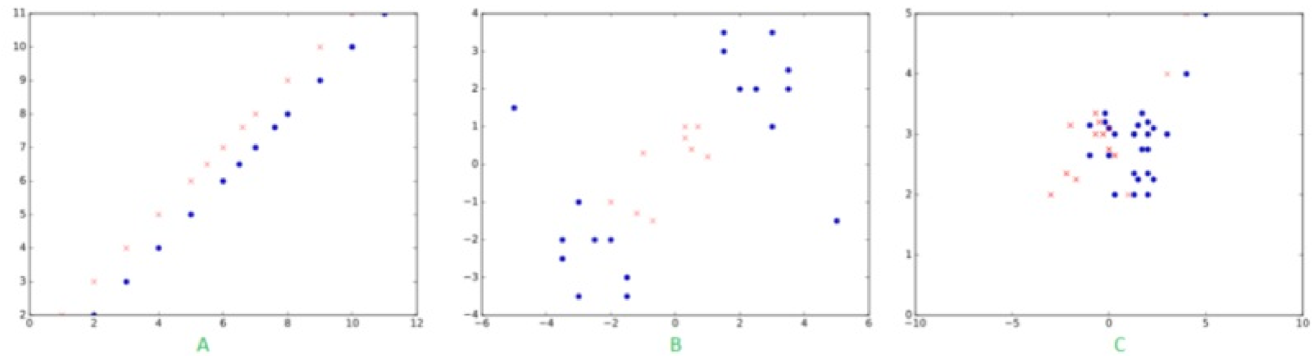
(b) What activation function is used for the output layer?

(c) Note that your friend introduces the (unknown) coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that have to be fine-tuned like any hyperparameter. What kind of a binary classification task might your friend be solving? Give a real-life problem setup and reasonable values for the two parameters.

(d) Is there any other approach without modifying the loss function for solving the real-life problem you just came up with?

6) (10)

[10 points] [TLO 2.1, CO 1] For the given data plots below, choose (1) a method you could use to classify the data, and (2) a method that is not reasonable to use for the given dataset. Briefly explain your choices.



7) [10 points] [TLO 1.2, CO 1] Suppose we have a dataset with n samples. We use the following model: (10)

$$p\left(y^{(i)} \mid \theta\right) = \frac{e^{-\theta}\theta^{y^{(i)}}}{y^{(i)}!}.$$

Recall that is called the Poisson model.

(a) Write down the likelihood $L(\theta)$ as a *product* of terms.

(b) Write down the log-likelihood $LL(\theta)$ as a *sum* of terms.

(c) Derive the MLE estimate for θ and interpret the result in plain English

(c) Derive the MLE estimate for θ and interpret the result in plain English.

8) [10 points] [TLO 2.1, CO 1] Suppose we want to predict how a reader feels about a book based on three binary features X_1, X_2, X_3 and one continuous feature (age) X_4. The response variable Y can take four possibles values: {like, love, haha, sad}. (10)

We have access to a dataset with 10,000 users. Each user in the dataset has a value for X_1, X_2, X_3, X_4 and Y. You can use a special query method `count` that returns the number of users in the dataset that satisfy certain conditions as shown in the following example usage:

> `count`($X_1 = 1, Y =$ haha) returns the number of users where $X_1 = 1$ and $Y =$ haha;
> `count`($Y =$ love) returns the number of users where $Y =$ love.

You are given a new user with $x_1 = 1, x_2 = 1, x_3 = 0$ and $x_4 = 20$. Suppose we want to use Naive Bayes for the prediction.

Write down a model for Y as:

$$P(y \mid \theta) = ?$$

Using the model, explain how you will compute $P(y = \text{like} \mid \theta), P(y = \text{love} \mid \theta), P(y = \text{haha} \mid \theta)$, and $P(y = \text{sad} \mid \theta)$ <u>only</u> using the `count` method.

9) [10 points] [TLO 2.1, CO 1] Continuing from the previous problem, write down a model for X_1 as: (10)

$$P(x_1 \mid y, \theta) = ?$$

Using the model, explain how you will compute $P(x_1 = 1 \mid y = \text{love}, \theta)$ <u>only</u> using the `count` method.

10) (10)

[10 points] [TLO 2.1, CO 1] Continuing from Question-8, write down a model for X_4 as:

$$P(x_4 \mid y, \mu, \Sigma) = ?$$

Note the following regarding the model above:

- X_4 (age) is a continuous random variable, and therefore here we are referring to the probability density and not the probability.
- There are two model parameters μ and Σ. How can they be estimated?

Using the model, briefly explain how you will compute $P(x_4 = 22 \mid y = \text{love}, \theta)$. Finally, conclude how you will predict the output label for the new user data given in Question-8 using the Naive Bayes algorithm.

-----End-----