# Question Paper

Exam Date & Time: 14-Feb-2024 (09:30 AM - 11:00 AM)

## MANIPAL ACADEMY OF HIGHER EDUCATION

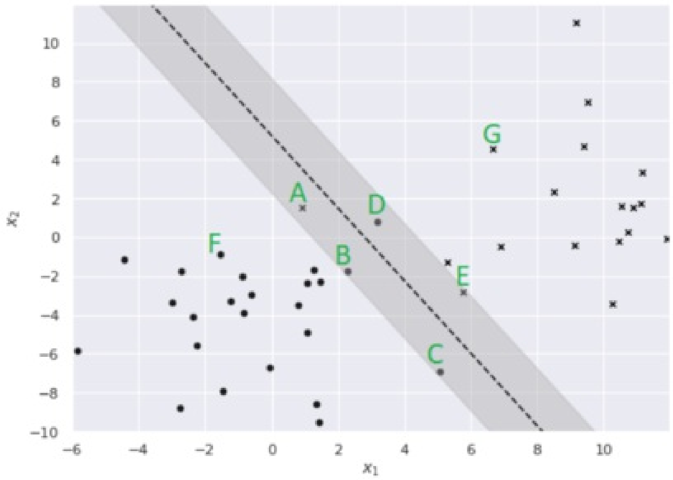
**Machine Learning Principles and Applications [AML 5203]**

**Marks: 50** **Duration: 90 mins.**

**Keep answers short and precise**

**Answer all the questions.**

1) [10 points] [L5, CO 1] Answer the questions based on the SVM linear decision boundary shown below: (10)




(a) Justify briefly if the data is linearly separable or not.

(b) How many support vectors are there?

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

$$\xi = 0 \text{ or } 0 < \xi < 1 \text{ or } \xi \geq 1.$$

(d) Calculate the full-margin width if the equation of the separating line (hyperplane) is $x_2 = -1.8x_1 + 5$.

2) (10)

[10 points] [L5, CO 2] Consider solving the MNIST classification problem (labels 0 through 9) using the hinge loss function-based formulation of SVM.

(a) Fill in the missing entries below in the $i$th sample's loss:

$$L_i = \sum_{\substack{j=? \\ j \neq y^{(i)}}}^{\boxed{?}} \max\left(0, z_j^{(?)} - z_{\boxed{?}}^{(i)} + \boxed{?}\right)$$

(b) Rewrite the $i$th sample's loss in terms of the weight vectors assuming that the bias trick is done.

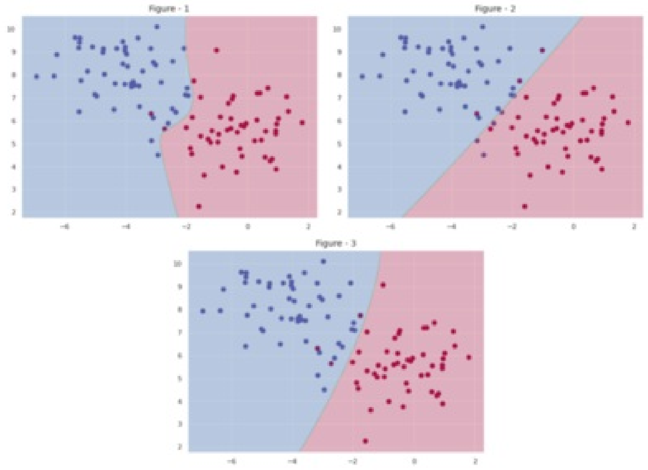
(c) What is the computational advantage of using a loss function-based formulation of SVM over the optimization-based formulation of SVM (such as SVC or LinearSVC)?

(d) The SVM algorithm presented here results in a linear decision boundary similar to the hyperplane generated by the optimization-based formulation of linear SVM. How can such a linear model over-fit? How can you avoid over-fitting? Justify using one or two lines for both questions.

3) [10 points] [L5, CO 2] Select the correct option in each of the following: a large value of the SVC hyperparameter $C$ results in (10)

- more/less misclassifications;
- more/less regularization;
- more/less overfitting;
- more/less number of support vectors;
- a decision boundary that is close to linear/nonlinear.

4) [10 points] [L2, CO 3] ] Identify and match the Kernel-SVC hyperparameter values $C = 10^5, 10^1$, and $10^{-1}$ with the corresponding figures below with a brief justification: (10)




5) (10)

[10 points] [L5, CO 3] Consider a dataset where a generic sample $\mathbf{x}$ has 2 features and can correspond to 2 output labels. Your friend tries to fit a linear SVM model to the dataset and concludes that introducing new features as follows would be needed for a more accurate classification model:

$$\mathbf{x}_{\text{old}} = \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\text{old features}} \mapsto \mathbf{x}_{\text{new}} = \underbrace{\begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}}_{\text{new features}}.$$

(a) Argue briefly as to why your friend is introducing these new features. What is a potential disadvantage in your friend's approach?

(b) Which form of SVM – the primal or dual – is your friend using. Explain briefly.

(c) Compute the dot product between samples $i$ and $j$ in the the new feature space: $\mathbf{x}_{\text{new}}^{(i)} \cdot \mathbf{x}_{\text{new}}^{(j)}$.

(d) Show that the dot product result you derived in the previous part can also be efficiently computed by first computing $\mathbf{x}_{\text{old}}^{(i)} \cdot \mathbf{x}_{\text{old}}^{(j)}$ in the old feature space and making use of that result. This is the kernel trick.

-----End-----