Question Paper

Exam Date & Time: 24-Apr-2024 (02:00 PM - 05:00 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 - April / May 2024

Machine Learning Principles and Applications [AML 5203]

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

Wednesday, April 24, 2024

Answer all the questions.

1) [C01, L5] Multiple choice question (justify your answer briefly) (10)

In linear SVMs, the decision boundary is a:
a) Sphere
b) Hyperplane
c) Parabola
d) Ellipse

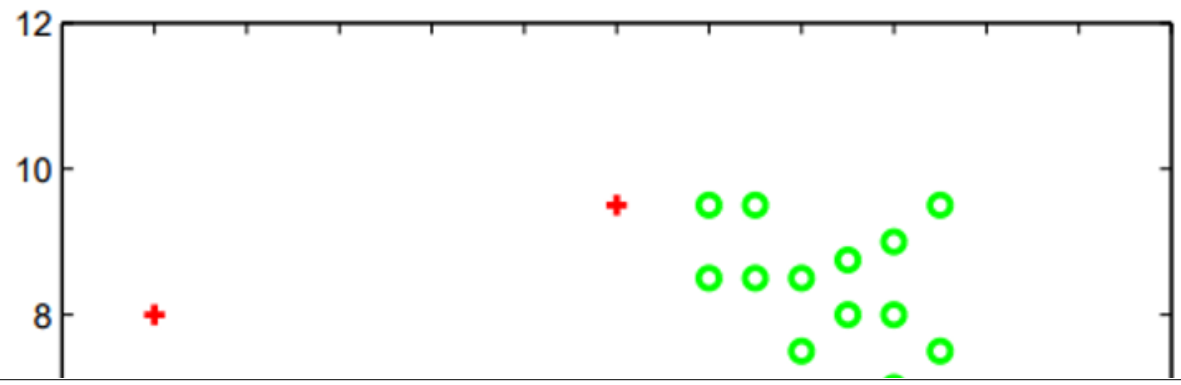
The main difference between hard and soft SVMs lies in:
a) Kernel function used
b) Margin maximization
c) Handling errors
d) All of the above

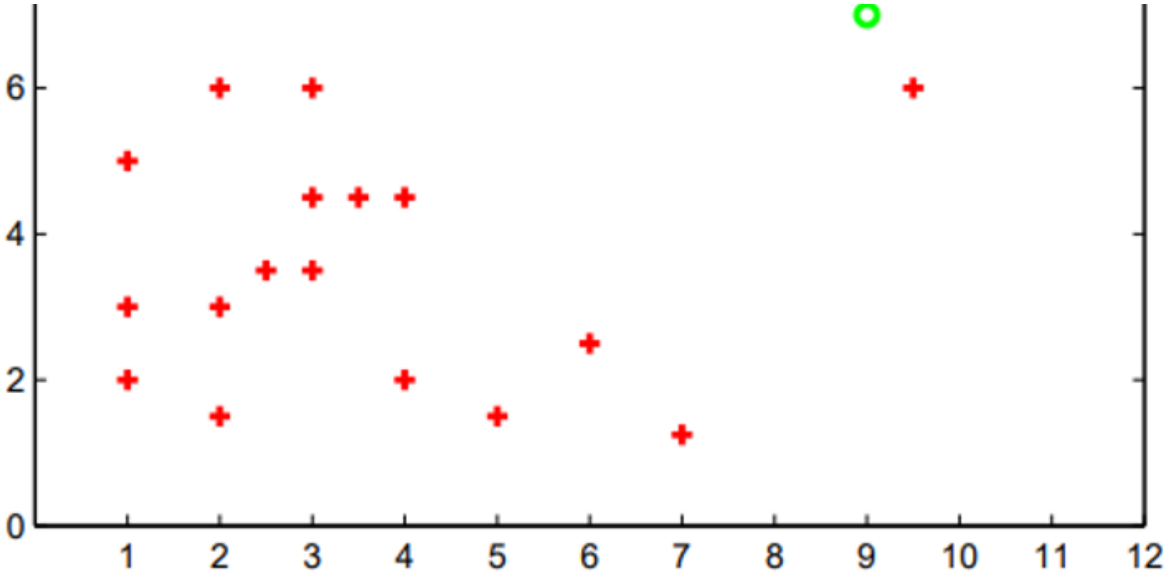
What is the main advantage of the kernel trick?
a) Reduces computational cost
b) Enables non-linear decision boundaries
c) Improves overfitting
d) All of the above

The minimum time complexity for training an SVM is $O(n^2)$. According to this fact, what sizes of datasets are not best suited for SVMs?
a) Large datasets
b) Small datasets
c) Medium-sized datasets
d) Size does affect the training

5) You are given a labeled binary classification data set with N data points and D features. Suppose that N < D. In training an SVM on this data set, which of the following kernels is likely to be most appropriate?
(a) Linear kernel
(b) Quadratic kernel
(c) Higher-order polynomial kernel
(d) RBF kernel

2) [C01, L5] Assume that we are training an SVM with a quadratic kernel- that is, our kernel function is a polynomial kernel of degree 2. You are given the data set presented in Figure. The slack penalty C will determine the location of the separating hyperplane. Give a one sentence answer/justification for each and draw your solution. (10)





- Where would the decision boundary be for very large values of C (i.e., C → ∞)? (Remember that we are using an SVM with a quadratic kernel.) Draw on the figure above. Justify your answer.
- For C ≈ 0, indicate in the figure below, where you would expect the decision boundary to be? Justify your answer.

3) [C02, L4] (10)

Define the concept of a soft margin in SVM and explain why it is necessary in cases where the data is not perfectly separable. Given a soft-margin SVM optimization problem with the objective:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi^{(i)},$$

subject to

$$y^{(i)}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(i)} + b\right) \geq 1 - \xi^{(i)},\ i = 1, 2, \ldots, N.$$

Explain the role of the slack variable $\xi^{(i)}$ and the parameter C in controlling the trade-off between maximizing the margin and minimizing the classification error.

4) [CO2, L4] Explain the decision criteria of Naive Bayes Classifier and how it is calculated. (10)

5) [CO3, L4] (10)

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{\quad} \qquad \text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

$p(\mathbf{x})$ evidence

Choose the correct option in each of the following with a brief justification:

1) In a Naive Bayes classifier, a higher prior probability for a class results in.

More/Less likely to classify new data points as belonging to that class.
More/Less impact of the likelihood on the final decision.
More/Less biased predictions towards that class.
More/Less balanced classification across different classes.

2) In a Naive Bayes classifier, assuming independence between features leads to:

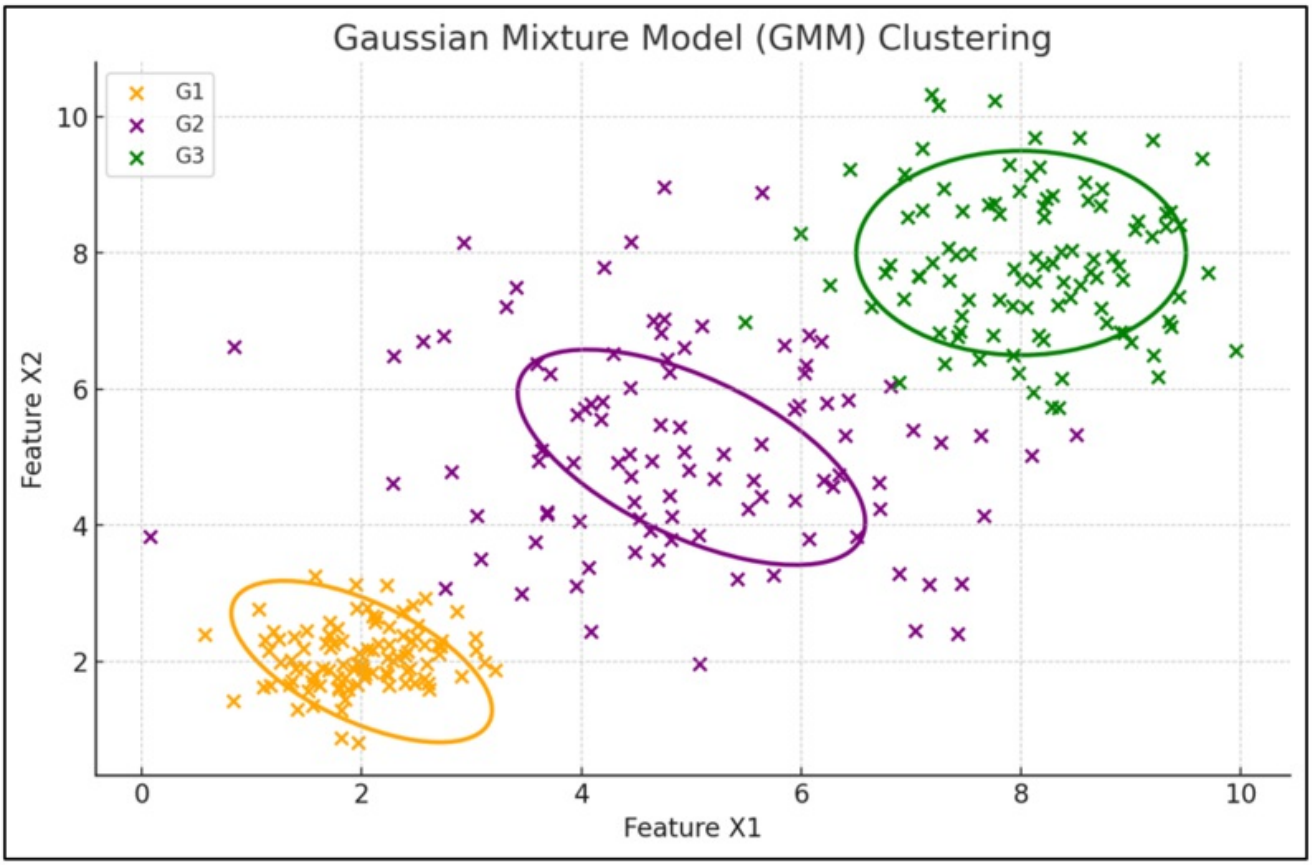
More/Less computational efficiency;
More/Less robust predictions when features are correlated.
More/Less complex model structure.

6) [CO3, L4] (10)

1) What are some advantages and disadvantages of using a Gaussian mixture model compared to other clustering algorithms like K-means?

2) In the Sklearn implementation of the Gaussian mixture model there are settings for Covariance_type ('full', 'tied', 'diag', 'spherical'). Explain any 2 of these

7) [CO3, L5] The following diagram represents a dataset plotted in a two-dimensional space, with points that seem to cluster into three distinct groups. A Gaussian Mixture Model (GMM) has been fitted to this data, resulting in three Gaussian distributions, each representing a cluster. (10)



(1) Which Gaussian (G1, G2, or G3) appears to have the highest variance? Justify how you can tell based on the diagram.

(2) If a new data point is added to the plot and falls within the overlap between G2 and G3, how does the GMM decide which cluster to assign the point to? Discuss the concept of "soft clustering" in your explanation.

8) [CO2, L4] (10)

1) Explain the efficiency of Gaussian mixture models in dealing with outliers.
2) How can we use Gaussian mixture models for generating data?

9) [CO3, L4] What is Laplace smoothing and why is it performed? Explain the effect of the "alpha" parameter. (10)

10) [CO4, L4] Explain in brief the difference between Principal Component Analysis and Independent Component Analysis. (10)

-----End-----