Department of Data Science
IIT Palakkad
DS5006 : Machine Learning (August 2025)

09:30-12:30 **End Semester** Marks : 45

Instructions

1. Write your answers neatly in Blue/ Black ink. Do not use pencil / Red ink. Make sure your answers are legible.
2. If you have to make any assumption about unspecified things, write the assumption clearly with justification.
3. Write the question number clearly for each answer. Draw a line after the answer.
4. There will be partial markings for the questions, so even if you are not able to solve the entire problem be sincere with the steps.
5. **Be precise.**

1. Solve the following constrained optimization problem: (3)

$$\text{maximize } f = -\sum_{i=1}^{N} P(x_i) log P(x_i)$$

$$\text{subject to} \sum_{i=1}^{N} P(x_i) = 1$$

2. (Source: B. Ravindran[1]) If a test of hypothesis has a Type I error probability (α) of 0.01, we mean: (2)

   a If the null hypothesis is true, we don't reject it 1% of the time.
   b If the null hypothesis is false, we don't reject it 1% of the time.
   c If the null hypothesis is true, we reject it 1% of the time.
   d If the null hypothesis is false, we reject it 1% of the time.

3. (Source: Tom Mitchell[2]) (4)

   Answer in one or two sentences:

   1. When would you prefer Decision Tree learning over Logistic Regression?
   2. When would you prefer Logistic Regression over Naive Bayes?

4. (Source: [Han et al., 2011]) (5)

   The data tuples of Table 1 are sorted by decreasing probability value, as returned by a classifier. For each tuple, compute the values for the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Compute the true positive rate (TPR) and false positive rate (FPR). Plot the ROC curve for the data.

---

[1] `https://dsai.iitm.ac.in/~ravi/`
[2] `https://www.cs.cmu.edu/~tom/`

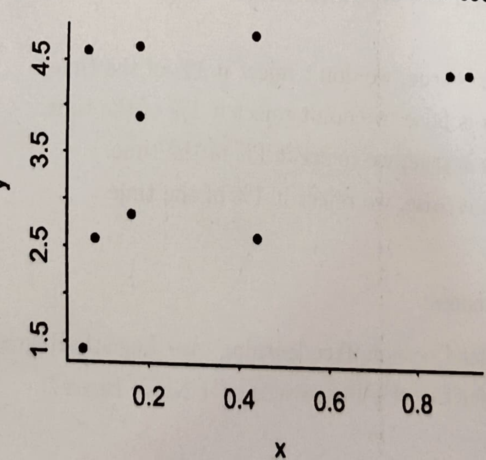
Tuple #	Class (True class)	Probability
1	p	0.95
2	n	0.85
3	p	0.78
4	p	0.66
5	n	0.60
6	p	0.55
7	n	0.53
8	n	0.52
9	n	0.51
10	p	0.4

Table 1: Tuples sorted by decreasing score, where the score is the value returned by a probabilistic classifier.

5. Discuss a real-world scenario in which the methodology of [Banerjee et al., 2025]: **Children's arithmetic skills do not transfer between applied and academic mathematics** can be adapted. (4)

6. (Source: [Gelman et al., 2021]) (5)

Influence of individual data points: A linear regression is fit to the data below. Which point has the most influence on the slope (give its approximate coordinate)?



7. What is boosting? Why is it likely to improve the performance of decision tree induction? (3)

8. You have a large amount of high-dimensional data, of which only 30% is labeled. How would you train a classifier using such data? (Assume that the labeled and unlabeled data come from the same source.) (5)

9. (Source: Eric Xing[3]) (5)

Consider the following statements about the EM algorithm in general. State if they are True (T) or False (F).

(a) The EM Algorithm maximizes a lower bound on the log likelihood at each iteration.

(b) The EM Algorithm minimizes an upper bound on the log likelihood at each iteration.

(c) The EM algorithm never decreases the log likelihood.

(d) EM is guaranteed to converge to a unique globally optimal solution.

(e) It is *impossible* to use gradient descent/ ascent to learn the parameters instead of EM.

10. (Source: [Han et al., 2011]) (4)

The task is to cluster points (with (x, y) representing location) into three clusters, where the points are —

$A_1(2, 10)$, $A_2(2, 5)$, $A_3(8, 4)$, $B_1(5, 8)$, $B_2(7, 5)$, $B_3(6, 4)$, $C_1(1, 2)$, $C_2(4, 9)$.

The distance function is Euclidean distance. Suppose initially we assign A_1, B_1, and C_1 as the center of each cluster, respectively. Use the k-means algorithm to show only—

Q1: the three cluster centers after the first round of execution.

Q2: the final three clusters.

11. (Source: Zichao Yang, 2014) (5)

Support Vector Machines are powerful tools for classification. SVM can perform non-linear classification using the kernel trick. In this question, you have to examine the decision boundaries of SVM with different kernels.

Figure 1 plots of SVM decision boundaries resulting from using different kernels and/or different slack penalties. In Figure 1, there are two classes of training data, with labels $y_i \in \{-1, 1\}$, represented by circles and squares respectively. The SOLID circles and squares represent the support vectors. The following are the different optimization strategies for identifying SVM-based decision boundaries. Identify a subplot in Figure 1 corresponding to each optimization strategy. Briefly explain rationale behind your decision. (Note that there are 6 plots, but only 5 problems, so one plot does not match any of the problems.)

Optimization strategy:

1. A soft-margin linear SVM with $C = 0.1$
2. A soft-margin linear SVM with $C = 10$
3. A hard-margin kernel SVM with $K(u, v) = u.v + (u.v)^2$
4. A hard-margin kernel SVM with $K(u, v) = \exp(-\frac{1}{4}||u - v||^2)$
5. A hard-margin kernel SVM with $K(u, v) = \exp(-4||u - v||^2)$

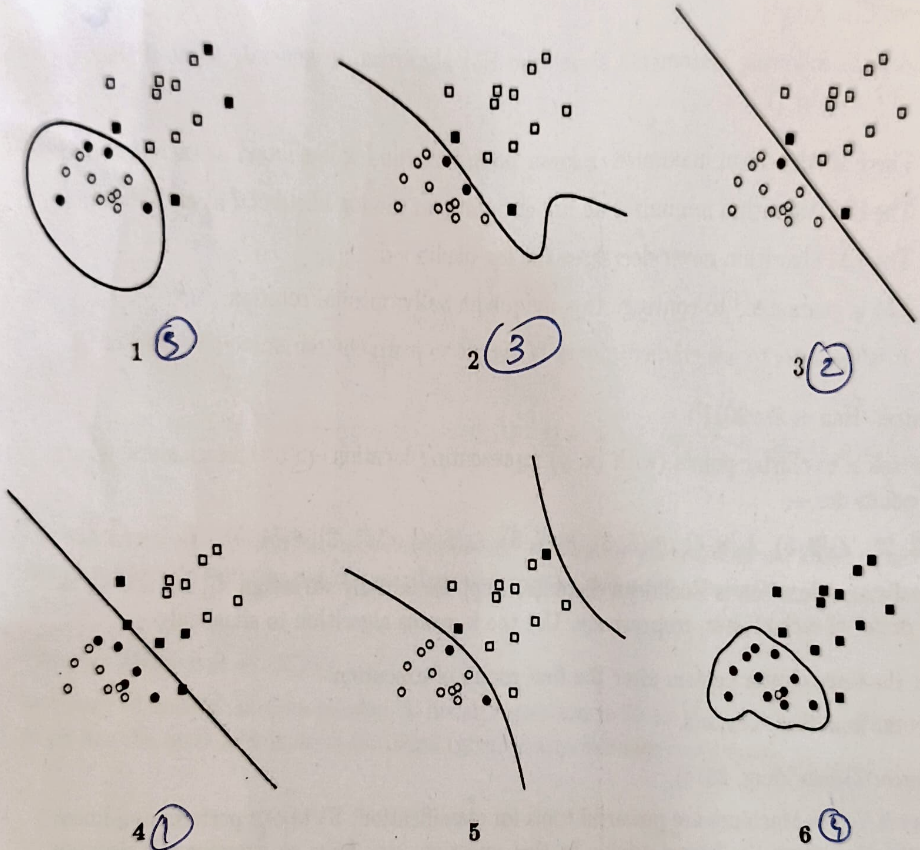


Figure 1: SVM Induced Decision Boundaries

END

References

[Banerjee et al., 2025] Banerjee, A. V., Bhattacharjee, S., Chattopadhyay, R., Duflo, E., Ganimian, A. J., Rajah, K., and Spelke, E. S. (2025). Children's arithmetic skills do not transfer between applied and academic mathematics. *Nature*, 639(8055):673–681.

[Gelman et al., 2021] Gelman, A., Hill, J., and Vehtari, A. (2021). *Regression and other stories*. Analytical methods for social research. Cambridge University Press, Cambridge.

[Han et al., 2011] Han, J., Kamber, M., and Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 3rd edition.

[3] `https://sailing-lab.github.io/`