Department of Data Science
IIT Palakkad
DS5006 : Machine Learning

0800-0850 | Test 2 (14 Oct 2024) | Marks : 15

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. Consider the following text classification tasks. Each classification task has two classes (*C1* vs *C2*) and the goal is to classify a document as either *C1* or *C2*: (5)

   1. *politics* vs *religion*
   2. *politics* vs *sports*
   3. *hardware* vs *software*
   4. *mathematics* vs *biology*

   Assume that there is no class imbalance. Explain which classification tasks are likely to have high performance and which tasks are likely to have low performance, irrespective of any specific classification approach.
   How can you verify your explanation, if you employ Naive Bayes and SVM?

2. Radial Basis Function (RBF) Kernel is defined as: $K(x, x') = exp(-\gamma||x - x'||^2)$. (4)

   How does the γ in RBF Kernel influence the resulting decision boundary?
   What could be a possible relation between the parameter γ in RBF kernel and parameter k in k-nearest neighbor classification?

3. Let. (2)

$$g(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

   Prove: $g'(x) = 1 - g(x)^2$

4. A company wants to automate its recruitment process and is planning to deploy an intelligent bot developed using the recent advances in AI. The goal of the bot is to screen and interview candidates solely on the basis of their resumes. (4)

   How will you design experiments to evaluate the effectiveness of the bot? What inputs should the bot provide to facilitate the experiments? What should be the characteristics of an evaluation metric in this case?