DS4023 Machine Learning

SVM Exercise1

Q1. What linear function is used by a SVM for classification? How is an input vector $\mathbf{x_i}$(instance) assigned to the positive or negative class?

Q2. If the training examples are linearly separable, how many decision boundaries can separate positive from negative data points? Which decision boundary does the SVM algorithm calculate? Why?

Q3. Use Lagrange multiplier method to answer the following questions.

3.1 Consider the Entropy definition**:**

- If we are given a probability distribution $P = (p_1, p_2, \ldots, p_n)$, then the information conveyed by this distribution, also called the Entropy of P, is $I(P) = -(p_1 \times \log p_1 + p_2 \log p_2 + \cdots + p_n \log p_n)$ (The base of the logarithm is 2)

What is the range of entropy of P and which distribution gives maximum entropy? Show the details of your answer.

3.2 Given the probability distribution $P = (p_1, p_2, \ldots, p_n)$, Gini index is another way to measure the uncertainty $Gini(P) = 1 - \sum_i^n {p_i}^2$.

What is the range of Gini index and which distribution gives maximum

Gini index value? Show the details of your answer.

Q4. Given the SVM optimization problem:

$$\min_{\mathrm{w},b} \frac{1}{2}\|\mathrm{w}\|^2$$
$$s.t. \quad y_i(\mathrm{w}^T \mathrm{x}_i + b) \geq 1, \qquad i = 1,2,\dots,m$$

Derive the dual optimization problem and show the detail steps.

Q5. For a solution $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_m^*)^T$ for the dual problem, if exist α_j^* that $\alpha_j^* > 0$, then the solution for primal problem is:

- $w^* = \sum_{i=1}^{m} \alpha_i^* y_i \mathbf{x}_i$
- $b^* = y_j - \sum_{i=1}^{m} \alpha_i^* y_i \mathbf{x}_i{}^T \mathbf{x}_j$

Show the detail steps for deriving the solution given α^*.