MDSC - 301

Sri Sathya Sai Institute of Higher Learning

(Deemed to be University)

Department of Mathematics and Computer Science

Muddenahalli Campus

Course: **M.Sc. Data Science and Computing** Date: **September 10, 2022**

Subject: **Machine Learning** Module : **Linear Models for Classification**

Answer the following:

(1) Write the objective function for the primal optimization problem for maximal margin classifier. Define the terms.

(2) State the optimization problem solved to train the soft margin vector machine. Define each term.

(3) Assume that we have hyperplane in separating two class using SVM is $\mathbf{w}^T\mathbf{x} + b = 0$. Answer the following:

(a) Let $\mathbf{x}_j$ be a support vector in a hard margin SVM then what is the value of $|\mathbf{w}^T\mathbf{x}_j + b|$

(b) Suppose $(\mathbf{x}_j, y_j)$ are the support vectors and λ_js the corresponding Lagrange multiplies, the what is the value of $\mathbf{w}$?

(c) Suppose $(\mathbf{x}_j, y_j)$ are the support vectors and λ_js the corresponding Lagrange multiplies, the what is the value of $\sum_j \lambda_j y_j$?

(4) Consider the following training data:

X	(1,0)	(0,1)	(0,-1)	(-1,0)	(0,2)	(0,-2)	(-2,0)
y	-1	-1	-1	+1	+1	+1	+1

We perform the following non-linear of the input vector $\mathbf{x} = (x_1, x_2)^T$ to obtain the transformed feature space $\mathbf{z} = (\phi_1(x), \phi_2(x))^T$, with $\phi_1(\mathbf{x}) = x_2^2 - 2x_1 + 3,\ \phi_2(\mathbf{x}) = x_1^2 - 2x_2 - 3$.

Write the equation of the optimal separating hyperplane in transformed space $\mathbf{z}$. Explain your answer.

(5) A linear with two-dimensional input was trained using the following training set.

feature space	labels
(1,1)	1
(1,0)	1
(0,1)	1
(0,0.5)	-1

The following Lagrange multipliers were obtained after solving Quadratic Programming problem:

$$\lambda_1 = 0, \lambda_2 = 2, \lambda_3 = 0, \lambda_4 = 2.$$

Find the equation of the optimal separating hyperplane.

(6) Consider the following 2-class (C_1, C_2) classification problem involving a single feature x. Assume equal class priors. The class conditional distributions are:

$$p(x|C_1) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad p(x|C_2) = \begin{cases} 2-2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Derive the Bayes decision boundary?

(b) What is the new decision boundary if the apriori probability of class C_1 is changed to 0.7?

(7) Consider two classes C_1 and C_2. We want to classify a variable X into one of these two classes. Suppose $p(X|C_1)$ and $p(X|C_2)$ are defined as follows:

$$p(X|C_1) \sim N(0,1)$$
$$p(X|C_2) = \frac{1}{4}, \ -2 < X < 2$$

(a) Derive the Bayes decision boundary?

(b) Find the maximum error classification rule $g(x)$ for this two-class problem, assuming $p(C_1) = p(C_2) = 0.5$.

(8) Consider a SVM whose input space in $\mathbb{R}^2$, and in which the kernel function is computed as $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T\mathbf{y}+1)^2 - 1$. Find the $\phi(x)$ to the feature space corresponding to this kernel. Show your derivation.

(9) Consider the following data:

feature space	labels
(1,-1,-1)	-1
(-3,1,1)	1
(-3,1,-1)	-1
(1,2,1)	-1
(-1,-1,2)	1

These points are separable. Derive the optimum margin classifier and the margin.

(10) A hypothetical SVM model has the following values of Lagrange multipliers λ and support vectors:

α	Support Vectors	labels
1	(1,-1,1)	+1
0.5	(0,2,-1)	-1
1	(-1,0,2)	-1

Suppose that the linear kernel used. Compute the output y of this SVM model when the input feature vector is $(0.3, 0.8, 0.6)$.

(11) Define the convex function. Let f and g are two convex functions. Prove that $f + g$ is convex.

(12) Prove that the following functions are convex.

(a) $f(x) = -x \log x$

(b) $f(x) = x^2$

(Hint: Use the fact that the double derivative of convex functions assumes non-negative values)

(13) Define the loss function. How it will be helpful in Machine Learning algorithms.

(14) Which loss functions you prefer in Linear models for classification. Explain.

(15) Define Hinge loss. Where and how it is useful to train the data.

$$- - \times \times \times - -$$