AI21MTECH14011

Foundations of Machine Learning
Assignment - 2

Q.1] SVM:

1. In the expression $W^T x + b$, W and b are unknowns.
For Ex. $10x + 100 = 1000$, $\frac{x}{100} + \frac{1}{10} = 1$
both are same lines. Since W and b are unknowns fixed right side to $+1$ & -1, rescale $W \to kW$, $b \to kb$ according to that.
2. Making $+1$ & -1 just for our convenience. It makes calculations easier.
In this way, if margin boundaries replaced by any arbitrary constants, solution is unchanged.

Q.3) Kernels.

a) $k(x,z) = k_1(x,z) + k_2(x,z)$

$\Rightarrow k(x,z) = \alpha k_1(x,z) + \beta k_2(x,z)$. for $\alpha, \beta \geq 0$
here $\alpha = \beta = 1$

$k_1(x,z) = (\sqrt{1}\,\Phi_1(x), \sqrt{1}\,\Phi_1(z))$, $k_2(x,z) = (\sqrt{1}\,\Phi_2(x), \sqrt{1}\,\Phi_2(z))$

$$\Rightarrow k(x,z) = k_1(x,z) + k_2(x,z)$$
$$= (\sqrt{1}\,\Phi_1(x), \sqrt{1}\,\Phi_1(z)) + (\sqrt{1}\,\Phi_2(x), \sqrt{1}\,\Phi_2(z))$$
$$= [\sqrt{1}\,\Phi_1(x), \sqrt{1}\,\Phi_2(x)], [\sqrt{1}\,\Phi_1(z), \sqrt{1}\,\Phi_2(z)]$$

So, $k(x,z)$ can be expressed as inner product.
Hence it is valid.

b) $k(x,z) = k_1(x,z) \cdot k_2(x,z)$

k is Hadamard product i.e. element-by-element multiplication of k_1 and $k_2 \Rightarrow (k = k_1 \odot k_2)$.

k_1 and k_2 are covariance matrices of let $(a_1 a_2 \dots a_n)$ and $(b_1 b_2 \dots b_n)$. Then k simply covariance matrix of $(a_1 b_1, a_2 b_2, \dots a_n b_n)$.

Hence it is valid kernel

c) $k(x,z) = h(k_1(x,z))$ ... $h \Rightarrow$ poly. functn with +ve coefficients

Polynomial term is a product of kernels with positive coefficients

$\therefore k(x,z) = h k_1(x,z) + h k_2(x,z)$ $-\mathrm{I}$ .. from first

$k(x,z) = k_1(x,z) \cdot k_2(x,z)$ $-\mathrm{II}$ ... from second.

From I & II

$$k(x,z) = h(k_1(x,z))$$

Hence it is a valid kernel.

d) $k(x,z) = \exp(k_1(x,z))$

Using taylor series, $\exp(x) = \lim_{n \to \infty} \left(1 + x + \dots \frac{x^n}{n!}\right)$

$$k(x,z) = h(k_1(x,z)) \Rightarrow k(x,z) = \lim_{n \to \infty} k_i(x,z)$$

Implies it is a valid kernel.

e) $k(x,z) = \exp\left(\frac{-\|x - z\|_2^2}{\sigma^2}\right)$

$$= \exp\left(\frac{-\|x\|^2 - \|z\|^2 + 2x^T z}{\sigma^2}\right)$$

$$= \left(\exp\left(\frac{-\|x\|^2}{\sigma^2}\right) \exp\left(\frac{-\|z\|^2}{\sigma^2}\right)\right) \exp\left(\frac{2x^T z}{\sigma^2}\right)$$

$$= f(x)\, f(z)\, \exp(k_1(x,z))$$

$\underbrace{f(x)\,f(z)}_{\text{kernel}}$ is one kernel $\exp(k_1(x,z))$ is one kernel and hence product of two kernels is a valid kernel.

Q.2) $\rho = \dfrac{1}{\|w\|}$, S.T. $\dfrac{1}{\rho^2} = \sum_{i=1}^{N} \alpha_i$

$$\Rightarrow \rho^2 = \frac{1}{\|w\|^2}$$

$$\therefore \frac{1}{\rho^2} = \|w\|^2 \quad \text{—I}$$

From SVM hard-margin duality, $\bar{w} = \sum_{i=1}^{n} \bar{\alpha}_i y_i x_i$ —II

Here

Lagrangian, $L(w,\alpha) = \|w\|_2^2 + \sum_{i=1}^{n} \alpha_i (1 - y_i x_i^T w)$.

Primal maximum margin was.

$$P(w) = \sup_{\alpha \geq 0} L(w,\alpha) = \sup_{\alpha \geq 0} \left[\|w\|_2^2 + \sum_{i=1}^{n} \alpha_i (1 - y_i x_i^T w) \right]$$

Dual is given by

$$D(\alpha) = \min_{w \in R^d} L(w,\alpha) = \sum_{i=1}^{n} \alpha_i - \left\| \sum_{i=1}^{n} \alpha_i y_i x_i \right\|_2^2$$

For max. margin SVM, $D(\alpha) = 0$

$$\therefore \sum_{i=1}^{n} \alpha_i = \left\| \sum_{i=1}^{n} \alpha_i y_i x_i \right\|_2^2$$

$$\Rightarrow \sum_{i=1}^{n} \alpha_i = \|\bar{w}\|_2^2 \quad \text{... From II}$$

$$\Rightarrow \sum_{i=1}^{n} \alpha_i = \frac{1}{\rho^2} \quad \text{... From I}$$

Hence, $\dfrac{1}{\rho^2} = \sum_{i=1}^{N} \alpha_i$ ... thus proved.

Programming Questions

Q.4)

Ans. In this question, initially training and testing data are loaded, then they are filtered according to one-versus-one classifier for the digits ‘1’ (+1) and ‘5’ (-1). Values are stored as features and labels. Data is trained and tested using different kernel implementation techniques.

a) Model is train and tested using Linear Kernel method. Overall accuracy calculated over entire test set is 0.97877 or 97.877%. Total no. of support vectors calculated are 28.

b) Model is train and tested using Linear Kernel method for first given points.
The given number of samples, their respective test accuracy and their respective support vectors are given below.

Samples	Test Accuracy	Total Support Vectors
50	0.98	2
100	0.99	4
200	0.99	8
800	0.981	14

c) Model is train and tested using Polynomial Kernel method.
True or False.

i. False.
At C=0.0001, For Q=2, training error is 0.02242 and for Q=5, it is 0.00641. Hence, this statement is False.

ii. True.
At C=0.001, For Q=2, total number of support vectors are 152 and for Q=5, they are 28. Hence, this statement is True.

iii. False.
At C=0.01, For Q=2, training error is 0.00448 and for Q=5, it is also 0.00448. Hence, this statement is False.

iv. True.
At C=1, For Q=2, test error is 0.01887 and for Q=5, it is 0.01651. Hence, this statement is True.

d) Model is train and tested using RBF Kernel method.
Training and Testing error values for given values of C are listed below.

C-Value	Training Error	Testing Error
0.01	0.00384	0.02123
1	0.00448	0.02123
100	0.0032	0.01887
10000	0.00256	0.01887
1000000	0.00128	0.02123

Lowest training error results for C=10^6, which is 0.00128. Lowest testing error results for C=100 and C=10000, which is 0.01887.

Q.5)

Ans. In this question, initially training and validation data are loaded. Values are stored as features and labels, converted into arrays. Data is trained and tested using different kernel implementation techniques.

a) Model is train and tested using Linear Kernel method. Training error calculated as 0, Testing error as 0.024 or 2.4%. Total number of support vectors calculated are 1084.
b) In the first part, Model is train and tested using RBF Kernel method. Training error calculated as 0, Testing error as 0.5 or 50%. Total number of support vectors calculated are 6000.
   In the second part, Model is train and tested using Polynomial Kernel method. Training error calculated as 0, Testing error as 0.021 or 2.1%. Total number of support vectors calculated are 1755.

   All Kernels have lowest training error i.e., 0.