Assignment 1

Dishank Jain
AI20BTECH11011

1 (4 marks)

In the derivation for the Support vector machine, we assumed that the margin boundaries are given by $\vec{w}.\vec{x} + b = +1$ and $\vec{w}.\vec{x} + b = -1$. Show that, if the $+1$ and -1 are replaced by some arbitrary constants $+\gamma$ and $-\gamma$ for some $\gamma > 0$, the solution for the maximum margin hyperplane is unchanged.

1.1 Solution

In the above situation, for the positive and negative support vectors, we will have,

$$\vec{w}.\vec{x} + b = +\gamma;\ \vec{w}.\vec{x} + b = -\gamma \tag{1}$$

Hence the margin is given by $\dfrac{2\gamma}{\|\vec{w}\|}$. To maximize the margin, we have to maximize $\dfrac{2\gamma}{\|\vec{w}\|}$ or alternately minimize $\dfrac{\|\vec{w}\|^2}{2\gamma^2}$ subject to $y_i(\vec{w}.\vec{x} + b) - \gamma \geq 0$. We can use method of Lagrange multipliers and obtain the Lagrangian and the Karush-Kuhn-Tucker conditions as follows.

$$L(\vec{x}, \vec{\eta}) = \frac{\|\vec{w}\|^2}{2\gamma^2} - \sum_i \eta_i(y_i(\vec{w}.\vec{x}_i + b) - \gamma) \tag{2}$$

$$\vec{w} = \gamma^2 \sum_j \eta_j y_j \vec{x}_j \tag{3}$$

$$\sum_j \eta_j y_j = 0 \tag{4}$$

We have used η instead of α to keep the notations clear. α is still being used for the solution of the original maximization problem. The Lagrangian will be obtained as

$$\gamma \sum_j \eta_j - \frac{\gamma^2}{2} \sum_{i,j} y_i y_j \eta_i \eta_j (\vec{x}_i.\vec{x}_j)$$

Originally, the Lagrangian was

$$\sum_j \alpha_j - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j (\vec{x}_i.\vec{x}_j)$$

Therefore, we obtain

$$\eta_i = \frac{\alpha_i}{\gamma} \tag{5}$$

Substituting in equation 2, we obtain

$$\vec{w} = \gamma \sum_j \alpha_j y_j \vec{x}_j \tag{6}$$

Now, b is given by

$$b = -y_i(y_i(\vec{x}_i.\vec{w}) - \gamma) = -\gamma y_i(y_i(\vec{x}_i.\sum_j \alpha_j y_j \vec{x}_j) - 1)) \tag{7}$$

Hence we note that if the original equation of the hyperplane was $\vec{W}.\vec{x} + B = 0$, then $\vec{w} = \gamma\vec{W}$ and $b = \gamma B$. Therefore the new equation of the hyperplane is $\gamma\vec{W}.\vec{x} + \gamma B = 0$. On dividing by γ on both sides, we get $\vec{W}.\vec{x} + B = 0$ which is same as original equation. Thus the solution of the maximum margin hyperplane is unchanged.

2 (4 marks)

Consider the half margin of the maximum-margin SVM defined by ρ, i.e. $\rho = \frac{1}{\|\vec{w}\|}$. Show that ρ is given by

$$\frac{1}{\rho^2} = \sum_i \alpha_i \tag{8}$$

2.1 Solution

$$\|\vec{w}\|^2 = \vec{w}.\vec{w} \tag{9}$$

$$= \sum_j \alpha_j y_j \vec{w}.\vec{x}_j \tag{10}$$

α_j will be non-zero only when $\vec{x}_j$ is a support vector. Therefore, in equation 9,

$$y_j \vec{w}_j.\vec{x}_j = 1 - b y_j \tag{11}$$

$$\implies \|\vec{w}\|^2 = \sum_j \alpha_j - \sum_j \alpha_j (b y_j) \tag{12}$$

Using second KKT condition

$$\|\vec{w}\|^2 = \sum_j \alpha_j \tag{13}$$

$$\implies \frac{1}{\rho^2} = \sum_i \alpha_i \tag{14}$$

3 (5 marks)

Let k_1 and k_2 be valid kernel functions. Comment about the validity of the following kernel functions.

(a) $k(\vec{x}, \vec{z}) = k_1(\vec{x}, \vec{z}) + k_2(\vec{x}, \vec{z})$

(b) $k(\vec{x}, \vec{z}) = k_1(\vec{x}, \vec{z})k_2(\vec{x}, \vec{z})$

(c) $k(\vec{x}, \vec{z}) = h(k_1(\vec{x}, \vec{z}))$ where h is a polynomial function with positive coefficients.

(d) $exp(k_1(\vec{x}, \vec{z}))$

(e) $exp(\frac{-\|\vec{x} - \vec{z}\|^2}{\sigma^2})$

3.1 Solution

(a) The kernel $k(\vec{x}, \vec{z}) = k_1(\vec{x}, \vec{z}) + k_2(\vec{x}, \vec{z})$ is a valid kernel. Since $k_1(\vec{x}, \vec{z})$ and $k_2(\vec{x}, \vec{z})$ are valid kernels, Mercer's condition holds for them, i.e. for all square integrable functions $f(\vec{x})$,

$$\int\int K_1(\vec{x}, \vec{z})f(\vec{x})f(\vec{z})dxdy \geq 0 \tag{15}$$

$$\int\int K_2(\vec{x}, \vec{z})f(\vec{x})f(\vec{z})dxdy \geq 0 \tag{16}$$

Using equations 15 and 16, we can say

$$\int\int (K_1(\vec{x}, \vec{z}) + k_2(\vec{x}, \vec{z}))f(\vec{x})f(\vec{z})dxdy \geq 0 \tag{17}$$

$$\implies \int\int K(\vec{x}, \vec{z})f(\vec{x})f(\vec{z})dxdy \geq 0 \tag{18}$$

Therefore the given kernel satisfies Mercer's condition. Thus it is a valid kernel.

(b) The given kernel is a valid kernel. Since $k_1(\vec{x}, \vec{z})$ and $k_2(\vec{x}, \vec{z})$ are valid kernels, there exist some functions $\phi_1(\vec{x})$ and $\phi_2(\vec{x})$ such that

$$k_1(\vec{x}, \vec{z}) = \phi_1(\vec{x}).\phi_1(\vec{z});\ k_2(\vec{x}, \vec{z}) = \phi_2(\vec{x}).\phi_2(\vec{z}) \tag{19}$$

$$Let\ \phi_1(\vec{x}) = \begin{pmatrix} a_1(\vec{x}) \\ a_2(\vec{x}) \\ : \\ : \\ a_m(\vec{x}) \end{pmatrix};\ \phi_2(\vec{x}) = \begin{pmatrix} b_1(\vec{x}) \\ b_2(\vec{x}) \\ : \\ : \\ b_n(\vec{x}) \end{pmatrix} \tag{20}$$

$$\implies k_1(\vec{x},\vec{z}) = \sum_{i=1}^{m} a_i(\vec{x})a_i(\vec{z});\ k_2(\vec{x},\vec{z}) = \sum_{j=1}^{n} b_j(\vec{x})b_j(\vec{z}) \tag{21}$$

$$\implies k(\vec{x},\vec{z}) = \left(\sum_{i=1}^{m} a_i(\vec{x})a_i(\vec{z})\right)\left(\sum_{j=1}^{n} b_j(\vec{x})b_j(\vec{z})\right) \tag{22}$$

$$\implies k(\vec{x},\vec{z}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i(\vec{x})a_i(\vec{z})b_j(\vec{x})b_j(\vec{z}) \tag{23}$$

$$\implies k(\vec{x},\vec{z}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i(\vec{x})b_j(\vec{x})a_i(\vec{z})b_j(\vec{z}) \tag{24}$$

$$\implies k(\vec{x},\vec{z}) = \sum_{k=1}^{mn} c_k(\vec{x})c_k(\vec{z}) \tag{25}$$

$$where\ c_k(\vec{x}) = a_p(\vec{x})b_q(\vec{x}) \tag{26}$$

$$where\ p = \left\lceil \frac{k}{n} \right\rceil;\ q = k + 1 - \left\lceil \frac{k}{n} \right\rceil \tag{27}$$

Thus the given kernel is a valid kernel

(c) Note that following the same argument as in option (a), we can conclude that the kernel $k(\vec{x},\vec{z}) = ak_1(\vec{x},\vec{z}) + bk_2(\vec{x},\vec{z})$ is a valid kernel for any $a, b > 0$. Also, using (b), we can conclude that the kernel $k(\vec{x},\vec{z}) = k_1^n(\vec{x},\vec{z})$ is a valid kernel for any $n \in \mathcal{N}$. Using these two conclusions, we can satisfactorily conclude that the kernel $h(k_1(\vec{x},\vec{z}))$ is a valid kernel.

(d) The i^{th} co-efficient of the taylor series expansion of e^x is $\dfrac{1}{i!}$. Thus we observe that $k(\vec{x},\vec{z}) = e^{k_1(\vec{x},\vec{z})}$ is nothing but a polynomial in $k_1(\vec{x},\vec{z})$ with positive co-efficients which extends to infinity. Thus using our result from option (c), we conclude that $e^{k_1(\vec{x},\vec{z})}$ is a valid kernel.

(e)

$$e^{-\dfrac{\|\vec{x}-\vec{z}\|^2}{\sigma^2}} = exp\left(-\frac{\|\vec{x}\|^2}{\sigma^2}\right) exp\left(\frac{2\vec{x}.\vec{z}}{\sigma^2}\right) exp\left(-\frac{\|\vec{z}\|^2}{\sigma^2}\right) \tag{28}$$

We know that $\vec{x}.\vec{z}$ is a valid kernel. $\dfrac{2}{\sigma^2}$ is greater than 0. Thus using option (c), $\dfrac{2\vec{x}.\vec{z}}{\sigma^2}$ is a valid kernel. Further, using option (d) $exp\left(\dfrac{2\vec{x}.\vec{z}}{\sigma^2}\right)$

is a valid kernel. Thus for some $\phi(\vec{x})$, we can write

$$exp\left(\frac{2\vec{x}.\vec{z}}{\sigma^2}\right) = \phi(\vec{x}).\phi(\vec{z}) \tag{29}$$

Substituting this in 28, we get

$$e^{-\dfrac{\|\vec{x}-\vec{z}\|^2}{\sigma^2}} = exp\left(-\frac{\|\vec{x}\|^2}{\sigma^2}\right)\phi(\vec{x}).\phi(\vec{z})exp\left(-\frac{\|\vec{z}\|^2}{\sigma^2}\right) \tag{30}$$

We define $\Phi(\vec{x}) = exp\left(-\frac{\|\vec{x}\|^2}{\sigma^2}\right)\phi(\vec{x})$. Using this, we can write

$$e^{-\dfrac{\|\vec{x}-\vec{z}\|^2}{\sigma^2}} = \Phi(\vec{x}).\Phi(\vec{z}) \tag{31}$$

Thus $e^{-\dfrac{\|\vec{x}-\vec{z}\|^2}{\sigma^2}}$ is a valid kernel.

4 (10 marks)

4.1 Solution

(a) Testing Accuracy = 0.9811, Number of support vectors = 36

(b)
- For first 50 records, Testing Accuracy = 0.9741, Number of support vectors = 5
- For first 100 records, Testing Accuracy = 0.9764, Number of support vectors = 6
- For first 200 records, Testing Accuracy = 0.9811, Number of support vectors = 12
- For first 800 records, Testing Accuracy = 0.9811, Number of support vectors = 21

(c)
- **False**
- **False**
- **True**
- **True**

(d)
- For C = 0.01, Training error = 0.0045, Testing error = 0.0165
- For C = 1, Training error = 0.0032, Testing error = 0.0189
- For C = 100, Training error = 0.0013, Testing error = 0.04
- For C = 10000, Training error = 0, Testing error = 0.3656
- For C = 1000000, Training error = 0, Testing error = 0.3774
- C = 1000000 and C = 10000 result in lowest Training error
- C = 0.01 results in lowest Testing error

5 (7 marks)

5.1 Solution

(a) For linear kernel, Training error = 0.0, Testing error = 0.0240, Number of support vectors = 1084

(b)
- For rbf kernel, Training error = 0.0, Testing error = 0.5 , Number of support vectors = 6000
- For polynomial kernel, Training error = 0.0, Testing error = 0.0210, Number of support vectors = 1755
- Both kernels yield same Training error = 0.0