Foundations of Machine Learning

Assignment-2

ES19BTECH11003
Hemanth. K

2) Given,

to consider half Margin of maximum-margin SVM defined by ρ, i.e, $\rho = \frac{1}{\|w\|}$

to show :- $\frac{1}{\rho^2} = \sum_{i=1}^{N} \alpha_i$, α_i = Lagrange Multiplier

As given $\rho = \frac{1}{\|w\|}$

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

$\Rightarrow \boxed{\frac{1}{\rho^2} = w \cdot w}$ — ①

we already know, $w = \sum \alpha_i x_i y_i$. As y_i, α_i are Scalars x_i is vector

$w = \sum y_i \alpha_i x_i$

$\boxed{w \cdot w = \sum y_i \alpha_i (x_i \cdot w)}$ — ②

as ~~xxxx~~ . we know that $y_i(x_i \cdot w + b) = 1$

$x_i \cdot w + b = \frac{1}{y_i}$, $\boxed{x_i w = \frac{1}{y_i} - b}$ — ③

Substitute eq-③ for w·w (eq-②)

i.e, $w \cdot w = \sum \alpha_i y_i \left(\frac{1}{y_i} - b\right) = \sum(\alpha_i - b\alpha_i y_i)$

$w \cdot w = \sum \alpha_i - b \sum \alpha_i y_i$

as $\alpha_i y_i$ Scalar so $\sum_{i=1}^{N} \alpha_i y_i = 0$

$$\Rightarrow w \cdot w = \sum_{i=1}^{N} \alpha_i$$

ie, $\boxed{\frac{1}{\rho^2} = \sum_{i=1}^{N} \alpha_i}$ (from eq ①)

1) we assumed, margin boundaries given by $w \cdot x + b = +1$, $w x + b = -1$

to show:- if $+1$, & -1 are replaced by $+\gamma$ and $-\gamma$ solution for maximum margin hyperplane is unchanged.

As we know original margin boundaries $w_0 x + b_0 = 1$

$w_0 x + b_0 = -1$

and $+1$ & -1 are replaced by $+\gamma$ & $-\gamma$

So $w_1 x + b_1 = \gamma$

$w_1 . x + b_1 = -\gamma$ ①

so now new L_p will be, $L_p = \frac{1}{2}\frac{\|w_1\|^2}{\gamma} - \sum \alpha_i [y_i (w_1 x_i + b_1) - \gamma]$ ①

$$\left(\frac{\partial L_p}{\partial w}\right)_{replaced} = \frac{\|w_1\|}{\gamma} - \sum \alpha_i x_i y_i$$ (so as to for L_p to be min)

$$0 = \frac{\|w_1\|}{\gamma} - \sum \alpha_i x_i y_i$$

$$\frac{\|w_1\|}{\gamma} = \sum \alpha_i x_i y_i \quad — ②$$

from ①

$$\left(\frac{\partial L_p}{\partial b}\right)_{rep} = 0 - \sum \alpha_i y_i$$

$$\boxed{\sum \alpha_i y_i = 0} \quad — ③$$

Scanned with CamScanner

As we know as per original conditions

$$\left(\frac{\partial L_p}{\partial w}\right)_{original} = ||w_0|| - \Sigma \alpha_i x_i y_i$$

$$\boxed{\therefore ||w_0|| = \Sigma \alpha_i x_i y_i} \quad - (4)$$

$$\left(\frac{\partial L}{\partial b}\right)_{original} = 0 - \Sigma \alpha_i y_i \quad (\text{from eq } (1))$$

(2)/(4) (dividing eq (2) by eq (4))

$$\frac{||w_1||}{\gamma ||w_0||} = \frac{\cancel{\Sigma \alpha_i x_i y_i}}{\cancel{\Sigma \alpha_i x_i y_i}} = 1$$

$$||w_1|| = \gamma ||w_0|| \quad - (5)$$

from question As already mentioned that is a hard Margin SVM

so for SV's $\quad y_i(w_0 x_i + b_0) = 1 \quad - (6)$

$\quad y_i(w_1 x_i + b_1) = \gamma \quad - (7)$

dividing eq's (6) & (7)

$$\frac{y_i(w_0 x_i + b_0)}{y_i(w_1 x_i + b_1)} = \frac{1}{\gamma}$$

$$\gamma(w_0 x_i + b_0) = w_1 x_i + b_1 \quad \text{from eq (5)} \quad w_1 = \gamma w_0$$

$$\cancel{\gamma w_0 x_i} + \gamma b_0 = \cancel{\gamma w_0 x_i} + b_1$$

$$\boxed{b_1 = \gamma b_0} \quad - (8)$$

Now substituting eq (5) & (8) in replaced Margin's

$w_1 x_i + b_1 = \gamma$	$w_1 x_i + b_1 = -\gamma$
$\gamma w_0 x_i + \gamma b_0 = \gamma$	$\gamma w_0 x_i + \gamma b_0 = -\gamma$
$\cancel{\gamma}(w_0 x_i + b_0) = \cancel{\gamma}$	$\cancel{\gamma}(w_0 x_i + b_0) = -\cancel{\gamma}$

$w_0 x + b_0 = 1$ & $w_0 x + b_0 = -1$

So we can clearly see that if even original margins are replaced by γ & $-\gamma$, the maximum margin hyperplane will be unchanged.

3) Given, k_1, k_2 valid kernels, comment about validity of following kernel functions

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

Lets assume $k_1(x,z) = y_1(x) \cdot y_1(z)$

$k_2(x,z) = y_2(x) \cdot y_2(z)$

$$k(x,z) = y_1(x) \cdot y_1(z) + y_2(x) \cdot y_2(z)$$

$$= (y_1(x), y_2(x)) \cdot (y_1(z), y_2(z)) \quad \text{(dot product)}$$

$$k(x,z) = y'(x) \cdot y'(z)$$

Hence, $k(x,z)$ is a valid kernel function

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

$$= y_1(x) y_1(z) \cdot y_2(x) y_2(z)$$

if we consider $y_1(x) = l_1, l_2, \ldots$ & $y_1(z) = m_1, m_2, \ldots$

similarly $y_2(x) = n_1, n_2, \ldots$ & $y_2(z) = o_1, o_2, \ldots$

Hence, $k(x,z) = \sum_{i,j} l_i m_i n_j o_j$

Scanned with CamScanner

So the feature map will be,

$$\phi(x,z) = \sum_{i,j} l_i m_i n_j o_j$$

$\therefore$ $K(x,z)$ is also a valid kernel

c) $K(x,z) = h(K_1(x,z))$ Given, h is a polynomial function with all +ve coefficients

In this, we have to prove that summation over polynomial functions is valid

as already we have proven that kernel function is valid (in ⓐ) for summation & also for polynomial function (in ⓑ)

So even $K(x,z)$ is also valid for summation over polynomial functions.

d) $K(x,z) = e^{(K_1(x,z))}$

as we know e^x expansion $= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} - - - -$ (from taylor series)

$$e^{(K_1(x,z))} = 1 + K_1(x,z) + \frac{K_1(x,z)^2}{2!} + \frac{K_1(x,z)^3}{3!} + - - -$$

as from ⓐ, ⓑ & ⓒ we have proven that kernel function is valid for polynomials, summation

So $K(x,z)$ is also valid for exponential function as well

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

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

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

$$= \underbrace{e^{\frac{-|x|^2}{\sigma^2}}}_{①} \cdot \underbrace{e^{\frac{-|z|^2}{\sigma^2}}}_{②} \cdot \underbrace{e^{\frac{2|x||z|}{\sigma^2}}}_{③}$$

① → this is valid from d

② → this is also valid from d

③ → this is also valid from b & d

∴ Hence $K(x,z)$ is also valid