CS771- Assignment 2

Shubham Jain - 13683

1)
Standard formulation
Batch number 0 as the test data the accuracy is: 0.8454
Batch number 1 as the test data the accuracy is: 0.8687
Batch number 2 as the test data the accuracy is: 0.8618
Batch number 3 as the test data the accuracy is: 0.8636
Batch number 4 as the test data the accuracy is: 0.8653

Hinge Loss
Batch number 0 as the test data the accuracy is: 0.9674
Batch number 1 as the test data the accuracy is: 0.9983
Batch number 2 as the test data the accuracy is: 0.9845
Batch number 3 as the test data the accuracy is: 0.9896
Batch number 4 as the test data the accuracy is: 0.9827

Standard formulation Average: 0.8609
Hinge Loss Average: 0.9845

2)
One versers One
Batch number 0 Accuracy: 0.76412
Batch number 1 Accuracy: 0.81807
Batch number 2 Accuracy: 0.71904
Batch number 3 Accuracy: 0.77078
Batch number 4 Accuracy: 0.67296
One versers Rest
Batch number 0 Accuracy: 0.76101
Batch number 1 Accuracy: 0.80490
Batch number 2 Accuracy: 0.71512
Batch number 3 Accuracy: 0.76804
Batch number 4 Accuracy: 0.65349

One versers One Average: 0.7489
One versers Rest Average: 0.7405

Assignment - 2 , Shubham Jain
13683.

Q3) (a) K_1 is a kernel on $\mathcal{X}$

$$K(x,z) = e^{k_1(x,z)}$$

$$= 1 + K_1(x,z) + K_1^2(x,z) + K_1^3(x,z) + \cdots$$

Now, $K_1(x,z)$ is a kernel
$1 + K_1(x,z)$ is a kernel
$K_1^2(x,z)$ is a kernel
$\Rightarrow 1 + K_1(x,z) + K_1^2(x,z)$ is a kernel
similarily can be shown that $e^{K_1(x,z)}$ is a kernel

(b) $K(x,z) \Rightarrow e^{(\|x\|^2 + \|z\|^2)} \cdot \left(\frac{x^T z}{\|x\| \|z\|}\right)$.

Let $k_2 = e^{\|x\|^2 + \|z\|^2} \Rightarrow e^{\|x\|^2} \cdot e^{\|z\|^2}$
$\Rightarrow f(x) \cdot f(z)$

$\Rightarrow k_2$ is a kernel. (taking $\cancel{f(x)}$ & $f(x_j)$ common is the way)

Now, $K_3 \Rightarrow \frac{x^T z}{\|x\|^2 \|z\|^2}$

Gram matrix $\Rightarrow$
$$\begin{bmatrix} \frac{x_1^T x_1}{\|x_1\|^2 \|x_1\|} & & - \\ \frac{x_2^T x_2}{\|x_1\|^2 \|x_2\|^2} & & \\ \vdots & & \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \frac{1}{\|x_1\|^2} & & & \\ & \frac{1}{\|x_2\|^2} & & \\ & & \ddots & \\ & & & \frac{1}{\|x_n\|^2} \end{bmatrix} \begin{bmatrix} \frac{x_1^T x_1}{\|x_1\|^2} & \frac{x_2^T x_1}{\|x_2\|^2} & \cdots \\ \frac{x_1^T x_1}{\|x_1\|^2} & & \cdots \\ \vdots & & \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \frac{1}{\|x_1\|^2} & & \\ & \ddots & \\ & & \frac{1}{\|x_n\|^2} \end{bmatrix} \underbrace{\begin{bmatrix} x_1^T x & x_2^T x_1 & \cdots \\ & & \end{bmatrix}}_{K'} \begin{bmatrix} \frac{1}{\|x_1\|^2} & & \\ & \ddots & \\ & & \frac{1}{\|x_n\|^2} \end{bmatrix}$$

$$\Rightarrow W^T K_3 W \Rightarrow \left(W^T diag\left(\frac{1}{\|x_1\|^2} - \frac{1}{\|x_n\|^2}\right)\right)^T K' \left(W^T diag\left(\frac{1}{\|x_1\|^2} - \right)\right)^T$$

$$\Rightarrow Z^T K' Z$$

Now since inner product are kernels

$\Rightarrow$ K_3 is a kernel

Then $K_2 \times K_3$ is also a kernel.

(C) = $K(x, z) = \sum_{i \geq 1}^{d} \min(|x_i|, |z_i|)$ is a kernel

Let x_1 & x_2 be two points, we will show this a kernel by giving a ϕ function. Let D be the value that can normalize all x_i so that Dx_i is an integer. We will just convert this

~~x_1 x_{n1} to Dx_n~~ Dx_{i} to $D\times d$ binary vector

which is just $\langle \overbrace{\frac{1}{\sqrt{D}}, \cdots \frac{1}{\sqrt{D}}}^{Dx_{i1}}, 0 \cdots 0,$

Also we take the ~~norm~~ absolute value of x_{ik} $\forall k$

$\underbrace{\frac{1}{\sqrt{D}} \frac{1}{\sqrt{D}} \cdots \frac{1}{\sqrt{D}}}_{|x_{i2}| \times D} 0 0 \cdots,$

$1 1 0 0 \cdots 0 \rangle$ $D x_{..} \times D$

Now, $\phi(x_1) \cdot \phi(x_2) \Rightarrow \frac{1}{D} \times \sum_{k}$ ~~time~~ minimum number of ones in $|x_{ik}|, |x_{jk}|$

$$\Rightarrow \frac{1}{D} \sum_{k} \min(D|x_{ik}|, D|x_{jk}|)$$

$$\Rightarrow \cancel{\frac{1}{D}} \sum_{k} \min(|x_{ik}|, |x_{jk}|)$$

Hence we can see that this ϕ function Serves for the kernel.

hence K is a Kernel for rationals.

(Q4) $$\min_{w, \xi} \sum_{i=1}^{n} \xi_i^2$$

subject to $y_i - w^T x_i = \xi_i, \forall i = 1 - n$

$\|w\|_2 \leq B$

Let us take $\|w\|_2^2 \leq B^2$ to make thing easier

$$L'(\xi, w, \alpha, \overline{B}) \Rightarrow \sum_{i=1}^{n} \xi_i^2 + \alpha(\|w\|_2^2 - B^2) + \sum_{j=1}^{n} \beta_j (y_j - w^T x_j - \xi_j)$$

For dual problem:- $\frac{\partial L'}{\partial \xi_i} = 0 \; \forall i$, $\frac{\partial L'}{\partial w} = 0$

$$2\xi_i - \beta_i = 0$$

$$\beta_i = 2\xi_i$$

$$2\alpha w - \sum_{j=1}^{n} \beta_j x_j = 0$$

$$w = \frac{1}{2\alpha}\sum_{j=1}^{n} \beta_j x_j$$

then dual problem becomes

$$\sum_{i=1}^{n}\left(\frac{\beta_i}{2}\right)^2 + \alpha\left(\frac{1}{4\alpha^2}\left\langle \sum_{i=1}^{n} \beta_i x_i, \sum_{i=1}^{n} \beta_i x_i \right\rangle - B^2\right)$$

$$+ \sum_{j=1}^{n} \beta_j\left(y_j - \frac{1}{2\alpha}\sum_{i=1}^{n} \beta_i x_i^T x_j - \frac{\beta_j}{2}\right)$$

$$\Rightarrow \frac{1}{4}\sum_{i=1}^{n} \beta_i^2 + \frac{1}{4\alpha}\left\langle \sum_{i=1}^{n} \beta_i x_i, \sum_{i=1}^{n} \beta_i x_i \right\rangle - \alpha B^2 + \sum_{j=1}^{n} \beta_j y_j$$

$$- \frac{1}{2\alpha}\left\langle \sum_{j=1}^{n} \beta_j x_j, \sum_{j=1}^{n} \beta_j x_j \right\rangle - \frac{1}{2}\sum_{j=1}^{n} \beta_j^2$$

$$\max_{\alpha, \bar{\beta}} \; -\frac{1}{4}\sum_{i=1}^{n} \beta_i^2 - \frac{1}{4\alpha}\left\langle \sum_{j=1}^{n} \beta_j x_j, \sum_{j=1}^{n} \beta_j x_j \right\rangle - \alpha B^2 + \sum_{j=1}^{n} \beta_j y_j$$

$$\Rightarrow \min_{\alpha, \bar{\beta}} \; \frac{1}{4}\sum_{i=1}^{n} \beta_i^2 + \frac{1}{4\alpha}\left\langle \sum_{j=1}^{n} \beta_j x_j, \sum_{j=1}^{n} \beta_j x_j \right\rangle + \alpha B^2 - \sum_{j=1}^{n} \beta_j y_j$$

Now, KKT conditions.

$$\frac{\partial L}{\partial \beta_i}(\xi_i, w_i^*, \alpha^*, \beta^*) = 0$$

$$y_i - w^{T*} x_i - \xi_i^* = 0 \quad \forall i$$

$$\alpha^* \geq 0$$

$$\alpha^*\left(\frac{1}{4\alpha^{*2}}\left\langle \sum_{i=1}^{n} \beta_i^* x_i, \sum_{i=1}^{n} \beta_i^* x_i \right\rangle - B^2\right)$$

$$\alpha^*\left(\frac{1}{4\alpha^{*2}}\left\langle \sum_j \beta_j^* x_j, \sum_j \beta_j^* x_j \right\rangle - B^2\right) = 0$$

$\Rightarrow \alpha^* = 0$ or $\alpha^* = \frac{1}{2B} \sqrt{\langle \sum_i B_i x_i, \sum_j B_j x_j \rangle}$

(b) $\xi_i^* = y_i - w^{*T} x_i$

Now, it is possible that $\xi_i^* \neq 0$ for all i's & hence no x_i's lie on the margin plane in that case there is no concept of support but it is possible for same x_i's that ξ_i are zero & those can still be called as supports. Hence, it is possible to have support vector which are rare & in general there may be no support vectors.

(c) The basic disadvantage as compared to SVM is the sparsity that is lacking here. That it means it that in SVM the α_i's were zero & hence ~~that~~ made it less computational expensive but here B_i's are mostly non zero leading to high computation cost.