AI20BTECH11006
Assignment-2
AI2000

Assignment -2

1) $L_p = \|w_\gamma\|^2 + \cancel{C\sum_i \xi_i} - \sum \alpha_{\gamma i}(y_i(x_i^T \underline{w}_\gamma + b_\gamma) - \gamma + \cancel{\xi_i}) - \cancel{\sum \beta_i \xi_i}$

$\frac{\partial L_p}{\partial w_\gamma} = 2w_\gamma - \sum_i \alpha_{\gamma i} y_i x_i = 0$

$\Rightarrow w_\gamma = \frac{1}{2}\sum \alpha_{\gamma i} y_i x_i$

$\frac{\partial L_p}{\partial b_\gamma} = -\frac{1}{2}\sum_i \alpha_{\gamma i} y_i = 0 \Rightarrow \sum_i \alpha_{\gamma i} y_i = 0$

$\cancel{\frac{\partial L_p}{\partial \xi_i} = C - \beta_i - \alpha_i = 0 \Rightarrow C = \beta_i + \alpha_i}$

$L_D = \gamma \sum_i \alpha_{\gamma i} - \frac{1}{2}\sum_{i,j} \alpha_{\gamma i}\alpha_{\gamma j} y_i y_j (x_i^T x_j)$

The solution ($\alpha_\gamma$) to the above lagrangian is just $\gamma\alpha$ where $\alpha$ is the solution when $\gamma = 1$

$\therefore L_D = \gamma^2 \left(\sum_i \alpha_i - \frac{1}{2}\sum \alpha_i \alpha_j y_i y_j (x_i^T x_j)\right)$

$\therefore \quad w_\gamma = \gamma w$

Now since $\alpha_\gamma = \gamma\alpha$, the support vectors are the same

$\therefore y_i(x_i^T \underline{w}_\gamma + b_\gamma) = \gamma$ for all support vectors

clearly $\boxed{b_\gamma = \gamma b}$

$\therefore$ The solution set remains same.

2) $\alpha_i ( y_i (w^T x_i + b) - 1) = 0 \quad \forall \; i$

$$w^T(\alpha_i y_i x_i) + \alpha_i y_i b - \alpha_i = 0$$

summing over all i

$$w^T \sum \alpha_i y_i x_i + b \underbrace{\cancel{\sum \alpha_i y_i}}_{0} - \sum \alpha_i = 0$$

$$w^T w = \sum \alpha_i$$

$$\|w\|^2 = \sum \alpha_i$$

Hence Proved

3)

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

$K_1(x,z)$, $K_2(x,z)$ are valid kernels,

$\therefore$ $K_1(x,z)$, $K_2(x,z)$ both satisfy mercer condition

$$\int f(x) f(z) K_1(x,z) \, dx dz \geq 0$$

$$\int f(x) f(z) K_2(x,z) \, dx dz \geq 0$$

adding both

$$\int f(x) f(z) (K_1(x,z) + K_2(x,z)) \, dx \, dz \geq 0$$

$$\int f(x) f(z) K(x,z) \, dx \, dz \geq 0$$

b) kernel functions can be written as dot products

$k_1(x,z) = \phi_1(x)^T \phi_1(z)$

$k_2(x,z) = \phi_2(x)^T \phi_2(z)$

$\cancel{k_1(x,z)k_2(x,z) = (\phi_1(x)^T\phi_2(x))}($

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

$= \phi_1(x)^T \phi_1(z)\, \phi_2(x)^T \phi_2(z)$

$= (\phi_1(x)^T \phi_2(x))\, (\phi_1(z)^T \phi_2(z))$.

$\therefore$ $k$ is a valid vernel

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

using ⓐ & ⓑ, we can argue that $k(x,z)$ is a valid kernel, because all ~~the~~ terms would be sum of product of some power of given kernel functions

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

$\exp(k_1(x,z))$ can be expanded as

$$1 + k_1(x,z) + \frac{k_1^2(x,z)}{2!} + \cdots$$

following from ⓑ, we can infer that $\exp(k_1(x,z))$ is a valid kernel

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

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

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

Now $\exp\left(-\frac{\|x\|^2}{\sigma^2}\right)$, $\exp\left(-\frac{\|z\|^2}{\sigma^2}\right)$ are valid kernels &

$\exp\left(\frac{2x^Tz}{\sigma^2}\right)$ can also be represented as a dot product in expansion.

$\therefore$ by (a), (b), this is a valid kernel.

4)

**A)**

```
The accuracy is 0.9787735849056604
Number of support vectors is  28
```

**B)**

```
for first 50 data points
The accuracy is 0.9811320754716981
Number of support vectors is  2

for first 100 data points
The accuracy is 0.9811320754716981
Number of support vectors is  4

for first 200 data points
The accuracy is 0.9811320754716981
Number of support vectors is  8

for first 800 data points
The accuracy is 0.9811320754716981
Number of support vectors is  14
```

**C)**

**i) false**
**ii) true**
**iii) false**
**iv) false**

*The values are given below*

```
C: 0.000100 ,degree = 2
The training accuracy is 0.7463164638052531
The test accuracy is 0.7429245283018868
Number of support vectors is  1112

C: 0.000100 ,degree = 5
The training accuracy is 0.9814221652786675
The test accuracy is 0.9716981132075472
Number of support vectors is  188

C: 0.001000 ,degree = 2
The training accuracy is 0.985906470211403
The test accuracy is 0.9740566037735849
Number of support vectors is  456

C: 0.001000 ,degree = 5
```

```
The training accuracy is 0.9935938500960922
The test accuracy is 0.9811320754716981
Number of support vectors is  72

C: 0.010000 ,degree = 2
The training accuracy is 0.9948750800768738
The test accuracy is 0.9811320754716981
Number of support vectors is  132

C: 0.010000 ,degree = 5
The training accuracy is 0.9955156950672646
The test accuracy is 0.9834905660377359
Number of support vectors is  34

C: 1.000000 ,degree = 2
The training accuracy is 0.9955156950672646
The test accuracy is 0.9787735849056604
Number of support vectors is  28

C: 1.000000 ,degree = 5
The training accuracy is 0.9961563100576554
The test accuracy is 0.9764150943396226
Number of support vectors is  25
```

**D)**

The lowest training error is obtained for C=$10^6$

The lowest test error is obtained for C=100

Here are all the obtained values

```
C = 0.010000
The training error is 0.0038436899423446302
The test error is 0.02358490566037741
Number of support vectors is  406

C = 1.000000
The training error is 0.004484304932735439
The test error is 0.02122641509433964 6
Number of support vectors is  31

C = 100.000000
The training error is 0.0032030749519538215
The test error is 0.018867924528301883
```

Number of support vectors is  22

C = 10000.000000
The training error is 0.002562459961563124
The test error is 0.02358490566037741
Number of support vectors is  19

C = 1000000.000000
The training error is 0.0006406149903908087
The test error is 0.02358490566037741
Number of support vectors is  17

5)

a) **The training error is : 0.0**
**The Validation error is 0.02400000000000002**
**Number of support vectors is  1084**

b) The rbf kernel yields lower training error
Here are the values obtained
**RBF**
**The training error is : 0.0**
**The validation error is :  0.5**
**Number of support vectors is  6000**

**Polynomial**
**The training error is : 0.0004999999999999449**
**The validation error is 0.020000000000000018**
**Number of support vectors is  1332**