Assignment - 2

DATE ___/___/___

MACHINE LEARNING

- Abhishek Agarwal (2016126)

Q.2

The RBF kernel is of the form:

$$K(x_1, x_2) = e^{(-\gamma\|x_1 - x_2\|^2)}$$

where $\gamma\|x_1 - x_2\|^2$ represents the squared euclidean distance b/w the two datapoints x_1 and x_2.

The value of the RBF kernel is dependent on the distance between the points.
If the distance is huge, the value tends to 0, and if the points are very near the value tends towards 1.

The dimension to which rbf maps increases with more number of training samples.
However, γ also plays a role in determining overfitting of the data.
If the value of γ is taken large, the value of kernel function tends to zero for every (x_1, x_2) with $x_1 != x_2$ and it tends to 1 for every (x_1, x_2) with $x_1 = x_2$.

Thus the kernel matrix obtained is of the form:

$$\begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \\ 0 & \vdots & 0 & \ddots \\ \vdots & & \vdots & \end{bmatrix}$$

which is the identity matrix.

Hence, every point gets mapped to its own dimension, and we have a total of (num_training_samples) dimensions.

This is the case of overfitting as train accuracy will be 100% whereas on test set, poor results will be obtained.

Whereas if we choose a gamma which has a very low value.
In this case, the kernel function will tend to 1.
Thus, the kernel matrix obtained will be like

$$K \quad \begin{bmatrix} 1 & 1 & \cdots & \cdots \\ 1 & 1 & 1 & \cdots \\ 1 & 1 & \cdots & \cdots \\ 1 & & 1 & \end{bmatrix}$$

and hence, every point is getting mapped to ~~the~~ almost the same point.
This is the case of underfitting

∴ We should choose an appropriate value of γ to avoid over/under fit of data.

Q.3 According to the definition of a convex function:
$f(x)$ is a convex function on interval $[a, b]$
if $\forall\ x_1, x_2 \in [a, b]$ and $\lambda \in (0, 1)$

$$f[\lambda x_1 + (1-\lambda)x_2] \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

Now, assume that f and g are 2 convex functions.

T.P.: $f+g$ is a convex function

i.e. we have to prove that
$$(f+g)(\lambda x_1 + (1-\lambda)x_2) \leq \lambda (f+g)(x_1) + (1-\lambda)(f+g)(x_2)$$

convexity of f $\Rightarrow$
$$f[\lambda x_1 + (1-\lambda)x_2] \leq \lambda f(x_1) + (1-\lambda) f(x_2) \quad -①$$
convexity of g $\Rightarrow$
$$g[\lambda x_1 + (1-\lambda)x_2] \leq \lambda g(x_1) + (1-\lambda) g(x_2) \quad -②$$

Adding the eq^n ① and ②

$$(f+g)[\lambda x_1 + (1-\lambda)x_2] \leq \lambda (f+g)(x_1) + (1-\lambda)(f+g)(x_2)$$

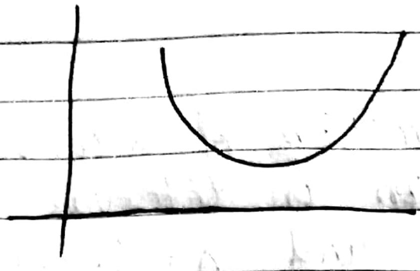
$\therefore$ Which is what we required to prove

(ii) T.P. Loss function for L_1 is convex function.

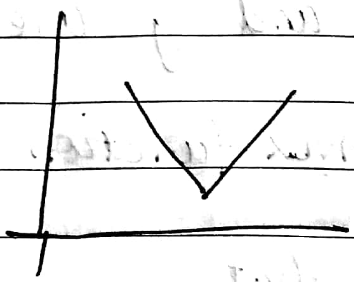
$$L = \underbrace{\sum_{\text{over all samples}} (y_i - f(x_i, w))^2}_{①} + \underbrace{\sum |w_j|}_{②}$$

Now, since we know from the graph, that quadratic function is a convex function

DATE ___/___/___



Similarly, for the ② part
It is the sum of absolute values
which has a graph like



which is again convex.

∴ Both of them are convex
⇒ From part (i) that loss function is convex function