Name - Mudit Samal
Roll No - 203100068
EE769

Ans - 1 given $J(w) = \frac{1}{2}\sum_{n=1}^{N}\{w^T\phi(x_n) - t_n\}^2 + \frac{\lambda}{2}w^Tw$ — (1)

where $\lambda \geq 0$

$\Rightarrow \frac{\partial J}{\partial w} = 0$

$\Rightarrow w = -\frac{1}{\lambda}\sum_{n=1}^{N}\{w^T\phi(x_n) - t_n\}\phi(x_n)$

$= \sum_{n=1}^{N} a_n \phi(x_n) = \phi^T a$

where ϕ is design matrix, whose n^{th} Row is given by $\phi(x_n)^T$

$a_n = -\frac{1}{\lambda}\{w^T\phi(x_n) - t_n\}$

put $w = \phi^T a$ in eq^n (1)

$J(a) = \frac{1}{2} a \phi\phi^T\phi\phi^T a - a^T\phi\phi^T t + \frac{1}{2}t^T t + \frac{\lambda}{2}a^T\phi\phi^T a$

Since $\boxed{K_{nm} = \phi(x_n)^T\phi(x_m) = K(x_n, x_m)}$

$J(a) = \frac{1}{2}a^T K K a - a^T K t + \frac{1}{2}t^T t + \frac{\lambda}{2}a^T K a$

now $\frac{\partial J(a)}{\partial a} = 0$

$\Rightarrow \boxed{a = (K + \lambda I_N)^{-1} t}$

Ans 2   Proove $k(x_i, x_j) = \exp(-\alpha\|x_i - x_j\|^2)$ is a Kernal

$\Rightarrow \quad \|x_i - x_j\|^2 = x_i^T x_i + x_j^T . x_j - 2x_i^T x_j$

So Putting this expanded form in given eqn

$$K(x_i, x_j) = \underbrace{\exp\left(-\frac{x_i^T x}{2\sigma^2}\right)}_{I} \underbrace{\exp\left(\frac{x_i^T x_j}{\sigma^2}\right)}_{II} \underbrace{\exp\left(-\frac{x_j^T . x_j}{2\sigma^2}\right)}_{III}$$

From Property 6.16 :

$$k(x, x') = \exp(k_1(x, x'))$$

$$x_i^T x_j = k(x_i, x_j)$$

So (II) term is a kernal $k(x_i, x_j)$

Now the eqn become

$$k(x_i, x_j) = \underset{I}{f(x_i)}\ \underset{II}{k(x_i, x_j)}\ \underset{III}{g(x_j)}$$

from Property 6.14 → this eqn matches.

So Hence proove that the given eqn is a Valid kernel

Scanned with CamScanner

Ans-(3) ~~Increase~~ Increase in K will Increase the no. of Support vectors eventually decreases the smoothness. and Viceversa when decreasing the K.

Ans-4

(a) Number of Input Neurons = 50×50 = 2500

(b) Number of output Neurons

= No.s of animal to be classified + None of those

= 5 Ans

(c) Activation function :- Sofmax (for multiple classification)

(d) Loss function :- "Cross entropy loss function"

$$C.E.L = -\sum_{i=1}^{J} t_i \log(f(s)_i)$$

J = No. of classes

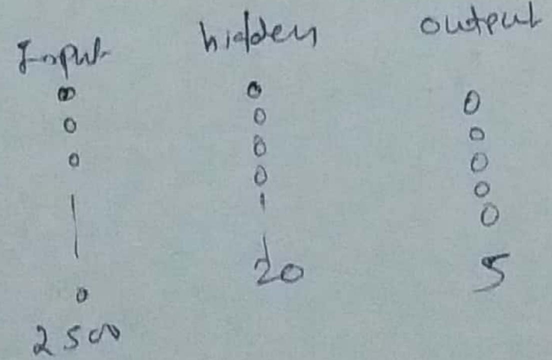
t_i = target

where, $f(s)_i = \dfrac{\exp(s_i)}{\sum_{j=1}^{J} \exp(s_j)}$

(e) examples of target : $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ - -

(f) Activation function on Hidden Layer = "ReLU"

(g)



for a single hidden layer with 20 Neurons

=) weights are

= 2500×20 + 20×5

= 50,100

(h) No. of biases => hidden layer Neuron and output Neuron have their biases.

=> 20+5 = 25

Ans 1 (5) Single Continuous Variable (Time) is the output

output Neuron Changes from 5 to 1

(a) No. of Input Neuron → 50×50 = 2500

(b) No. of output Neurons → 1

(c) Activation function on output Neuron → linear

(d) Loss function :- Mean Square Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - \hat{y}_i)^2$$

t_i = Actual Label

$\hat{y}_i$ = Predicted Label

(e) [one Single Real Value] → Since time is target value eg. → [1]

(f) Activation function on Hidden Layer :- ReLU

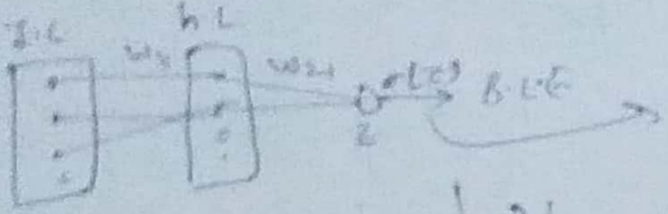
(g) No. of weights = 2500×20 + 20×1

= 50020

(h) No. of biases = 20 + 1 = 21

Ans-6 Derivation for the Gradients in B.C.E Cases for one Samples

$B.C.E = \{ t \log(p) + (1-t)\log(1-p) \}$



$\sigma(z)$ — sigmoid

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_{21}} \cdot \frac{\partial w_{21}}{\partial w_{11}}$$

$\frac{\partial L}{\partial z} \longrightarrow$ Two Case

$t = 0$

$L = \log(1-p)$

$$\frac{\partial(\log(1-p))}{\partial z} = \frac{\partial L}{\partial z}\Big|_{t=0}$$

$t = 1$

$$\frac{\partial L}{\partial z}\Big|_{t=1} = \log p$$

$$\frac{\partial L}{\partial z}\Big|_{t=1} = \frac{\partial(\log(\sigma(z)))}{\partial z} = \frac{1}{\sigma(z)} \times \sigma(z)[1-\sigma(z)]$$

$\{\because \sigma'(x) = \sigma(x)[1-\sigma(x)]$

$$= [1-\sigma(z)]$$

$$\frac{\partial L}{\partial z}\Big|_{t=0} = \frac{\partial}{\partial z}(\log(1-p)) = \frac{1}{1-\sigma(z)} \times -\sigma(z)[1-\sigma(z)] = -\sigma(z)$$

Assume → Relu Activation function is used. B/w Layers

$$\frac{\partial z}{\partial w_{21}} = \frac{\partial[\text{Relu}(w_{21}x_1 + b_1)]}{\partial w_{21}} \longrightarrow \text{Two Cases}$$

I

$$\frac{\partial z}{\partial w_{2i}} = x_i$$

II (−ve Relu)

$$\frac{\partial z}{\partial w_{21}} = 0$$

$$\frac{\partial L}{\partial w_{21}} = [1-\sigma(z)] \cdot x_i \quad \text{for } t=0$$

where

$$\sigma(z) = \frac{1}{1+\exp(w_{21}x_1+b)}$$

for $t = 1$

$$= -\sigma(z)\, x_i$$

$$\frac{\partial w_{21}}{\partial w_{11}} = \frac{\partial[\text{Relu}(w_{11}x_1 + b_1)]}{\partial w_{11}}$$

→ x_i → +ve Relu

→ 0 → −ve Relu

Similarly for biases;

If we do not put t as 0 and 1, and wants to put in the eq^n and carry on then

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial b_2} \cdot \frac{\partial B_2}{\partial b_1}$$

$$\frac{\partial L}{\partial z} = \frac{\partial}{\partial z}\left(t \log(\sigma(z)) + (1-t) \log(1-\sigma(z))\right]$$

$$\frac{\partial z}{\partial b_2} = 1$$

$$\frac{\partial L}{\partial z} = t\,[1-\sigma(z)) + (1-t)\,\sigma(z)$$

$$\frac{\partial L}{\partial b} = t\left[1 - \frac{1}{1+\exp(w_{ij}x_{ij}+b_1)}\right] + (1-t)\,\frac{1}{1+\exp(w_{ij}x_{ij}+b_i)}$$

Scanned with CamScanner