Section D — Ridge Regression

Q.5 If $E(w)$ is the objective, which is differentiable in the neighbourhood of point a, in order to decrease the error function we move in the direction −gradient i.e $-\nabla E(a)$

$$\therefore \quad b = a - \eta \nabla E(a) \quad -(1)$$

where η - learning rate

$\nabla E(a)$ - gradient of error

Steps

① Initialize $w(0)$ and $t = 1$.

② while $\eta' > \epsilon$

$$w = w^{(t-1)} - \eta' \nabla E(w^{(t-1)})$$

If $E(w) < E(w^{(t-1)})$, break

$$\eta' = \eta'/2$$

③ $w^{(t)} = w$

④ $t = t+1$

Let the original loss function be L_0, to add L_2 regularization term,

$$L(w) = L_0(w) + \lambda \|w\|_2^2$$

or

$$\nabla E(w) = -\sum_{n=1}^{N} (t_n - w \cdot \phi(x_n)) \phi(x_n) + 2\lambda w$$

where λ - regularization parameter.

SGD : We don't require whole data set in one go, we update the model parameter after accessing each data pt.

Error function for SGD

$$E_n(w) = \frac{1}{2}(t_n - w^T \phi(x_n))^2 + \lambda \|w\|_2^2$$

Steps -

1) Initialize the parameter to $w^{(0)} = 1 \quad 2 = 1$

2) while :

→ select random pt (x_n, y_n)

→ $w^{(2)} = w^{(2-1)} - \eta^{(2)} \nabla E_n(w^{(2-1)}) - \lambda \|w\|^2$

→ $2 = 2 + 1$

where $2 \rightarrow$ Iteration

$\lambda \rightarrow$ Regularization parameter.