4) Given function is –

$$\boxed{Y = w_0 + w_1 e^{-x_1} + w_2 x_1 + w_3 x_1 x_2}$$

∴ Let us take Sum of Squared Error

as loss function :

$$\therefore \quad SSE = \left(Y - w_0 - w_1 e^{-x_1} - w_2 x_1 - w_3 x_1 x_2\right)^2$$

∴ For Stochastic Gradient Descent (SGD),

we have –

$$\boxed{w_i^{T+1} = w_i^{T} - \alpha \nabla_{w_i} SSE \, .} \quad \rightarrow (i)$$

where α = Learning Rate

P.T.O

Now we will find gradient descent w.r.t of Loss Function w.r.t all weights –

$$\nabla_{w_0} = \frac{\partial SSE}{\partial w_0} = -2\left(y - w_0 - w_1 e^{-x_1} - w_2 x_1 - w_3 x_1 x_2\right)$$

$$\boxed{\Rightarrow \nabla_{w_0} = -2\sqrt{SSE}} \rightarrow \text{(ii)}$$

$$\nabla_{w_1} = \frac{\partial SSE}{\partial w_1} = -2e^{-x_1}\left(y - w_0 - w_1 e^{-x_1} - w_2 x_1 - w_3 x_1 x_2\right)$$

$$\boxed{\Rightarrow \nabla_{w_1} = -2e^{-x_1}\sqrt{SSE}} \rightarrow \text{(iii)}$$

Similarly,

$$\boxed{\nabla_{w_2} = \frac{\partial SSE}{\partial w_2} = -2x_1\sqrt{SSE}} \rightarrow \text{(iv)}$$

$$\boxed{\nabla_{w_3} = \frac{\partial SSE}{\partial w_3} = -2x_1 x_2\sqrt{SSE}} \rightarrow \text{(v)}$$

So, we have found the gradient descent w.r.t all weights

P.T.O

So, coefficient update are-

(using (ii), (iii), (iv), (v) in (i))

$$\boxed{w_0^{T+1} = w_0^T + 2\alpha\sqrt{SSE}}$$

$$\boxed{w_1^{T+1} = w_1^T + 2\alpha e^{-x_1}\sqrt{SSE}}$$

$$\boxed{w_2^{T+1} = w_2^T + 2\alpha x_1\sqrt{SSE}}$$

$$\boxed{w_3^{T+1} = w_3^T + 2\alpha x_1 x_2\sqrt{SSE}}$$

If we add a ridge regularization penalty, we get -

$$\boxed{w_0^{T+1} = w_0^T + 2\alpha\sqrt{SSE}} \rightarrow \text{No ridge penalty for intercept term}$$

$$\boxed{w_1^{T+1} = w_1^T + 2\alpha e^{-x_1}\sqrt{SSE} + 2\lambda w_1^T}$$

$$\boxed{w_2^{T+1} = w_2^T + 2\alpha x_1\sqrt{SSE} + 2\lambda w_2^T}$$

$$\boxed{w_3^{T+1} = w_3^T + 2\alpha x_1 x_2\sqrt{SSE} + 2\lambda w_3^T}$$

where α = learning rate

λ = regularization constant