Stochastic Gradient Descent - Part a.

$$\hat{y} = w_0 + w_1 e^{-x_1} + w_2 x_1 + w_3 x_1 \cdot x_2$$

$$\text{Loss} = \sum (y - \hat{y})^2$$

Note: we are not using ridge penalty here

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w}$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}} \left(y^2 + \hat{y}^2 - 2y\hat{y} \right) = 2(\hat{y} - y)$$

$$\frac{\partial L}{\partial w_0} = 1 \qquad \frac{\partial L}{\partial w_1} = e^{-x_1}$$

$$\frac{\partial L}{\partial w_2} = x_1 \qquad \frac{\partial L}{\partial w_3} = x_1 \cdot x_2$$

$$w^{t+1} = w^t - \underbrace{\eta}_{\text{learning rate}} \cdot \underbrace{\nabla}_{\text{gradient vector}}$$

$$\nabla = \left[\frac{\partial L}{\partial w_0}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \frac{\partial L}{\partial w_3} \right]$$

$$= 2(\hat{y} - y) \left[1, e^{-x_1}, x_1, x_1 \cdot x_2 \right]$$

$$w^t = [w^t_0, w^t_1, w^t_2, w^t_3]$$

$$w^{t+1} = [w_0^{t+1}, w_1^{t+1}, w_2^{t+1}, w_3^{t+1}]$$

$$\nabla = 2(\hat{y}-y)\cdot[1, e^{-x_1}, x_1, x_1 \cdot x_2]$$

$$w^{t+1} = w^t - \eta \cdot \nabla$$

Suppose we add ridge penalty

$$\text{Loss} = (\hat{y}-y)^2 + \lambda \sum w^2$$

↳ regularization parameter

in the summation of w^2 → we don't include w_0

so $\sum w^2 = w_1^2 + w_2^2 + w_3^2$

$$\frac{\partial \hat{y}}{\partial w_0} = 1 \qquad \frac{\partial \hat{y}}{\partial w'} = e^{-x_1} + 2\lambda w_1$$

$$\frac{\partial \hat{y}}{\partial w_2} = x_1 + 2\lambda w_2 \qquad \frac{\partial \hat{y}}{\partial w_3} = x_1 \cdot x_2 + 2\lambda w_3$$

$$\nabla = [1, e^{-x_1} + 2\lambda w_1, x_1 + 2\lambda w_2, x_1 \cdot x_2 + 2\lambda w_3]$$

$$w^{t+1} = w^t - \eta \cdot \nabla$$