Problem Set 2 DL 17CS10058

Ans: 1.1)

$$\hat{y}_i = \tanh(w \cdot x_i)$$

$$\hat{y}_i = \tanh\left(\sum_{j=1}^{d} w_j x_{ij}\right)$$

$$\therefore \frac{\partial \hat{y}_i}{\partial w_k} = \left(1 - \tanh\left(\sum_{j=1}^{d} w_j x_{ij}\right)^2\right) \times \sum_{j=1}^{d} \left(\frac{\partial w_j x_{ij}}{\partial w_k}\right)$$

$$\left[\frac{\partial \tanh(x)}{\partial x} = 1 - (\tanh(x))^2 \text{ \& chain rule} \right]$$

$$\therefore \frac{\partial \hat{y}_i}{\partial w_k} = (1 - \hat{y}_i^2)\, x_{ik} \quad \text{— (1)} \qquad \left[\frac{\partial w_j}{\partial w_k} = 1 \text{ if } j = k, = 0 \text{ else} \right]$$

, where w_k is k^{th} component of w

For stochastic gradient descent,

Loss, $J(w) = l(y_i, \hat{y}_i) + \lambda \|w\|^2$, where (x_i, y_i) is a single example

$$J(w) = \log_e(1 + \exp(-y_i \cdot \hat{y}_i)) + \lambda\left(\sum_{j=1}^{d} w_j^2\right)$$

$$\therefore \frac{\partial J(w)}{\partial w_k} = \frac{1 \times \exp(-y_i \cdot \hat{y}_i)}{1 + \exp(-y_i \cdot \hat{y}_i)} \cdot \left(-y_i \cdot \frac{\partial \hat{y}_i}{\partial w_k}\right) + \lambda\left(\sum_{j=1}^{d} \frac{\partial w_j^2}{\partial w_k}\right)$$

$$= \frac{-y_i \times \frac{\partial \hat{y}_i}{\partial w_k}}{1 + \exp(y_i \cdot \hat{y}_i)} + \lambda(2 w_k)$$

$$\left[\because \frac{\partial w_j^2}{\partial w_k} = 2 w_k, \; k = j; \; = 0 \text{ else} \right]$$

$$\frac{\partial J(w)}{\partial w_k} = -\frac{y_i(1-\hat{y}_i^2)}{1+\exp(y_i \cdot \hat{y}_i)} x_{ik} + 2\lambda w_k$$ [Using ①]

w_k is kth component of w

$$\therefore \frac{\partial J(w)}{\partial w} = \frac{-y_i(1-\hat{y}_i^2)}{1+\exp(y_i \cdot \hat{y}_i)} x_i + 2\lambda w$$

, where $w \in R^d$ and $x_i \in R^d$

Stochastic gradient descent update rule

$$w_{t+1} = w_t - \eta \frac{\partial J}{\partial w_t}$$

$$\boxed{\therefore w_{t+1} = w_t - \eta\left(\frac{-y_i(1-\hat{y}_i^2)}{1+\exp(y_i \cdot \hat{y}_i)} x_i + 2\lambda w_t\right)}$$

where $\hat{y}_i = \tanh(w_t \cdot x_i)$

Ans: 2.1 Answer in Ipynb Notebook

Ans: 3.1 Answer in Ipynb Notebook