Aditya Sawant 17CS10060

Q1.17 $y_i \in \{-1, 1\}$ $i = 1, \dots N$ $x_i \in R^d$

$w \in R^d$ $\hat{y}_i = \tanh(\omega \cdot x_i)$

$$\text{loss} = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \lambda \|\omega\|^2$$

$$\frac{\partial \text{loss}}{\partial \omega} = \frac{\partial}{\partial \omega}\left[\lambda \|\omega\|^2 + \sum_{i=1}^{N} \log_e\left(1 + \exp(-y_i \cdot \hat{y}_i)\right)\right]$$

$$= 2\lambda\omega + \sum_{i=1}^{N}\left[\frac{\partial}{\partial \omega}\left(\log_e(1 + \exp(-y_i \cdot \hat{y}_i))\right)\right]$$

$$= 2\lambda\omega + \sum_{i=1}^{N} \frac{-y_i e^{-y_i \hat{y}_i}}{1 + e^{-y_i \cdot \hat{y}_i}} \frac{\partial \tanh(x_i \omega)}{\partial \omega}$$

$$= 2\lambda\omega + \sum_{i=1}^{N} \frac{(-y_i) e^{-y_i \hat{y}_i}}{1 + e^{-y_i \hat{y}_i}} (1 - \tanh^2(x_i \omega))(x_i)$$

$$\therefore \nabla_{\omega} = 2\lambda\omega + \sum_{i=1}^{N} \frac{(-y_i) e^{-y_i \hat{y}_i}}{1 + e^{-y_i \hat{y}_i}} (1 - \hat{y}_i^2)(x_i)$$

step size $= \eta$

$$\therefore \omega^{t+1} = \omega^t - \eta\left[2\lambda\omega^t + \sum_{i=1}^{N} \frac{-y_i e^{-y_i \hat{y}_i}}{1 + e^{-y_i \hat{y}_i}} (1 - \hat{y}_i^2)(x_i)\right]$$

$$\omega_{t+1} = \omega_t[1 - 2\lambda\eta] + \sum_{i=1}^{N} \eta \frac{y_i e^{-y_i \hat{y}_i}}{1 + e^{-y_i \hat{y}_i}} (1 - \hat{y}_i^2)(x_i)$$

For stochastic gradient descent take $N = 1$