AHMAD FALLAHPOUR

CSCI567 2017 Homework Assignment 2 10/8/2017

# 1 Neural networks

## 1.1

$$\frac{\partial l}{\partial \mathbf{u}} = \left(\mathbf{W}^{(2)}\right)^T \frac{\partial l}{\partial \mathbf{a}} . * H(\mathbf{u}) \tag{1}$$

$$\frac{\partial l}{\partial \mathbf{a}} = \mathbf{z} \sum_k y_k - \mathbf{y} \tag{2}$$

$$\frac{\partial l}{\partial \mathbf{W}^{(1)}} = \frac{\partial l}{\partial \mathbf{u}} \mathbf{x}^T \tag{3}$$

$$\frac{\partial l}{\partial \mathbf{b}^{(1)}} = \frac{\partial l}{\partial \mathbf{u}} \tag{4}$$

$$\frac{\partial l}{\partial \mathbf{W}^{(2)}} = \frac{\partial l}{\partial \mathbf{a}} \mathbf{h}^T \tag{5}$$

## 1.2

Because $\partial l / \partial W$ is zero, the gradient is zero. Therefore, the weights cannot be updated.

$$W^{(t+1)} = W^{(t)} - \eta * 0 = W^{(t)} \tag{6}$$

## 1.3

$$\mathbf{U} = \mathbf{W}^{(2)} \mathbf{W}^{(1)} \tag{7}$$

$$\mathbf{v} = \mathbf{W}^{(2)} \mathbf{b}^{(1)} + \mathbf{b}^{(2)} \tag{8}$$

# 2 Kernel methods

## 2.1

$$J(\omega) = \sum_n l\left(\omega^T \phi(x_n), y_n\right) + \frac{\lambda}{2} ||\omega||_2^2 \tag{9}$$

$$\frac{\partial J(\omega)}{\partial \omega} = \sum_n \frac{\partial l\left(\omega^T \phi(x_n), y_n\right)}{\partial \omega} + \lambda \omega = 0 \tag{10}$$

$$\frac{\partial J(\omega)}{\partial \omega} = \sum_n \frac{\partial l\left(\omega^T \phi(x_n), y_n\right)}{\partial (\omega^T \phi(x_n))} . \frac{(\omega^T \phi(x_n))}{\partial \omega} + \lambda \omega = 0 \tag{11}$$

$$\frac{\partial J(\omega)}{\partial \omega} = \sum_n \frac{\partial l\big(\omega^T \phi(x_n), y_n\big)}{\partial(\omega^T \phi(x_n))}.\phi(x_n) + \lambda\omega = 0 \tag{12}$$

$$\omega^* = \sum_n \frac{-1}{\lambda}.\frac{\partial l\big(\omega^T \phi(x_n), y_n\big)}{\partial(\omega^T \phi(x_n))}.\phi(x_n) \tag{13}$$

$$\omega^* = \sum_n \alpha_n.\phi(x_n) = \Phi^T \alpha \tag{14}$$

### 2.2

We can plug Eqn.(14) into Eqn.(9):

$$J(\omega) = \sum_j l\big(\sum_i \alpha_i.\phi(x_i)^T \phi(x_j), y_j\big) + \frac{\lambda}{2} || \sum_n \alpha_n.\phi(x_n)||_2^2 \tag{15}$$

$$J(\omega) = \sum_j l\big(\sum_i \alpha_i.\phi(x_i)^T \phi(x_j), y_j\big) + \frac{\lambda}{2} \sum_i \sum_j \alpha_i \alpha_j \phi(x_i)^T \phi(x_j) \tag{16}$$

$$J(\omega) = \sum_j l\big(\sum_i \alpha_i K_{ij}, y_j\big) + \frac{\lambda}{2} \sum_i \sum_j \alpha_i \alpha_j K_{ij} \tag{17}$$