CSCI 567 Assignment 4
Fall 2016

Snehal Adsule
2080872073
adsule@usc.edu

November 5,2016

1 Problem 1

1.1 1 (a)

Given that $L(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$

$$\frac{dL}{d\hat{y}_i} = g_i = 2(\hat{y}_i - y_i) \tag{1}$$

1.2 1 (b)

Given that h^*= arg min (min $\sum_{i=1}^{n}(-g_i - \gamma h(x_i))^2$)

Differentiating wrt γ for optimal h^*

$$\frac{dh}{d\gamma} = min_H(min_R \sum_{i=1}^{n} 2(-g_i - \gamma h(x_i))(-h(x_i))) = 0$$

$$\gamma = \sum_{i=1}^{n} \frac{-g_i h(x_i)}{h(x_i)^2}$$

$$\gamma = \sum_{i=1}^{n} \frac{-g_i}{h(x_i)}$$

To prove the optimal we should take the second derivative

$$\frac{d^2h}{d\gamma^2} = \sum_{i=1}^{n} 2(-h(x_i))(-h(x_i)) = 0$$

$$=> \sum_{i=1}^{n} 2h(x_i)^2 >= 0$$

As the above equation is always going to be positive, we can say that h^* is optimal.

1.3 1 (c)

Given that a^*= arg min $\sum_{i=1}^{n} L(y_i, \hat{y}_i + \alpha h^*(x_i))$ differntiating wrt α

$$\frac{dL}{d\alpha} = \sum_{i=1}^{n} 2[y_i - (\hat{y}_i + \alpha h^*(x_i))](-h^*(x_i)) = 0$$

$$\alpha^* = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)}{h^*(x_i)}$$

Differentiating again , for optimal

$$\frac{d^2L}{d\alpha^2} = \sum_{i=1}^{n} 2(-h(x_i))(-h(x_i)) = 0$$

$$=> \sum_{i=1}^{n} 2h(x_i)^2 >= 0$$

As, second derivative is positive it will be optimal solution.

2 Problem 2

2.1 2 (a)

Conside the neural network with the linear activation for hidden layer and sigmoid output. We can think of it as a one single layer input , which is the output from the hidden layer with j units $a_j = \sum h(w_{ji}x_i)$.
$z_j = \sum w_j i x_i$,
$a_j = h(z_j)$
$y_k = \sum \sigma(v_{kj}z_j)$,where h is the linear activation function.

Hence, we can think of it as reduced input a_j along with weights v_{kj} , as linear input to the final sigmoid unit. We can represent the output in terms of the logistic regression, with a_j as the input .

$f(x) = \frac{1}{1+e^{\sum v_{jk}a_j}}$

Therefore, it is equivalent to the logistic regression.

2.2 2 (b)

Given that $L(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$

The error in the output layer is given as

$$\frac{dL}{dy_j} = \delta_j = (y_j - t_j)$$

using chain rule, we can back propagate the error as

$$\begin{aligned}
\frac{dL}{da_k} &= \delta_k \\
&= \frac{dL}{da_j}\frac{da_j}{da_k}, \\
\delta_k &= \delta_j(1 - z^2)\sum v_{jk}
\end{aligned}$$

Now, derivative for weight update of v_{jk}, using chain rule and $\sigma^{'}(a) = 1 - z^2$ for tanh function

$$\begin{aligned}
\frac{dL}{dv_{jk}} &= \frac{dL}{dy_j}\frac{dy_j}{dv_{jk}} \\
&= \delta_j \frac{d\sum v_{jk} z_k}{dv_{jk}} \\
\frac{dL}{dv_{jk}} &= \delta_j z_k
\end{aligned}$$

Now, derivative for weight update update of w_{ki}, using chian rule

$$\begin{aligned}
\frac{dL}{dw_{ki}} &= \frac{dL}{da_k}\frac{da_k}{dw_{ki}} \\
&= \delta_k \frac{d\sum w_{ki} x_i}{dw_{ki}} \\
\frac{dL}{dw_{ki}} &= \delta_k x_i
\end{aligned}$$