COMP 632: Assignment 1

Due on Tuesday, January 27 2015

Presented to Dr. Doina Precup

Geoffrey Stanley
Student ID: 260645907

Question 1

A)

See source code.

B)

$$w = \begin{pmatrix} 0.88485161 & 11.02552359 & -9.35666242 \end{pmatrix} \tag{1}$$

$$w_{normalized} = \begin{pmatrix} 7.07067572 & 84.7476132 & -9.35666242 \end{pmatrix} \tag{2}$$

C)

See source code.

D)

E)

Fold #1 results are:
Training error is : [17633.45681241]
Testing error is : [4795.16430455]
Fold #2 results are:
Training error is : [16317.25933808]
Testing error is : [6362.29874466]
Fold #3 results are:
Training error is : [15395.564805]
Testing error is : [7339.05242593]
Fold #4 results are:
Training error is : [18457.0083633]
Testing error is : [3879.4511839]
Fold #5 results are:
Training error is : [20136.42642011]
Testing error is : [2188.4234707]

F)

Question 2

When simplifying the maximum likelihood equation for a regression whose variables maintain a constant normal distribution we are able to obtain the sum-squared-error function. However, when the standard deviation varies from one variable to another this simplification is no longer achievable.

$$L(w) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\left(\frac{y_i - h_w(x_i)}{\sigma_i}\right)^2} \tag{3}$$

$$\log L(w) = \sum_{i=1}^{m} \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right) - \sum_{i=1}^{m} \frac{1}{2}\left(\frac{y_i - h_w(x_i)}{\sigma_i}\right)^2 \tag{4}$$

$$\log L(w) = \sum_{i=1}^{m} \log\left((2\pi)^{-1/2}\sigma_i^{-1}\right) - \sum_{i=1}^{m} \frac{1}{2}\left(\frac{y_i - h_w(x_i)}{\sigma_i}\right)^2 \tag{5}$$

$$\log L(w) = -\frac{1}{2}\log(2\pi) - \sum_{i=1}^{m}\log\sigma_i - \sum_{i=1}^{m}\frac{1}{2}\left(\frac{y_i - h_w(x_i)}{\sigma_i}\right)^2 \tag{6}$$

$$\frac{\partial}{\partial w_j}\log L(w) = \frac{\partial}{\partial w_j}\left(-\frac{1}{2}\log(2\pi) - \sum_{i=1}^{m}\log\sigma_i - \sum_{i=1}^{m}\frac{1}{2}\left(\frac{y_i - h_w(x_i)}{\sigma_i}\right)^2\right) \tag{7}$$

$$\frac{\partial}{\partial w_j}\log L(w) = -\sum_{i=1}^{m}\frac{x_i}{\sigma_i}\left(\frac{y_i - h_w(x_i)}{\sigma_i}\right) \tag{8}$$

Simplying further while converting standard deviation to variance. Where $\sigma^2 = \Omega$.

$$\frac{\partial}{\partial w_j}\log L(w) = \sum_{i=1}^{m}\frac{x_i y_i}{\Omega_i} - \sum_{i=1}^{m}\frac{h_w(x_i)x_i}{\Omega_i} \tag{9}$$

$$\frac{\partial}{\partial w_j}\log L(w) = X^T\Omega^{-1}Y - wX^T\Omega^{-1}X \tag{10}$$

Now setting the gradient to zero.

$$0 = X^T\Omega^{-1}Y - wX^T\Omega^{-1}X \tag{11}$$

$$w^* = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}Y \tag{12}$$

Question 3

Given the following huberized loss function:

$$L_H(w,\delta) = \begin{cases} (y_i - w^T x_i)^2/2 & \text{if } |y_i - w^T x_i| \le \delta \\ \delta|y_i - w^T x_i| - \delta^2/2 & \text{otherwise} \end{cases} \tag{13}$$

A)

Find the derivative:

$$\frac{\partial}{\partial w}L_H(w,\delta) = \begin{cases} \frac{\partial}{\partial w}\left((y_i - w^T x_i)^2/2\right) & \text{if } |y_i - w^T x_i| \le \delta \\ \frac{\partial}{\partial w}\left(\delta|y_i - w^T x_i| - \delta^2/2\right) & \text{otherwise} \end{cases} \tag{14}$$

The equation $|y_i - w^T x_i|$ is equivalent to $\sqrt{(y_i - w^T x_i)^2}$ as such the derivative becomes:

$$\frac{\partial}{\partial w}L_H(w,\delta) = \begin{cases} (y_i - w^T x_i)x_i & \text{if } |y_i - w^T x_i| \le \delta \\ \delta\frac{y_i - w^T x_i}{|y_i - w^T x_i|}x_i & \text{otherwise} \end{cases} \tag{15}$$

The expression $\frac{y_i - w^T x_i}{|y_i - w^T x_i|}$ is equivalent to the sign of $y_i - w^T x_i$ as such the derivative can be written as:

$$\frac{\partial}{\partial w}L_H(w,\delta) = \begin{cases} (y_i - w^T x_i)x_i & \text{if } |y_i - w^T x_i| \le \delta \\ \delta sign(y_i - w^T x_i)x_i & \text{otherwise} \end{cases} \tag{16}$$

Where $sign(y_i - w^T x_i)$ is replaced by 1 or -1 depending on the result of the equation is positive or negative.

B)

The algorithm was implemented in python source code and is defined as function GradientDescent.

C)

Question 4

$$h_{w1,\ldots wk}(x) = \prod_{k=1}^{K} \phi_k(x) \tag{17}$$

Where:

$$\phi_k(x) = e^{w_k^T x} = e^{\sum_{i=1}^{m} w_{ik} x_i} \tag{18}$$

Replacing $\phi_k(x)$

$$h_{w1,\ldots wk}(x) = \prod_{k=1}^{K} e^{\sum_{i=1}^{m} w_{ik} x_i} \tag{19}$$

$$\log h_{w1,\ldots wk}(x) = \sum_{k=1}^{K} \sum_{i=1}^{m} w_{ik} x_i \tag{20}$$

$$\frac{\partial}{\partial w} \log h_{w1,\ldots wk}(x) = \frac{\partial}{\partial w} \left(\sum_{k=1}^{K} \sum_{i=1}^{m} w_{ik} x_i \right) \tag{21}$$

$$\frac{\partial}{\partial w} \log h_{w1,\ldots wk}(x) = \sum_{i=1}^{m} x_i \tag{22}$$