Introduction to Machine Learning - HW1

Professors: Abolghasemi & Erabi
Student: Mohamad Mahdi Samadi
Student ID: 810101465

Q4.

In a regression task, we want an exponential relationship between feature and target as it would be $y = e^{wx}$. Where $x \in R$ is the feature, $y \in R$ is the target and $w \in R$is the parameter. We are provided with a training dataset $D = \{\{x_1, y_1\}, \{x_2, y_2\}, ..., \{x_n, y_n\}\}$. Answer the following questions.

a. Form the sum of squared residuals.

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} (e^{wx_i} - y_i)^2$$

b. Using gradient descent algorithm to find the optimal w, how will w be updated?

The algorithm involves computing gradient of the loss function with respect to w and updating it in the opposite direction of the gradient.

1. Compute the partial derivative of the SSR with respect to w ($using\ the\ chain\ rule$):

$$\frac{\partial(SSR)}{\partial w} = \frac{\partial[\sum_{i=1}^{n}(e^{wx_i} - y_i)^2]}{\partial w} = \sum_{i=1}^{n} \frac{\partial((e^{wx_i} - y_i)^2)}{\partial w} = \sum_{i=1}^{n} \frac{\partial(e^{2wx_i} + y_i^2 - 2y_i e^{wx_i})}{\partial w} = 2\sum_{i=1}^{n} x_i e^{2wx_i} - x_i y_i e^{wx}$$

$$\frac{\partial(SSR)}{\partial w} = -2\sum_{i=1}^{n} (y_i - e^{wx_i}) x_i e^{wx_i}$$

2. The gradient descent updates w this way:

$$w_{t+1} = w_t - \alpha \frac{\partial(SSR)}{\partial w} = w_t + 2\sum_{i=1}^{n} (y_i - e^{w_t x_i}) x_i e^{w_t x_i}$$

Where α is the learning rate.

c. In order to minimize the loss function, which of the following expressions best describes the optimal value of w?

1. $\sum_{i=1}^{n} x_i e^{wx_i} = \sum_{i=1}^{n} x_i y_i e^{wx_i}$
2. $\sum_{i=1}^{n} e^{wx_i} = \sum_{i=1}^{n} x_i y_i e^{wx_i}$
3. $\sum_{i=1}^{n} x_i e^{2wx_i} = \sum_{i=1}^{n} x_i y_i e^{wx_i}$ (correct answer)

Optimal w comes from the expression that zeros the $\frac{\partial(SSR)}{\partial w}$ which was calculated in the b section.

$$\frac{\partial(SSR)}{\partial w} = -2\sum_{i=1}^{n}(y_i - e^{wx_i})x_i e^{wx_i} = -2\sum_{i=1}^{n} y_i x_i e^{wx_i} + 2\sum_{i=1}^{n} x_i e^{2wx_i}$$

$$\frac{\partial(SSR}{\partial w} = 0 \rightarrow -2\sum_{i=1}^{n} y_i x_i e^{wx_i} + 2\sum_{i=1}^{n} x_i e^{2wx_i} = 0 \rightarrow \sum_{i=1}^{n} y_i x_i e^{wx_i} = \sum_{i=1}^{n} x_i e^{2wx_i}$$