Machine Learning Foundations/hw3

連少甫/r07922107

January 4, 2019

1.

2.

根據第四張投影片所提到之PLA的權重更新方式為：

當$sign(w^Tx) \neq sing(y)$時需要更新w的權重，這等價於$sign(w^Txy) < 0$時更新權重。

因此當我們的$err(w) = max(0, -yw^Tx)$時，當某點分類正確時，其error為0，反之為$-yw^Tx$，接著我們嘗試將$-yw^Tx$對w進行微分，可得$\frac{\partial(-yw^Tx)}{\partial w} = -yx$。因此若使用SGD時更新公式為：

$$w_{t+1} = w_t + learningrate * -yx$$

所以可以將PLA視為：$err(w)$為$max(0, -yw^Tx)$的SGD

3.

根據地16題已知$E_{in} = \frac{1}{N}\sum_{n=1}^{N}(ln(\sum_{i=1}^{k} exp(w_i^Tx_n)) - w_{yn}^Tx_n)$

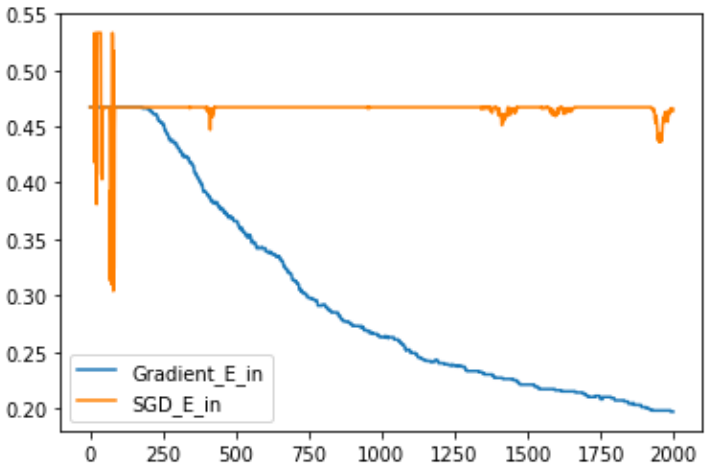
$$\frac{\partial E_{in}}{\partial w_i} = \frac{\partial \frac{1}{N}\sum_{n=1}^{N}(ln(\sum_{i=1}^{k} exp(w_i^Tx_n)) - w_{yn}^Tx_n)}{\partial w_i} = \frac{1}{N}\sum_{n=1}^{N}(\frac{\partial(ln(\sum_{i=1}^{k} exp(w_i^Tx_n)) - w_{yn}^Tx_n)}{\partial w_i})$$

$$= \frac{1}{N}\sum_{n=1}^{N}(\frac{\partial(ln(\sum_{i=1}^{k} exp(w_i^Tx_n))}{\partial w_i} - \frac{\partial w_{yn}^Tx_n}{\partial w_i}) = \frac{1}{N}\sum_{n=1}^{N}(\frac{exp(w_i^Tx_n)x_n}{\sum_{i=1}^{k} exp(w_i^Tx_n)} - \frac{\partial w_{yn}^Tx_n}{\partial w_i})$$

$$=\frac{1}{N}\sum_{n=1}^{N}\left(\frac{exp(w_i^T x_n)x_n}{\sum_{i=1}^{k} exp(w_i^T x_n)} - [[y_n = i]]x_n\right) = \frac{1}{N}\sum_{n=1}^{N}(h_i(x_n) - [[y_n = i]]x_n)$$

4.

根據下圖所示，藍色線為gradient decent在training data 中的error曲線(learning rate=0.01)，橘色線為SGD在training data 的error曲線(learning rate=0.001)，由圖中可以看到由於learning rate相差十倍的關係，導致error下降幅度的差異很大，所以learning rate在機器學習中扮演著相當重要的因素。



5.

根據下圖所示，藍色線為gradient decent在testing data 中的error曲線(learning rate=0.01)，橘色線為SGD在testing data 的error曲線(learning rate=0.001)， 可以發現testing呈現和training訓練時相似的成果。這代表這兩個群資料是很相似的。