

Machine Learning Foundations/hw3

連少甫/r07922107

January 4, 2019

1.

The screenshot shows a Coursera interface. At the top, there's a search bar and a user profile icon labeled 'lien,shaofu'. Below the navigation bar, the course title '此課程: 機器學習基石下 (Machine Learning Foundations)---Algorithmic Foundations' is displayed. On the left, a sidebar lists several quizzes: 'Quadratic Hypothesis' (23 min), 'Nonlinear Transform' (9 min), 'Price of Nonlinear Transform' (15 min), 'Structured Hypothesis Sets' (9 min), and '作業三' (20 個問題), which is currently selected and marked with a green checkmark. The main content area shows the '作業三' quiz results. It indicates a score of 100.00% and states '我們會保留您的最高分數。' (We will keep your highest score.) and '查看最新提交內容' (View latest submission). A blue button labeled '再次參加' (Retake) is visible. At the bottom right of the quiz area, there are icons for sharing, commenting, and printing.

2.

根據第四張投影片所提到之PLA的權重更新方式為：

當 $\text{sign}(w^T x) \neq \text{sign}(y)$ 時需要更新 w 的權重，這等價於 $\text{sign}(w^T x y) < 0$ 時更新權重。

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

$$w_{t+1} = w_t + \text{learningrate} * -y x$$

所以可以將PLA視為： $\text{err}(w)$ 為 $\max(0, -y w^T x)$ 的SGD

3.

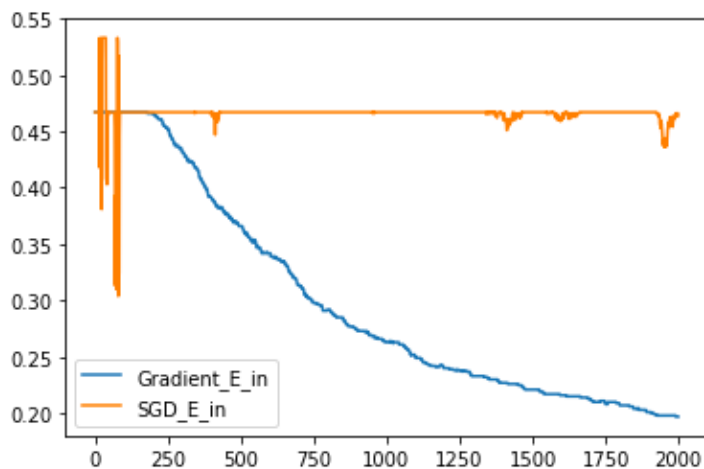
根據地16題已知 $E_{in} = \frac{1}{N} \sum_{n=1}^N (\ln(\sum_{i=1}^k \exp(w_i^T x_n)) - w_{yn}^T x_n)$

$$\begin{aligned} \frac{\partial E_{in}}{\partial w_i} &= \frac{\partial \frac{1}{N} \sum_{n=1}^N (\ln(\sum_{i=1}^k \exp(w_i^T x_n)) - w_{yn}^T x_n)}{\partial w_i} = \frac{1}{N} \sum_{n=1}^N \left(\frac{\partial (\ln(\sum_{i=1}^k \exp(w_i^T x_n)) - w_{yn}^T x_n)}{\partial w_i} \right) \\ &= \frac{1}{N} \sum_{n=1}^N \left(\frac{\partial (\ln(\sum_{i=1}^k \exp(w_i^T x_n)))}{\partial w_i} - \frac{\partial w_{yn}^T x_n}{\partial w_i} \right) = \frac{1}{N} \sum_{n=1}^N \left(\frac{\exp(w_i^T x_n) x_n}{\sum_{i=1}^k \exp(w_i^T x_n)} - \frac{\partial w_{yn}^T x_n}{\partial w_i} \right) \end{aligned}$$

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

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。

