機器學習 HW1　0751231　曾揚

Q1.

If use MSE as loss function :

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i(x) - y_i\right)^2$$

$$\nabla_\theta J = \nabla_\theta\left(\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2\right)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\nabla_\theta\left(\hat{y}_i - y_i\right)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m}2\left(\hat{y}_i - y_i\right)\nabla_\theta\left(\hat{y}_i - y_i\right)$$ ....常係數可忽略，放在 learning rate 中

$$※\quad \nabla_\theta\left(\hat{y}_i - y_i\right) = \nabla_\theta \hat{y}_i = \frac{\prod_{i=1}^{n}\mu_{A_i^l}(x_i)}{\sum_{l=1}^{M}\left(\prod_{i=1}^{n}\mu_{A_i^l}(x_i)\right)}\begin{bmatrix}\theta^1\\ \theta^2\\ \theta^3\\ \vdots\\ \theta^M\end{bmatrix}$$

$$\Rightarrow \theta_t \leftarrow \theta_{t-1} - \alpha\cdot 2\left(\hat{y} - y\right)\cdot\frac{\prod_{i=1}^{n}\mu_{A_i^l}(x_i)}{\sum_{l=1}^{M}\left(\prod_{i=1}^{n}\mu_{A_i^l}(x_i)\right)}\cdot\theta_{t-1}$$

\#

Q2.(d):

根據 Q2(b) and Q2(c)的結果，我們可以很明顯地看到，在使用單一特徵時做 Linear Regression 之 RMSE 的排序與各個特徵對於 Label 的相關係數排序順序相似度非常高，甚至在前段 RMSE 較小的區段幾乎完全一模一樣。表示用單一 feature 的 Linear Regression 之 RMSE 在一定程度上可以表示出該 feature 對於

label 的影響程度(或相關性)。