機器學習 HW1

0610719 賴知榆

1. Gradient Descent: (TS fuzzy model)

$$y_f(x)=\frac{\sum_{l=1}^{M}\theta^l\left(\prod_{i=1}^{n}\mu_{A_i^l}(x_i)\right)}{\sum_{l=1}^{M}\left(\prod_{i=1}^{n}\mu_{A_i^l}(x_i)\right)}, \quad \text{Let } \prod_{i=1}^{n}\mu_{A_i^l}(x_i)=K(l)$$

$$y_f(x)=\frac{\sum_{l=1}^{M}\theta^l\cdot K}{\sum_{l=1}^{M}K}$$, we MSE cost function

$$J(\theta)=\frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x_i)-y_i\right)^2, \quad \text{here } h_\theta(x)=y_f(x)$$

Target update law: $\theta^l := \theta^l-\alpha\cdot\frac{\partial}{\partial\theta^l}J(\theta)$, α: step size, $l=1,2,\dots,M$

$\Rightarrow$ so we find $\frac{\partial}{\partial\theta^l}J(\theta)=\frac{1}{m}\sum_{i=1}^{m}\left(y_f(x)-y_i\right)\cdot\frac{\partial}{\partial\theta^l}\left(y_f(x)-y_i\right)$

$$\Rightarrow \frac{\partial}{\partial\theta^l}y_f(x)=\frac{\partial}{\partial\theta^l}\frac{\sum_{l=1}^{M}\theta^l\cdot K}{\sum_{l=1}^{M}K}=\frac{K(l)}{\sum_{l=1}^{M}K(l)}, \quad l=1,2,\dots,M$$

$$=\frac{\prod_{i=1}^{n}\mu_{A_i^l}(x_i)}{\sum_{l=1}^{M}\left(\prod_{i=1}^{n}\mu_{A_i^l}(x_i)\right)}$$

$$\Rightarrow \theta^l := \theta^l-\frac{\alpha}{m}\sum_{i=1}^{m}\left(y_f(x)-y_i\right)\cdot\frac{\prod_{i=1}^{n}\mu_{A_i^l}(x_i)}{\sum_{l=1}^{M}\left(\prod_{i=1}^{n}\mu_{A_i^l}(x_i)\right)} \quad \#$$

2. (d) Compare the results in ranking of part (b) and (c), what do you observe?

從結果來看，可以發現 B 和 C 的 Ranking 結果一模一樣。以 X-Y 平面的單一 X 對單一 Y 來解釋。在 B 小題我們所做的事情就是用一個 feature 對應一個 label，使用 linear regression 求出一條最符合(X,Y)分布的直線。而在這邊，RMSE 越小，代表(X,Y)分布的趨勢越明顯，越符合我們求出的直線。

而 C 小題叫我們求出每一個 feature 和 label 的皮爾森相關係數。皮爾森相關係數用於度量兩個變數 X 和 Y 之間的相關程度（線性相依），其值介於-1 與 1 之間。在自然科學領域中，該係數廣泛用於度量兩個變數之間的線性相依程度。這個相關係數也稱作「皮爾森相關係數 r」。

由此可知，我們從 B 小題得到的結論其實和 C 小題在尋找的東西是一樣的。RMSE 越小，代表趨勢越明顯，X 與 Y 的相關性越高。反之，相關係數越高，則趨勢越明顯，RMSE 也就越小。