

結構化機器學習模型及其應用作業一

統研碩一 7107018013 郭又嘉

I. 問題與目標

這次的題目是要如何用 MLP 來做 regression 問題，目標要利用波士頓房地產資料庫來預測房價。

II. 資料讀取與前處理

我們使用從 sklearn.datasets 中得到的波士頓房地產資料庫，在這當中有 506 筆資料及 13 個變數，而目標 ($y^{(i)}$, $i = 1 \dots 506$) 是自行擁有房屋的房價，對它取中位數 (單位：1000 美元)，稱做 MEDV。我們現在的目標是要對房價做預測，而在這之前先對資料做標準化，再以 70/30 來切割 training 和 testing data。

III. 3-LAYER MLP 模型理論架構

對於這次的預測房價，我以 3-layer MLP 來解決 regression 問題，我所設置的架構是會有兩個 Affine 層，中間經過一個激活函數 (Sigmoid)，而此輸出就是我所預測的房價 ($Z^{(out)}$)，最後經過 cost function 可再對 training 和 testing data 做比較，查看是否會有 overfitting 的問題。

定義 Affine 層是 $y = x \cdot w + b$ ，Sigmoid 層是 $\phi = \frac{1}{1+\exp(-x)}$ ，cost function $J = \sum_{i=1}^n \|\hat{y}_i - y_i\|^2$ 。

IV. 前後傳播推導

運算 3-layer MLP 模型架構的 Forward 及 Backward Propagation 的推導。定義變數： $X^{(in)}$ 為輸入的資料， $W^{(h)}$ 、 $W^{(out)}$ 、 $B^{(h)}$ 、 $B^{(out)}$ 為 MLP 當中的權重與偏誤值， Y 為輸出值。

A. Forward Propagation

- 第一 Affine 層

$$Z^{(h)} = (X^{(in)} \cdot W^{(h)}) + B^{(h)}$$

- Sigmoid 層

$$A^{(h)} = \phi(Z^{(h)})$$

- 第二 Affine 層

$$Z^{(out)} = (A^{(h)} \cdot W^{(out)}) + B^{(out)}$$

- cost function

$$J = \sum_{i=1}^n (y_i - z_i^{(out)})^2$$

B. Backward Propagation

- cost function 的反傳播

$$Y - Z^{(out)}$$

- 第二 Affine 層的反傳播

$$\frac{\partial J(w)}{\partial W^{(out)}} = [A^{(h)}]^T \cdot (Y - Z^{(out)})$$

$$\frac{\partial J(w)}{\partial A^{(h)}} = (Y - Z^{(out)}) \cdot [W^{(out)}]^T$$

$$\frac{\partial J(w)}{\partial B^{(out)}} = Y - Z^{(out)}$$

- Sigmoid 層的反傳播

$$\frac{\partial J(w)}{\partial Z^{(h)}} = \frac{\partial J(w)}{\partial A^{(h)}} (1 - A^{(h)}) A^{(h)}$$

- 第一 Affine 層的反傳播

$$\frac{\partial J(w)}{\partial W^{(h)}} = [X^{(in)}]^T \cdot \frac{\partial J(w)}{\partial Z^{(h)}}$$

$$\frac{\partial J(w)}{\partial X^{(in)}} = \frac{\partial J(w)}{\partial Z^{(h)}} \cdot [W^{(h)}]^T$$

$$\frac{\partial J(w)}{\partial B^{(h)}} = \frac{\partial J(w)}{\partial Z^{(h)}} = \frac{\partial J(w)}{\partial A^{(h)}} (1 - A^{(h)}) A^{(h)}$$

V. 訓練模型

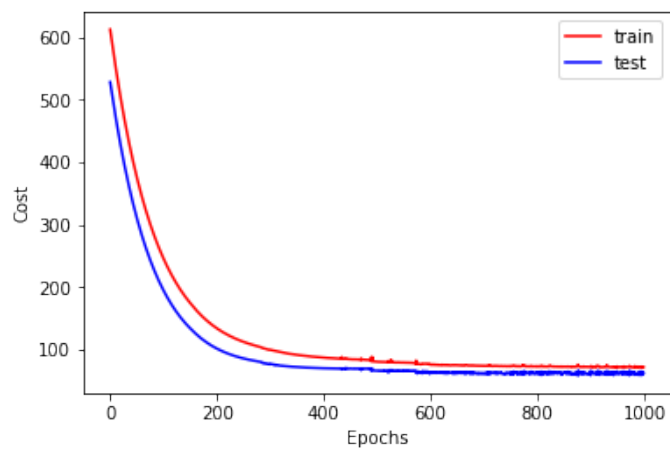
用 python 實作 3-layer MLP，在這之中我設定 hidden layer 為 20、output size 為 1、learning rate 為 1e-6，training 和 testing data 分別做迭代次數為 1000 次的運算。

VI. 結果評估

用 training data 所算出的 MSE 為 71.363、testing data 為 61.745。

用 training 和 testing data 所做出來的 cost 為下列圖所示，可看出兩者收斂速度非常快，線段

也幾乎相近，由此可知沒有 overfitting 的問題。



VII. 結論

訓練完模型之後，可預測出波士頓的房價，就是上面的 $Z^{(out)}$ ，由 MSE 可知 $Z^{(out)}$ 與真實值 Y 相差不大，所以在波士頓房地產資料庫下，3-layer MLP 模型的預測能力是好的。