結構化第一次作業

統研一 7107018025 張維浩

I. 簡介

這一個迴歸預測模型，資料來自 boston 的房價預測，資料共有 506 筆有 13 個變數，Fig1 從左至右分別是輸入層、一個隱藏層含有 30 個結點，隱藏層活化函數使用 sigmoid 函數、與輸出層，為三層的 MLP 模型。取全部資料的 7 成做為訓練資料共計 354 筆；取 3 成做為測試資料共計 152 筆。迭代次數 5000 次、學習率 0.005，以 SSE 做為誤差函數、R^2 做為預測準度的依據，訓練後得到 Cost=2823、$R^2 = 0.9017$。

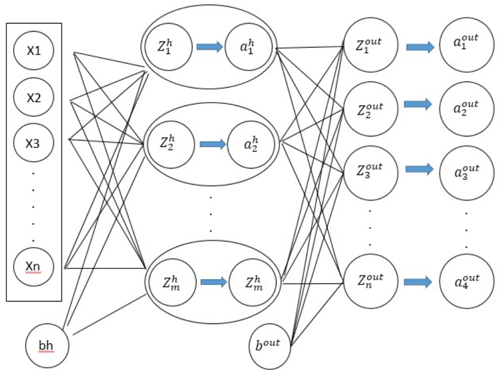

Fig. 1. 迴歸預測模型

II. 活化函數

隱藏層的活化函數使用 sigmoid 函數

$f(x) = \frac{1}{1+exp(-x)}$

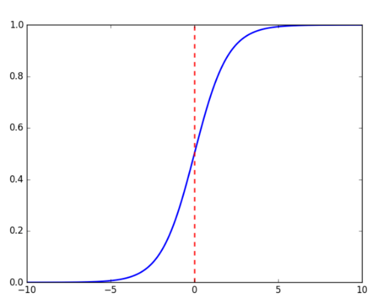

Fig. 2. sigmoid 函數

輸出層的活化函數使用恆等函數等同於輸出層不使用活化函數。
$f(x) = x$

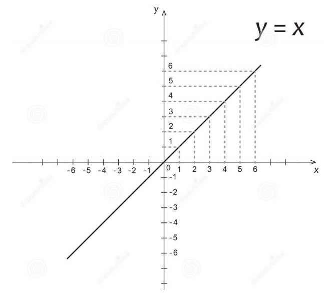

Fig. 3. 恆等函數

III. 正向傳播

輸入層到隱藏層:

$X \odot W^h + b^h = Z^h$

隱藏層的輸出 (活化函數為 sigmoid):

$sigmoid(Z^h) = a^h$

隱藏層到輸出層:

$a^h \odot W^{out} + b^{out} = Z^{out}$

輸出層 (活化函數為恆等含數):

$a^{out} = Z^{out}$

IV. 反向傳播

對誤差進行反向傳播，更新權重 W

誤差函數用 $SSE = \sum_i (\widehat{y^i} - y^i)^2$

$\delta^{out} = \frac{\partial SSE}{\partial Z^{out}} = 2 * (Z^{out} - y)$

$\delta^h = \frac{\partial SSE}{\partial Z^h} = W^{out}\delta^{out} * a^h(1 - a^h)$

$\frac{\partial SSE}{\partial W^{out}} = a^h \delta^{out}$

$\frac{\partial SSE}{\partial W^h} = a^{in} \delta^h$

$\frac{\partial SSE}{\partial b^{out}} = \delta^{out}$

$\frac{\partial SSE}{\partial b^h} = \delta^h$

使用梯度下降訓練參數

$W^h \to W^h - (A^{in})^T \delta^h$

$W^{out} \to W^{out} - (a^h)^T \delta$

$b^h \to b^h - \sum \delta^h$

$b^{out} \to b^h - \sum \delta^{out}$

V. 結果

透過向前傳播的過程可以得到一個預測值，而預測值和目標值的差距稱做誤差，再來透過反向傳播法不停的迭代更新權重，使所得到的誤差最小。這個地方第一次迭代後得到的誤差 (SSE) 為 180000 左右，經過 5000 次的迭代後誤差 (SSE) 降至 2823。這裡我用 R^2 來當做預測準度的方式，R^2 最後為 0.9017 可以說房價的高低有 9 成受到資料裡 13 個變數的影響，顯示出利用這條迴歸模式來推論房價有相當的可信度。

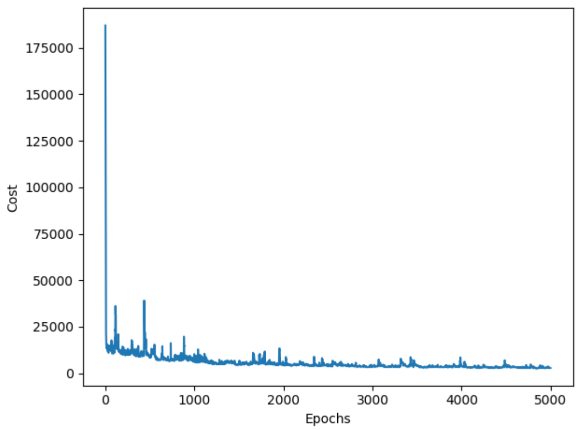

Fig. 4. cost-epochs

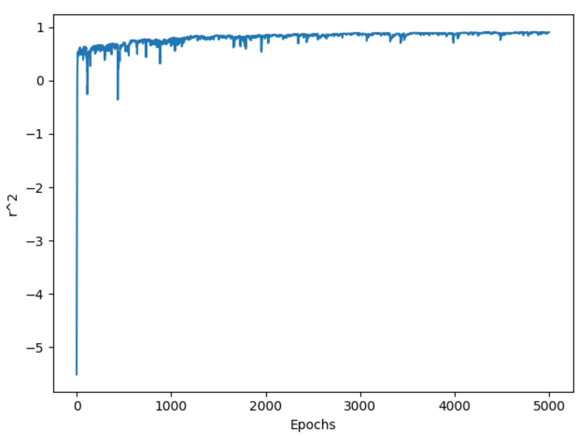

Fig. 5. R^2-epochs

接著我把輸出值跟預測值的圖做出來，加上一條 X=Y 的直線上去如圖所示，發現房價介於 0 至 50 的資料，有不錯的預測結果，當房價高於 50 其預測效果不佳，猜測 1: 高房價的樣本數太少導致此結果，猜測 2: 高房價的房屋有資料變數外的原因來影響房價，或是人為炒作的結果。