

Machine Learning Homework

Yu-Heng Lee

Department of Applied Mathematics

National Chung Hsing University

Taichung, Taiwan

I. REGRESSION PROBLEM

使用Regression來訓練neural network model去預估house dataset。

II. BACKWARD SCHEMES

我們的neural network model 總共有m層，所以我們會有k=m-1個W
MSE loss:

$$\begin{aligned} J(W) &= \sum_i \left\| \hat{y}^{(i)} - y^{(i)} \right\|^2 \\ &= Tr \left((\hat{Y} - Y)^T (\hat{Y} - Y) \right) \end{aligned}$$

1.我們首先推導第k層的W微分:

$$\begin{aligned} \frac{\partial J(W)}{\partial W^k} &= \frac{\partial Tr \left((\hat{Y} - Y)^T (\hat{Y} - Y) \right)}{\partial W^k} \\ &= \frac{\partial Tr \left((\hat{Y} - Y)^T (\hat{Y} - Y) \right)}{\partial (\hat{Y} - Y)} \frac{\partial (\hat{Y} - Y)}{\partial W^k} \\ &= 2 (\hat{Y} - Y) \frac{\partial (\hat{Y} - Y)}{\partial W^k} \\ &= 2 (\hat{Y} - Y) \frac{\partial (\hat{Y} - Y)}{\partial \hat{Y}} \frac{\partial \hat{Y}}{\partial W^k} \\ &= 2 (\hat{Y} - Y) \frac{\partial \hat{Y}}{\partial W^k} \\ &= 2 (\hat{Y} - Y) \frac{\partial (W^k A^{(m)})}{\partial W^k} \\ &= 2 A^{(m)} \delta^{(k)} \end{aligned} \quad (1)$$

2.再來推導第k-1層的W微分:

$$\begin{aligned} \frac{\partial J(W)}{\partial W^{k-1}} &= \frac{\partial Tr \left((\hat{Y} - Y)^T (\hat{Y} - Y) \right)}{\partial (\hat{Y} - Y)} \frac{\partial (\hat{Y} - Y)}{\partial W^{k-1}} \\ &= 2 (\hat{Y} - Y) \frac{\partial (\hat{Y} - Y)}{\partial \hat{Y}} \frac{\partial \hat{Y}}{\partial W^{k-1}} \\ &= 2 (\hat{Y} - Y) \frac{\partial W^{out} A^{(m-1)}}{\partial A^{(m-1)}} \frac{\partial A^{(m-1)}}{\partial W^{k-1}} \\ &= 2 W^{out} (\hat{Y} - Y) \frac{\partial (\phi(Z^{m-1}))}{\partial W^{k-1}} \\ &= 2 W^{out} (\hat{Y} - Y) \frac{\partial (\phi(Z^{m-1}))}{\partial Z^{m-1}} \frac{\partial A^{(m-1)} W^{k-1}}{\partial W^{k-1}} \\ &= 2 W^{out} A^{(m-1)} (\hat{Y} - Y) \frac{\partial (\phi(Z^{m-1}))}{\partial Z^{m-1}} \\ &= 2 W^{out} A^{(m-1)} (\hat{Y} - Y) \cdot (\phi(Z^{m-1})) (1 - \phi(Z^{m-1})) \\ &= 2 A^{(m-1)} W^k \delta^{(k)} \cdot (\phi(Z^{m-1})) (1 - \phi(Z^{m-1})) \\ &= 2 A^{(m-1)} \delta^{(k-1)} \end{aligned} \quad (2)$$

3.推導第k-2層的W微分

$$\begin{aligned} \frac{\partial J(W)}{\partial W^{k-2}} &= \frac{\partial Tr \left((\hat{Y} - Y)^T (\hat{Y} - Y) \right)}{\partial (\hat{Y} - Y)} \frac{\partial (\hat{Y} - Y)}{\partial W^{k-2}} \\ &= 2 (\hat{Y} - Y) \frac{\partial (\hat{Y} - Y)}{\partial \hat{Y}} \frac{\partial \hat{Y}}{\partial W^{k-2}} \\ &= 2 (\hat{Y} - Y) \frac{\partial W^{out} A^{(m-1)}}{\partial A^{(m-1)}} \frac{\partial A^{(m-1)}}{\partial W^{k-2}} \\ &= 2 W^{out} (\hat{Y} - Y) \frac{\partial (\phi(Z^{m-1}))}{\partial W^{k-2}} \\ &= 2 W^{out} (\hat{Y} - Y) \frac{\partial (\phi(Z^{m-1}))}{\partial Z^{m-1}} \frac{\partial A^{(m-1)} W^{k-1}}{\partial W^{k-2}} \\ &= 2 \delta^{(k-1)} \frac{\partial A^{(m-1)} W^{k-1}}{\partial A^{(m-1)}} \frac{\partial A^{(m-1)}}{\partial W^{k-2}} \\ &= 2 W^{k-1} \delta^{(k-1)} \frac{\partial \phi(Z^{m-2})}{\partial W^{k-2}} \\ &= 2 W^{k-1} \delta^{(k-1)} \frac{\partial \phi(Z^{m-2})}{\partial Z^{m-2}} \frac{\partial A^{(m-2)} W^{k-2}}{\partial W^{k-2}} \\ &= 2 A^{(m-2)} W^{k-1} \delta^{(k-1)} \cdot (\phi(Z^{m-2})) (1 - \phi(Z^{m-2})) \\ &= 2 A^{(m-2)} \delta^{(k-2)} \end{aligned} \quad (3)$$

4.

III. HOUSE DATA SET

我們對House data set中($X = \text{crim, zn, indus, chas, nox, rm, age, dis, rad, tax, ptratio, black, lstat}$)及 $\text{medv}(y)$ 做標準化。

IV. EXPERIMENTAL RESULTS

我們用了三種不同層數不同節點的neural network model

A. 3 Layer Neural Network

隱藏層的節點設30，epochs = 2000, eta = 0.002

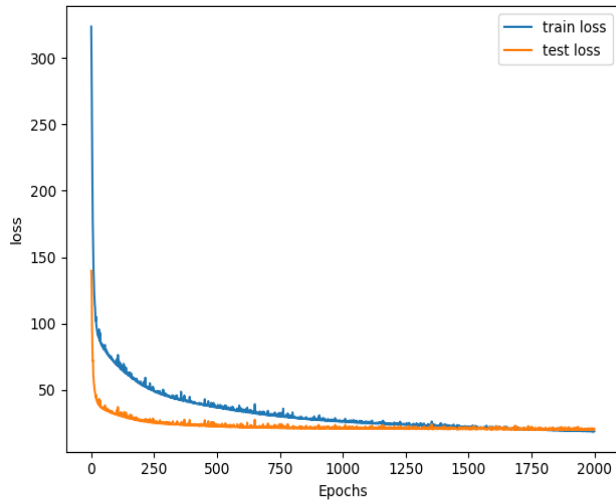


Fig. 1. 3 Layer Neural Network

B. 5 Layer Neural Network

隱藏層的節點設[30,30,30]，epochs = 2000, eta = 0.01

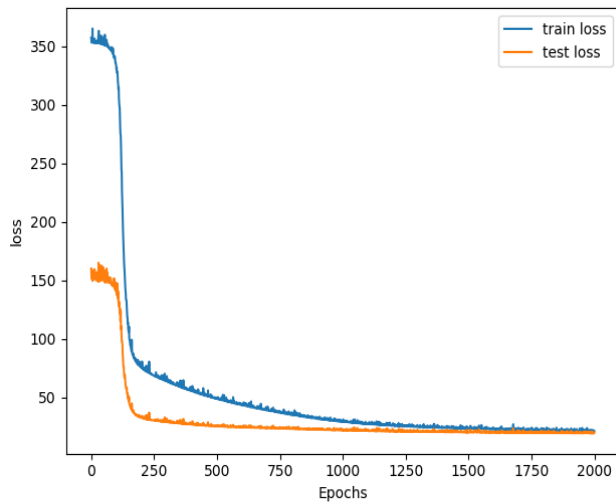


Fig. 2. 5 Layer Neural Network

C. 6 Layer Neural Network

隱藏層的節點設[10,10,10,20]，epochs = 2000, eta = 0.002

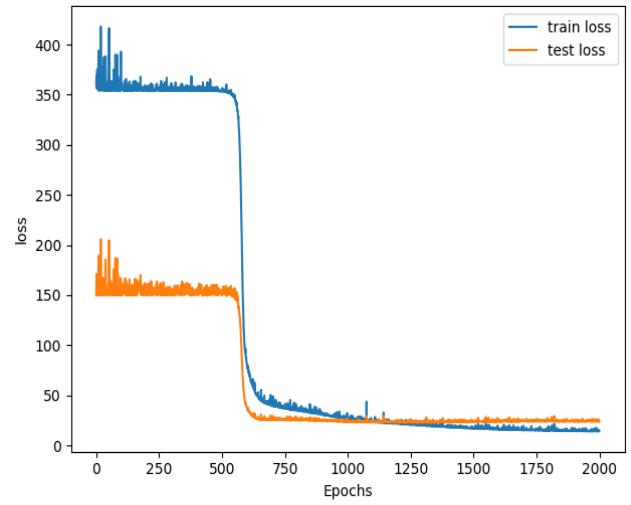


Fig. 3. 6 Layer Neural Network