

Homework 1

葉政義

I. EXPERIMENT RESULTS

建立 Multilayer perceptron 對 Boston housing dataset 進行預測；預測目標，自家用的住宅房價中位數 (medv)，且 對所有的 feature 進行標準化。Equation 1 是 cost function， 使用 Mean Squared Error(MSE) 進行模型比較。

$$J(w) = \sum_i \|\hat{y}^{(i)} - y^{(i)}\|^2 \quad (1)$$

A. The number of hidden layers

learning rate(lr) = 1e-6, hidden layer = 3 or 5, neurons = 30(in each layer), activation function = ReLU, p = 0.5, which is dropout rate.

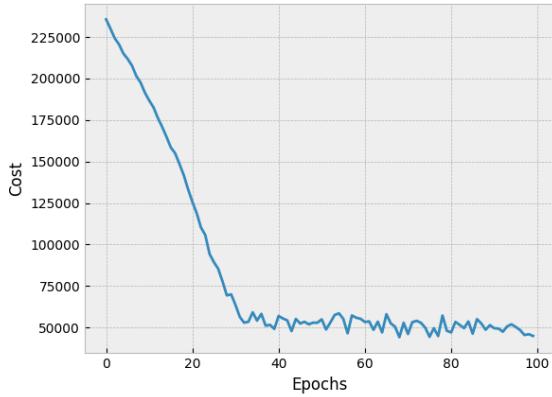


Fig. 1. Cost for 3-hidden layer.

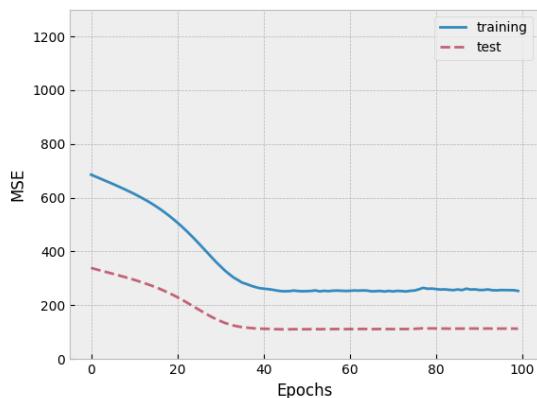


Fig. 2. MSE for 3-hidden layer.

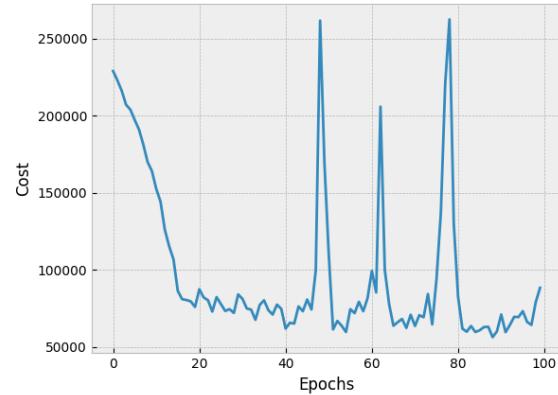


Fig. 3. Cost for 5-hidden layer.

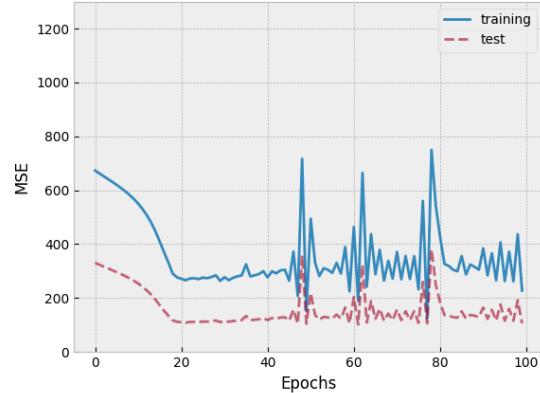


Fig. 4. MSE for 5-hidden layer.

從 Fig. 3 和 Fig. 4 得知，3-hidden layer 和 5-hidden layer 的 MSE 彼此差不多且都無 overfitting 的現象，但是 5-hidden layer 在第 40 個 epoch 後 MSE 突然上下起伏，觀察 Fig. 3 cost 也在這個時候有同一個現象，很有可能是在這時的 learning rate 太高所導致。

B. Without Dropout

learning rate(lr) = 1e-6, hidden layer = 3, neurons = 30(in each layer), activation function = ReLU, p = 0.5, which is dropout rate.

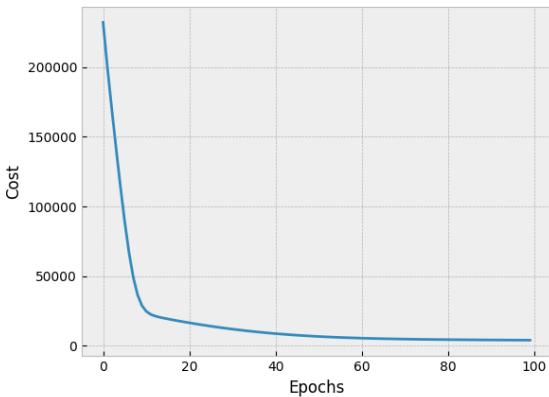


Fig. 5. Cost for 3-hidden layer without dropout.

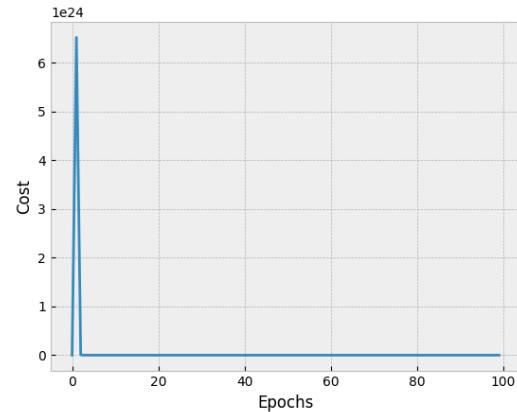


Fig. 7. Cost for 3-hidden layer, $lr = 1e-3$.

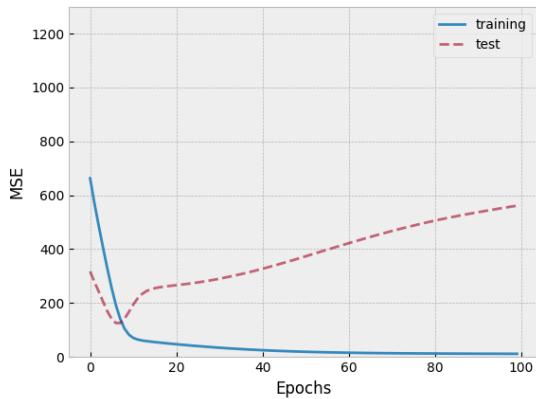


Fig. 6. MSE for 3-hidden layer without dropout.

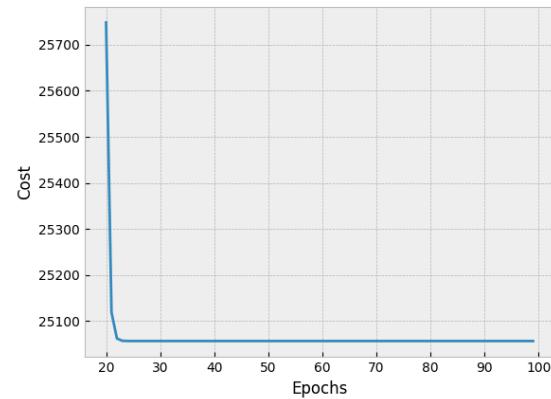


Fig. 8. MSE for 3-hidden layer, $lr = 1e-3$, from 20 epochs to 100 epochs.

不加 dropout 的模型, cost 能降到接近 0 但是由 Fig. 6 可以明顯的看到 overfitting, 也就是 training MSE 很低但是 test MSE 變高; 除此 Fig. 5 和 Fig. 1 相比, 不加 dropout 的模型 cost 比較平滑; 然而加了 dropout 的模型就比較曲折, 原因是在每次計算 cost 時使用的神經元都不同導致 cost 有高有低但是整體來看是有逐漸降低的。

C. Learning rate

learning rate(lr) = $1e-3$, hidden layer = 3, neurons = 30(in each layer), activation function = ReLU, $p = 0.5$, which is dropout rate.

從 Fig. 7 和 Fig. 1 可以得知, 當 learning rate = $1e-3$ 時, 前面幾個 epochs 的 cost 就會很大; 會這樣子的原因是 learning rate 太大, 有時 cost 就會在這時候顯示 inf 或者 nan。

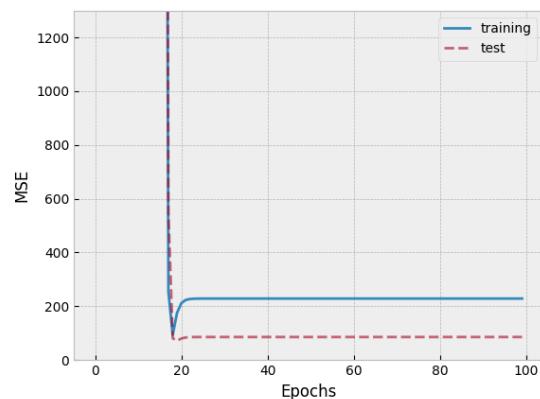


Fig. 9. MSE for 3-hidden layer, $lr = 1e-3$.

II. CONCLUSION

從上面這些圖可以知道對於這筆資料，層數的增加對 MSE 不會有明顯降低而使用 dropout 的模型在預測上效果很好，降低了 overfitting，learning rate 的選擇也很重要，過大的 learning rate 會讓 cost 無法降下來。