Part I. Regression Problem

1.Neural Network Structure:

Hidden layers: 2

# of neurons:

inputlayer:17 input, 10output

hidden layer:10 input, 16 output

output layer:16 input 1, 1 output.

Epoch number:1500

Learning rate:0.001

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

(2)Learning Curve一張含有 文字, 螢幕擷取畫面, 行, 圖表 的圖片

自動產生的描述

(3) (4)Training RMS error and Test RMS error

一張含有 文字, 字型, 螢幕擷取畫面, 行 的圖片

自動產生的描述

(5) (6)Regression result with training labels and test labels

一張含有 文字, 螢幕擷取畫面, 繪圖, 行 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 繪圖, 行 的圖片

自動產生的描述

(c)Feature Selection Procedure:使用sklearn套件計算Fisher’ s score作為feature

selection的依據。Fisher method本身為between-class variance 除以 within class

variance，可以視為一個參考指標衡量各個feature對預測結果(heating load)的貢

獻，在這次作業中我選擇Fisher Score較高的特徵作為模型輸入，試著看能不能

展現出優於使用所有feature的RMS error。

一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

由結果顯示，由fisher score最高的14個features組合所構成的training set 可

以打敗使用17個features的test RMSE 4.569，而且使用8，10，12、14

個feature都優於原先使用的17個features。

以下列出挑選出的feature set: (Glazing\_Area\_Distribution\_0.0和Orientation\_2.0

表示one hot encoding part的結果)

8 features: ['# Relative Compactness' 'Surface Area' 'Wall Area' 'Roof Area'

'Overall Height' 'Glazing Area' 'Cooling Load'

'Glazing\_Area\_Distribution\_0.0']

10 features: ['# Relative Compactness' 'Surface Area' 'Wall Area' 'Roof Area'

'Overall Height' 'Glazing Area' 'Cooling Load' 'Orientation\_2.0'

'Glazing\_Area\_Distribution\_0.0' 'Glazing\_Area\_Distribution\_1.0']

12 features: ['# Relative Compactness' 'Surface Area' 'Wall Area' 'Roof Area'

'Overall Height' 'Glazing Area' 'Cooling Load' 'Orientation\_2.0'

'Orientation\_4.0' 'Glazing\_Area\_Distribution\_0.0'

'Glazing\_Area\_Distribution\_1.0' 'Glazing\_Area\_Distribution\_3.0']

14 features: ['# Relative Compactness' 'Surface Area' 'Wall Area' 'Roof Area'

'Overall Height' 'Glazing Area' 'Cooling Load' 'Orientation\_2.0'

'Orientation\_4.0' 'Glazing\_Area\_Distribution\_0.0'

'Glazing\_Area\_Distribution\_1.0' 'Glazing\_Area\_Distribution\_3.0'

'Glazing\_Area\_Distribution\_4.0' 'Glazing\_Area\_Distribution\_5.0']

結果顯示僅使用8個fisher score最高的feature，就可以達到與使用所有feature

相近的預測效果，這在優化deep learning model上會有不錯的效果，在運算資

源有限的情況下也可以考慮以此方式降低模型對運算資源的配置要求。

Part II. Classification  
Network Architecture: 1 hidden layer, number of neurons controlled by n\_h

(1)34(input layer)-5(hidden layer)-1(output layer), learning rate = 0.005

(2)learning curve

一張含有 文字, 圖表, 行, 繪圖 的圖片

自動產生的描述

Part(c) Results of plotting the distribution of latent features at different training stage

一張含有 螢幕擷取畫面, 文字, 鮮豔, 圖表 的圖片

自動產生的描述 一張含有 螢幕擷取畫面, 文字, 鮮豔, 圖表 的圖片

自動產生的描述

(1)34(input layer)-10(hidden layer)-1(output layer)

(2)Learning curve

一張含有 文字, 繪圖, 行, 螢幕擷取畫面 的圖片

自動產生的描述

Part(c) Results of plotting the distribution of latent features at different training stage.

一張含有 螢幕擷取畫面, 文字, 圖表, 鮮豔 的圖片

自動產生的描述

一張含有 螢幕擷取畫面, 文字, 鮮豔, 圖表 的圖片

自動產生的描述

(1) 34(input layer)-15(hidden layer)-1(output layer)

(2)Learning Curve

一張含有 文字, 螢幕擷取畫面, 繪圖, 行 的圖片

自動產生的描述

Part(c) Results of plotting the distribution of latent features at different training stage.

一張含有 螢幕擷取畫面, 文字, 圖表, 鮮豔 的圖片

自動產生的描述

一張含有 螢幕擷取畫面, 文字, 圖表, 行 的圖片

自動產生的描述

(3)(4)Training error rate and test error rate

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

Conclusion:

在hidden layer的neuron數上升(5🡪10🡪15)的情況下，training 500次時loss有逐步下降，但在training 20次的時候是差不多的。而在training error rate也可以發現20次都落在0.45，test error rate也都一樣(0.00)，在training 500次以後似乎到15個neuron疑似有overfitting的現象，但test error rate相同所以也無法有明確的結論。而由epoch = 20和epoch = 500時latent features的散佈圖可以發現，在一開始epoch=20的訓練初期階段，latent features有集中的現象，在500次時則明顯分得比較開，雖然仍有重疊的區域，並不是非常漂亮的完全分離分布，但已經看得出明顯有差異，尤其在hidden layer neuron為5和10的時候明顯有不一樣，算是能夠觀察出差異。比較值得提的是，原本使用learning rate = 0.001是看不太出差異，調為0.01後learning curve又過於陡峭，latent feature的分布也沒有明顯的差異，調整為0.005以後發現有較好的結果，顯示這對deep learning neural network是相當重要的一個hyper parameter。