HW1 Report

學號:311512015 姓名:謝元碩

1. **Regression**
2. Network Architecture

* Input layer: 1, hidden layer: 2, output layer: 1
* Neurons of input layers: 16
* Neurons of hidden layers: 10\*10
* Neurons of output layers: 1
* Activation function design: hidden layers和output layer皆使用sigmoid function

1. Training parameters
   * 網路架構中的參數(w, b)使用高斯分布做隨機初始化
   * Learning rate: 1e-8
   * Epoch: 3000
   * Batch-size: training data的樣本數
2. Learning Curve

Loss大約在500個epoch內就逐漸收斂

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

1. Training RMS Error: 2.735003059442172
2. Test RMS Error: 2.914477191109333

* 使用自行設計之網路架構及參數進行訓練，訓練集和測試集的RMS誤差如上，可以看到測試集的誤差比訓練集稍微大一些，而測試集對於我們訓練的model屬於new data，因此測試集誤差較大為合理情形

1. Regression result with training labels

一張含有 藝術 的圖片

描述是以低可信度自動產生

1. Regression result with test labels

一張含有 印刷術, 設計 的圖片

描述是以低可信度自動產生

1. Design a feature selection procedure to find out which input features influence the energy load significantly.

* 為找出對模型影響較大的輸入資訊，我在每次訓練時移除其中一項資訊，使用移除該項資訊之輸入組合訓練該模型(其餘參數相同)，將訓練結果進行誤差的計算，並與原本無刪除的輸入組做比較。
* 可以發現將Glazing Area Distribution刪除後所訓練出之模型，計算出之訓練集和測試集誤差皆比其他情況大，估測效果也較差，因此認為此項目對於模型的訓練影響最大；另外，移除Glazing Area也會造成error上升，最後推測與Glazing Area相關的資訊對於heating load估測影響最大。

|  |  |  |
| --- | --- | --- |
| 刪除項目 | Train error | Test error |
| 無刪除資料 | 2.8358 | 2.9647 |
| Relative Compactness | 2.7496 | 2.7032 |
| Surface Area | 2.9711 | 3.2430 |
| Wall Area | 3.0644 | 3.0375 |
| Roof Area | 2.7663 | 2.8294 |
| Overall Height | 2.7971 | 2.8472 |
| Orientation | 2.8239 | 2.7540 |
| Glazing Area | **3.2130** | **3.4557** |
| Glazing Area Distribution | **4.6583** | **4.5092** |

1. **Classification**
   1. Network Architecture

* Input layer: 1, hidden layer: 2, output layer: 1
* Neurons of input layers: 34
* Neurons of hidden layers: 24\*16
* Neurons of output layers: 2 (將good, bad分類為[1,0], [0,1])
* Activation function design:
  + hidden layers皆使用sigmoid function
  + output layer使用softmax對前一層之輸出x數值做範圍限制，並以機率分布表示，適合用來做分類任務之訓練
  1. Training parameters
* 網路架構中的參數(w, b)使用高斯分布做隨機初始化
* Learning rate: 1e-5
* Epoch: 390
* Batch-size: training data的樣本數
  1. Learning curve
* 下圖分別為loss趨勢圖、使用訓練集之預測結果、使用測試集之預測結果
* 趨勢圖在400 epoch內就已經接近收斂，分類效果呈現在第4, 5點的數據以及下面的預測分布圖。
* 顏色為該資料的實際類別，從圖中可以看到，不同種類的顏色分布重疊度低，無論是訓練集或測試集的資料，分類效果皆較佳。

一張含有 行, 螢幕擷取畫面, 圖表, 繪圖 的圖片

自動產生的描述

一張含有 螢幕擷取畫面, 文字, 圖表, 視覺效果 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

* 1. Training error rate: 0.014285714285714285
  2. Test error rate: 0.05714285714285714
  3. Compare the results of choosing different numbers of nodes in the layer before the output layer by plotting the distribution of latent features at different training stage.
* 本作業除了使用390 epoch訓練，另外使用了三種epoch訓練並觀察比較結果，結果如下：
  + Epoch = 10

Train Error Rate: 0.16071428571428573

Test Error Rate: 0.18571428571428572

因迭代較少，模型尚未找到合適的參數、loss並未收斂，因此分類結果如下圖，兩種類別的資料分布重疊度高，導致分類效果差。

一張含有 行, 螢幕擷取畫面, 圖表, 繪圖 的圖片

自動產生的描述

一張含有 螢幕擷取畫面, 圖表, 視覺效果 的圖片

自動產生的描述

一張含有 螢幕擷取畫面 的圖片

自動產生的描述

* + Epoch = 100

Train Error Rate: 0.0035714285714285713

Test Error Rate: 0.07142857142857142

在100 epoch，loss已逐漸收斂，但估測結果分布成有部分重疊情形。

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

一張含有 螢幕擷取畫面, 圖表, 文字, 視覺效果 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

* + Epoch = 2000

Train Error Rate: 0.0

Test Error Rate: 0.1

使用較高的迭代數座訓練，可以看到使用訓練集的估測結果相當好，估測誤差為0；使用測試集的預測誤差則較高一些，但分類效果仍是較佳的。

因此較高的迭代數可以讓模型達到更好的效果，原本希望嘗試訓練到overfitting，經過多次調整後仍未嘗試出來。

一張含有 文字, 螢幕擷取畫面, Rectangle, 行 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

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