

機器學習專案作業一

摘要

由於傳統的葡萄酒的品質判斷需要非常繁雜的步驟，且需要用到化學的方式來測量，所以我們想利用已經測量完的資料集，來訓練出一個可以根據葡萄酒的化學成分、密度和 pH 值預測出葡萄酒的品質的模型，目的就是為了省去傳統判斷那些繁雜的步驟。而我們使用的方法是用前饋式神經網路去實作一個多類別分類器模型，我們先將資料集的數值欄位做正規化，並對正確解答的欄位做 One hot encoding，再將資料集分成 8 成的訓練集以及 2 成的測試集，之後建立前饋神經網路，並訓練。而最終模型測試集預測的正確率為 0.5694。

壹、緒論

1.1. 動機

由於傳統的葡萄酒的品質判斷需要非常繁雜的步驟，且需要用到化學的方式來測量，所以我們想利用已經測量完的資料集，來訓練出一個可以根據葡萄酒的化學成分、密度和 pH 值預測出葡萄酒的品質的模型。

1.2. 目的

為了能根據葡萄酒的化學成分、密度和 pH 值預測出葡萄酒的品

質，我們想分析化學成分與 pH 之間的關聯以及了解化學成分與品質之間是否存在關聯。由於品質是 0 到 10 之間的整數，所以我們使用多類別分類器。而 pH 值預測的部分則是使用回歸預測。

貳、 方法

程式的架構有四個部分，分別是讀取資料集、資料前處理、建構神經網路以及訓練神經網路。程式有用到 keras、tensorflow 以及 sklearn 套件，只需使用 jupyter 依序執行即可。

參、 實驗

3.1 資料集

我們所使用的資料集名稱是 Wine Quality Data Set，是來自葡萄牙北部的紅色和白色葡萄酒樣本，而我們使用白色葡萄酒的部分。筆數有 4898 筆，資料集總共有 12 個欄位，分別是 fixed acidity、volatile acidity、citric acid、residual sugar、chlorides、free sulfur dioxide、total sulfur dioxide、density、pH、sulphates、alcohol 和 quality，前面 11 個欄位是連續值，最後一個欄位 quality 則是整數類別值。

部分資料內容：

fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6
6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6
8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6
7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6
7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6

3.2 前置處理

由於欄位的值過大可能會影響到神經網路訓練的結果，所以我們對每個連續值欄位都做了正規化，每筆資料都除以整個資料集中該欄位的最大值，所以每個值都會是在 0~1 之間。由於我們要做多類別分類器，所以我們對目標欄位 target 做 one hot encoding。

3.3 實驗設計

多類別分類器的部分剛開始我們先使用了三層隱藏層，寬度分別是 16、32、32，激活函數皆是使用 relu，輸出層激活函數使用 softmax，優化器使用 adam，batch_size 為 16，epoch 為 50，結果測試集的準確度只有 0.5112，後來我們增加隱藏層的深度到四層，並將寬度改為 200、100、50、25，測試集的準確度提高到 0.5694。回歸預測的部分使用了四層隱藏層，寬度分別是 10、32、32、32，激活函數皆是使用 relu，輸出層沒有激

活函數，優化器使用 rmsprop，batch_size 為 16，epoch 為 50。

3.4 實驗結果

最後使用測試集評估模型，多類別分類器的結果是 loss:

1.0239 - accuracy: 0.5694 - f1_m: 0.5277 - precision_m:

0.5943 - recall_m: 0.4764，可以看出結果還是有待加強，不

過誤差並不會到太大，還是能大致分出葡萄酒的品質。回歸預

測的結果是 loss: 0.0174 - mae: 0.1020 - mape: 3.1743 -

rmse: 0.1020，可以看出結果相當不錯。

肆、 結論

由這次實驗的結果可以發現利用葡萄酒的化學成分去預測葡萄酒的品質是可行的，和確定葡萄酒的化學成分跟品質有一定的關聯，只要適當優化模型，就能將模型的準確率提高到可用的程度。

伍、 參考文獻

<https://datascience.stackexchange.com/questions/45165/how-to-get-accuracy-f1-precision-and-recall-for-a-keras-model>

<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

<https://stackoverflow.com/questions/43855162/rmse-rmsle-loss-function-in-keras>