Predicting Wine Quality

2017-2 Machine Learning Term Project

2015104226 YunMin Cho

Index

1. Introduction
2. Preparing data set
3. Training Model
4. Conclusion

Introduction

How do people know wine quality?

In the past, only wine connoisseurs could evaluate the quality.

➡ No one knows why they are different except for the wine connoisseurs.

In the late 1900's, Orley Clark Ashenfelter predicted relationship between wine quality and weather.

$$\begin{aligned} Winequality = {} & 12.145 + 0.00117 \times Winter\,Rainfall \\ & + 0.0614 \times Average\,growingseasontemp \\ & - 0.00386 \times Harvest\,rainfall \end{aligned}$$

와인의 품질 = 12.145 + 0.00117*겨울철 강수량 + 0.0614*재배철의 평균 기온 – 0.00386*수확기 강수량

Introduction

Is it impossible to predict the quality just using the information of wine?

➡ This project predicts wine quality only using the information of wine.

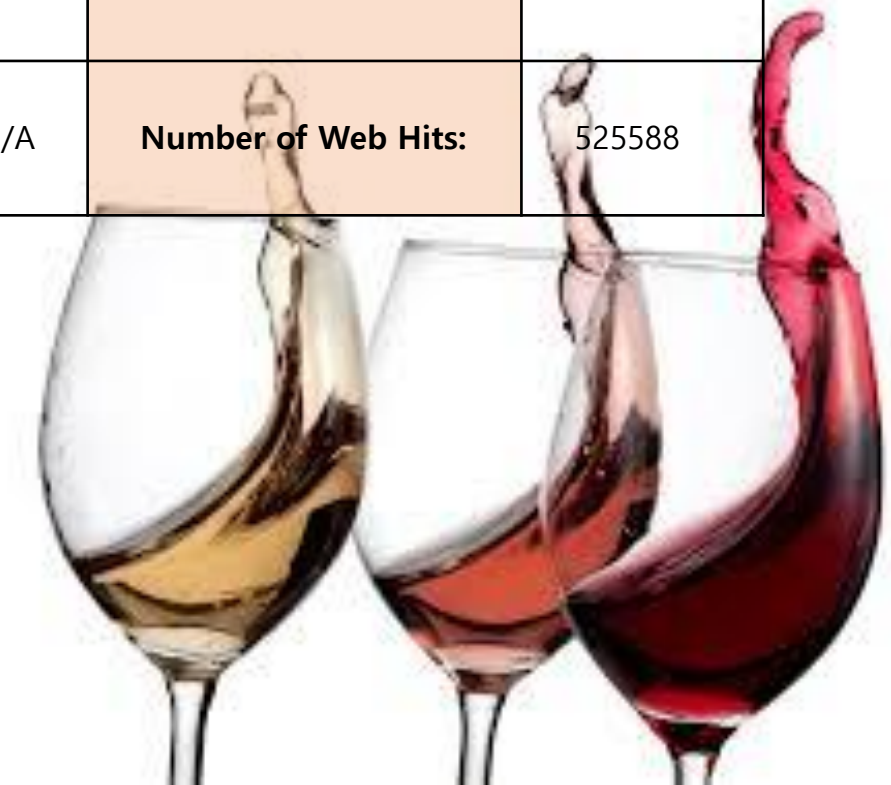
- Using Tensorflow as machine learning tool
- Data set file extension: CSV (I did data normalization.)
- Model: Multilayer Perceptron
- Using Sigmoid as activation function / Using Softmax at last layer
- Using Adam Optimizer as optimization function

Preparing Data Set

About Data Set

Data Set Characteristics:	Multivariate	**Number of Instances:**	4898	**Area:**	Business
Attribute Characteristics:	Real	**Number of attributes:**	12	**Date Donated:**	2009-10-07
Associated Tasks:	Classification, Regression	**Missing Values?**	N/A	**Number of Web Hits:**	525588

Fixed acidity(포도주 결합산) / volatile acidity(휘발성산) / Citric acid(구연산) /
Residual sugar(발효 후 와인 속에 남아있는 당분) / Chlorides(염화물) /
Free sulfur dioxide(유리 이산화황) / Total sulfur dioxide(총 이산화황) /
Density(농도) / pH(산도) / Sulphates(황산염) / Alcohol()알콜올 / Quality(품질)

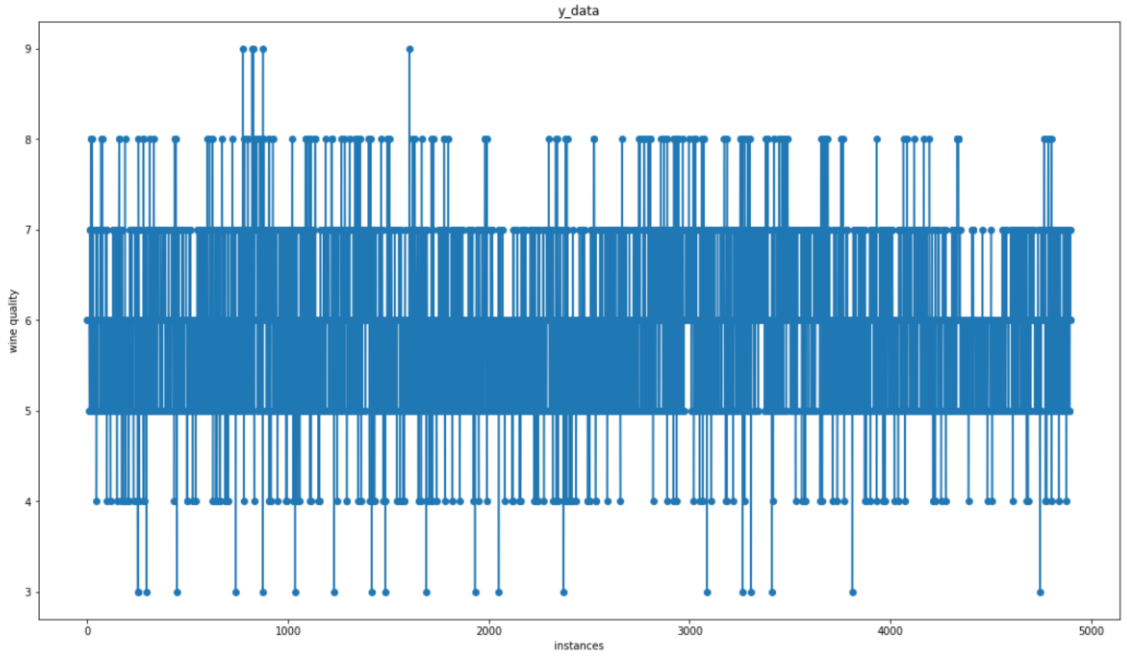


12th attribute is wine quality(score: 0 ~ 10)

→ Bad(0~5) / Normal(6) / Good(7~10)

Preparing Data Set

About Data Set



12th attribute is wine quality(score: 0 ~ 10)

→ Bad(0~5) / Normal(6) / Good(7~10)

Preparing Data Set

About Data Set

Two data files: White wine and Red wine

* There is no difference between the two except number of instances.

White wine: 4898 instances

Red wine: 1599 instances

Preparing Data Set

Data Preprocessing

Original data set is irregular in scope.

* First instance: 7;0.27;0.36;20.7;0.045;45;170;1.001;3;0.45;8.8;6

➡ Data preprocessing is needed.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

<Data normalization>

After normalization, all values of x_data are into the range [0, 1].

Preparing Data Set

Data Preprocessing

<Y_data>

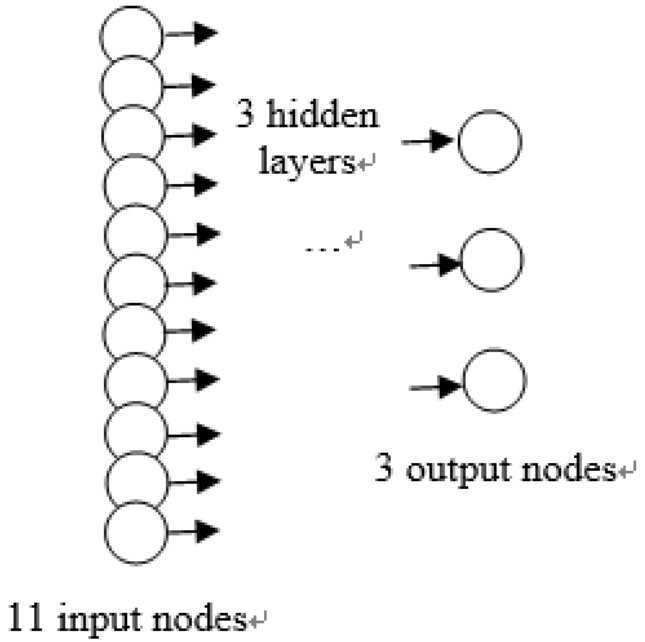
- Bad = 0
- Normal = 1
- Good = 2

Training Model

Multilayer Perceptron

Because I have 3 classes(bad, normal, good), I have to do multinomial classification.

➡ Multilayer Perceptron



Training Model

Training

$$hypothesis = x \times Weight + bias$$

$$C(H(x), y) = y \log(H(x)) - (1-y)\log(1-H(x))$$

Number of hidden layers	3
Activation function	Sigmoid / Softmax(only output layer)
Optimization function	Adam Optimizer
Learning rate	0.001
Training epoch	70

Conclusion

My project conclusion

```
0 12.4788
10 10.3911
20 8.25966
30 6.4207
40 5.27344
50 4.57267
60 3.98362
70 3.51598

Hypothesis:  [[ 0.70092028  0.20902337  0.09005642]
 [ 0.51616484  0.386419    0.09741624]
 [ 0.46361294  0.44086707  0.09552003]
 ...,
 [ 0.57079208  0.33016554  0.09904242]
 [ 0.35265219  0.54712731  0.10022057]
 [ 0.40198338  0.49708837  0.10092824]]
Prediction (Y):  [0 0 0 ..., 0 1 1]
Accuracy:  0.679665
```

Conclusion

Test Model

```
sample:  [[ 0.44884485  0.46769017  0.08346489]
 [ 0.54856944  0.36593911  0.08549138]
 [ 0.51596302  0.3977811   0.08625585]
 ...,
 [ 0.49540278  0.41305935  0.09153782]
 [ 0.50766593  0.4042373   0.08809683]
 [ 0.61535245  0.2963309   0.08831661]] [1 0 0 ..., 0 0 0] 0.628518
```

Conclusion

Related Preceding Research

Solvers	Entry	Last Submission	Leaderboard
jasonc	0.57528	Fri, Jun 3 2011 3:24 AM	1
Konrad	0.5783	Wed, May 25 2011 1:36 PM	2
mvp	0.58884	Wed, Jun 22 2011 6:02 PM	3
Zachary Mayer	0.59085	Sat, May 21 2011 12:37 AM	4

```
pred <- predict(model, newdata = test)
table(pred, test$taste)
pred      bad good normal
  bad     482   10    130
  good     14  252     85
  normal  171  149    667
```

We can test the accuracy as follow

```
(482 + 252 + 667) / nrow(test)
0.7147959
```

Question!


BOSCHENDAL
Jean Le Long
1685
BOSCHENDAL
CHARDONNAY
Appellation Series
Elgin Sauvignon Blanc
BOSCHENDAL
BOSCHENDAL
Chardonnay Pinot Noir
BRUT
BOSCHENDAL
CHARDONNAY
PINOT NOIR
2016
1685
BOSCHENDAL
MERLOT
Blanc de Noir
2016
BOSCHENDAL