CS 4375
ASSIGNMENT _____1_____

Names of students in your group:
Anthony Tran

Number of free late days used: _____0_____

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.
Gradient Descent.ipynb by Professor Anurag
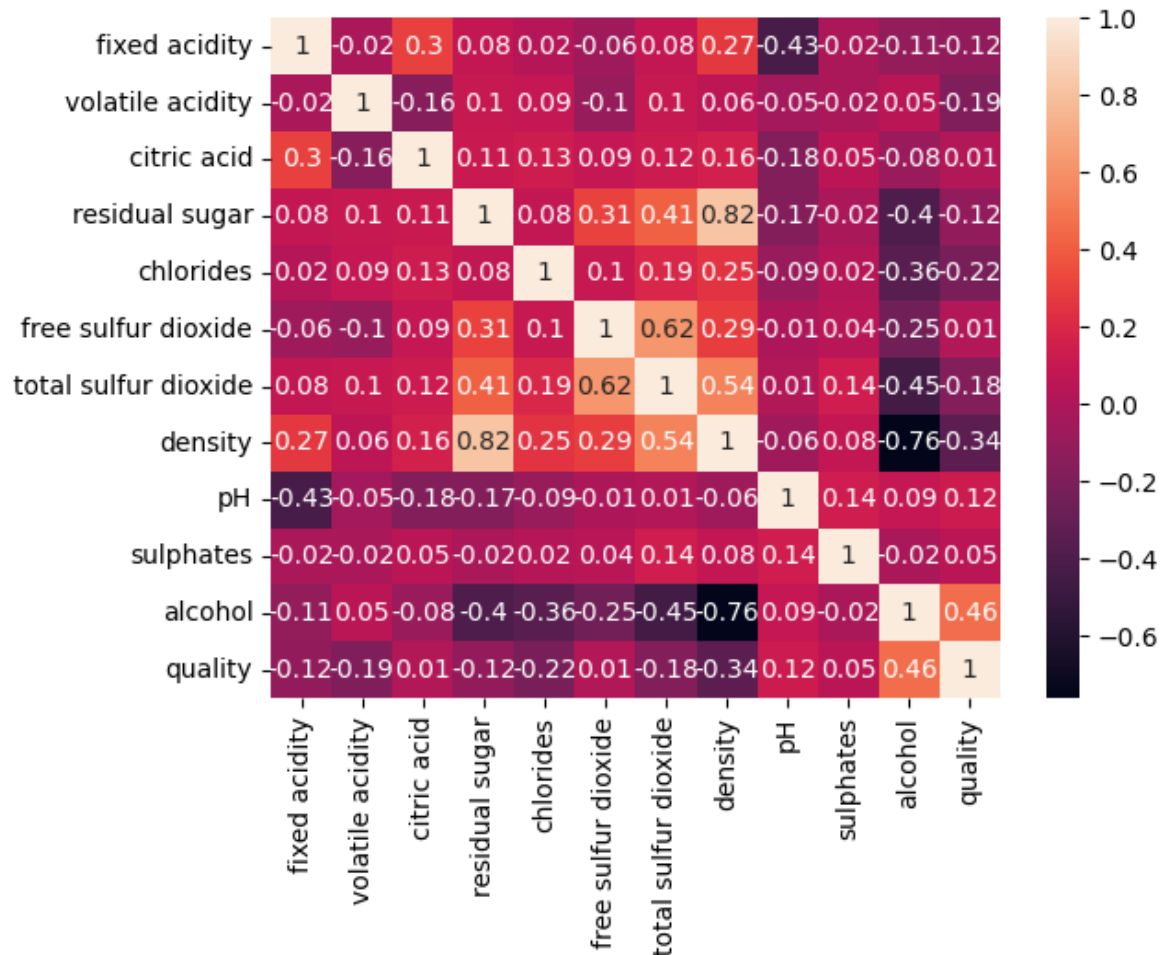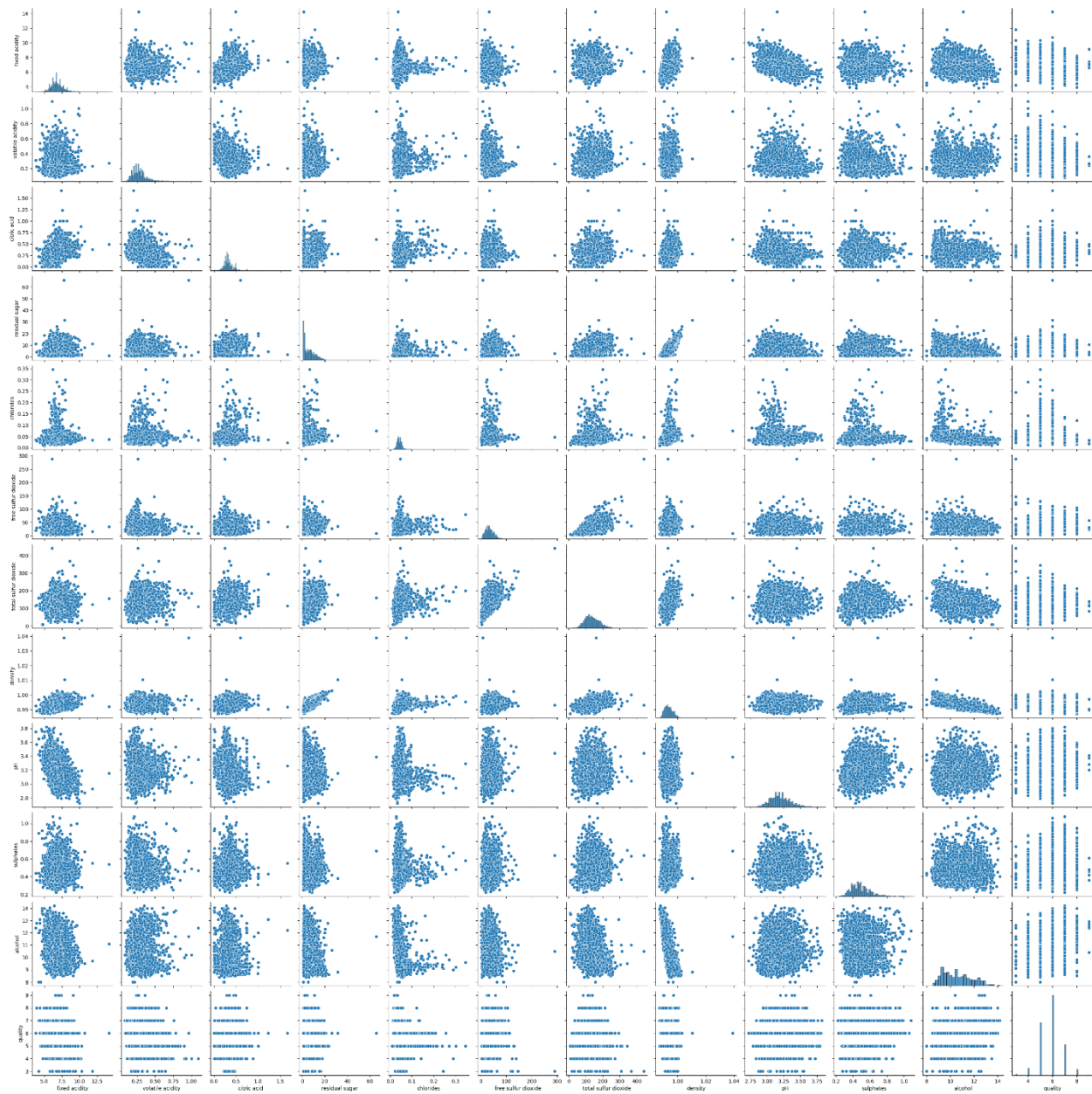
Anthony Tran
CS 4375.001

Libraries used:
-Pandas, seaborn, sklearn

Steps to run:
1. open code in google colab
2. Install all necessary libraries (should be already installed in google collab)
3. run cells in sequential order

Plots:

Question: Are you satisfied that you have found the best solution:

No, I am not satisfied that I found the best solution because the R^2 statistic is averaging 0.2 which is closer to 0 so it indicates my model isn't good at predicting the output. I've tried changing the parameters such as learning rate, max iterations, and alpha, but changing the parameters only negligibly changes the statistics so this could be because the parameters weren't changed properly, or a linear model isn't the best for training this dataset.

Trial Logs:

Model performance for: max_iter=1000, eta0=0.1, alpha=0.001, early_stopping=False

----

Training set:

MSE: 0.6131300360001383

MAE: 0.6120554655951832

EV: 0.2229923049479543

R2: 0.2229347719117627

Testing set:

MSE: 0.6486373723381651

MAE: 0.6257530174697788

EV: 0.1970801743814634

R2: 0.196209065160321

Model performance for: max_iter=1000, eta0=0.01, alpha=0.001, early_stopping=False

----

Training set:

MSE: 0.6085227605703343

MAE: 0.6087376895549884

EV: 0.22891514791373546

R2: 0.22877391421847748

Testing set:

MSE: 0.6469574267708788

MAE: 0.6242770968811797

EV: 0.20014806599792445

R2: 0.19829085242019062

Model performance for: max_iter=5000, eta0=0.01, alpha=0.01, early_stopping=False

----

Training set:

MSE: 0.6085227605703343

MAE: 0.6087376895549884

EV: 0.22891514791373546

R2: 0.22877391421847748

Testing set:

MSE: 0.6469574267708788

MAE: 0.6242770968811797

EV: 0.20014806599792445

R2: 0.19829085242019062

Model performance for: max_iter=10000, eta0=0.001, alpha=0.01, early_stopping=False

----

Training set:

MSE: 0.6089605371822776

MAE: 0.608829224184076

EV: 0.2289088094028252

R2: 0.22821908740713626

Testing set:

MSE: 0.6477073840814007
MAE: 0.6244088469605954
EV: 0.20066483389253043
R2: 0.1973615058955811

Model performance for: max_iter=10000, eta0=0.001, alpha=0.001, early_stopping=True
----
Training set:
MSE: 0.6087641482915878
MAE: 0.6089215371292854
EV: 0.22878681762994013
R2: 0.2284679856329246
Testing set:
MSE: 0.6477010545754824
MAE: 0.6245795725211467
EV: 0.1998310509739799
R2: 0.1973693494144657

Model performance for: max_iter=10000, eta0=0.001, alpha=0.1, early_stopping=True
----
Training set:
MSE: 0.6087641482915878
MAE: 0.6089215371292854
EV: 0.22878681762994013
R2: 0.2284679856329246
Testing set:
MSE: 0.6477010545754824
MAE: 0.6245795725211467
EV: 0.1998310509739799
R2: 0.1973693494144657

Model performance for: max_iter=10000, eta0=0.1, alpha=0.1, early_stopping=True
----
Training set:
MSE: 0.6250364239818462
MAE: 0.6207080261993804
EV: 0.21657865650149744
R2: 0.2078449222070079
Testing set:
MSE: 0.6685367615236025
MAE: 0.6390906700777297
EV: 0.18888582524800834
R2: 0.17154975732789723