

PS1

CS534 - Machine Learning

(25 points)

Notes

Put your homework in google drive under your name. Please mention in the beginning of your file the names of any students with whom you collaborated. If you didn't collaborate with anyone, mark your collaborators as "None." Remember, your goal is to communicate. Full credit will be given only to correct solutions which are described clearly. Convolved and obtuse descriptions will receive low marks.

To complete your homework, you may ONLY consult the following material:

1. Lecture slides
2. Course notes you or others took during lecture.
3. Websites that may clarify the concepts covered in the material but do not in any way provide complete solutions to the problems.

Deadline: 2/11/2021 11:59

Please provide an answer to the following questions and submit your solution in a colab notebook:

Question 1 (10 pts) (Scikit Allowed)

Create a script that:

1. Load the iris data.
2. Scale the independent features.
3. Compute the correlation coefficients and covariance among all the pair of independent features.
4. Compute the PCA and show the principal components (their coefficients).
5. Compute the correlation coefficient between each original feature and the new features generated by the PCA.

Question 2 (5 pts + (5 pts Extra))

1. What you can observe by comparing the results in subquestions 4 and 5 of Question1?
2. Can you define any property?
3. If there is any property are you able to formally prove it? (5 extra pts)

Question 3 (10 pts) (Scikit NOT Allowed)

Write a code in pure python, to create a decision tree (please show the results).

1. Create a toy example to run and test your code.
2. Provide simple explanation for your algorithm.

3. In input, you should provide the independent features and the dependent feature (the class/label) separately.
4. Also implement the gini index and the entropy gain ratio in your toy example.

Notes: You can assume that the dependent and independent features are binary and may use numpy and scipy. You can read the code available online, but you cannot plagiarize it.