Homework 4

DSO 530: Applied Modern Statistical Learning Methods

Spring 2022

Deadline. Thursday, April 21th, 5 pm, Los Angeles time.

Submission instruction. To submit your homework, please go to the **assessments** folder on Blackboard and find HW4. The submitted document should be in **pdf** format. Submission of other formats will incur a 20% penalty. Before you submit, please double check if your document is readable. You can submit your homework multiple times before the deadline, and the TA will grade the last submission.

Late submission policy. All late submission should be sent to xint@marshall.usc.edu. Late submission will incur 20% penalty within 24 hours after the deadline, and 40% penalty between 24 and 48 hours. **No submission after 48 hours of the deadline will be accepted**. As we have a large class this year, I wish that everybody respects this policy.

1. Did you review up to svm? If you haven't, please do so first.

2. Is the following statement correct? "K-means clustering and K-nearest neighborhood classifier are the same thing." (A simple Yes or No answer suffices)

3. Is the following statement correct? "With some choice of kernels, svm can create non-linear classifiers." (A simple Yes or No answer suffices)

4. There are different commonly ussed kernel functions, such ass linear, polynomial, radial basis function (rbf), and sigmoid. The kernel function $\exp(-\gamma\|x - x'\|^2)$ is

- A) linear
- B polynomial
- C) rbf
- D) sigmoid

5. To implement SVM for classification, we should call

- A) svm.SVC()
- B) svm.SVR()

6. (optional, not for grading) Recall the `height` data we used when we introduced simple linear regression. Now, we have learned quite some regression methods.Any of these more modern methods significantly outperform simple linear regression on this dataset in terms of (out-of-sample) R^2? (Note you do have much freedom in addressing this question.)