

Assignment 5

CS780/880: Introduction to Machine Learning

Due: By 12:40PM Thu Apr 20th, 2017

Submission: Turn in as a PDF on myCourses, or printed and turned in at class

Discussion forum: <https://piazza.com/unh/spring2017/cs780cs880>

Problem 1 [30%] Suppose we fit a curve with basis functions $b_1(X) = X$, $b_2(X) = (X - 1)^2 I(X \geq 1)$. (Note that $I(X \geq 1)$ equals 1 for $X \geq 1$ and 0 otherwise.) We fit the linear regression model

$$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \epsilon,$$

and obtain coefficient estimates $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = 2$. Sketch the estimated curve between $X = -2$ and $X = 2$. Note the intercepts, slopes, and other relevant information.

Problem 2 [30%] Fit some of the non-linear models investigated in Chapter 7 of ISL to the **Auto** data set. Is there evidence for non-linear relationships in this data set? Create some informative plots to justify your answer.

Problem 3 [40%] Apply *boosted trees*, *bagged trees*, *random forests*, and *SVMs* to a decision data set of your choice. Be sure to fit the models on a training set and to evaluate their performance on a test set. How accurate are the results compared to simple methods like linear or logistic regression? Which of these approaches yields the best performance?