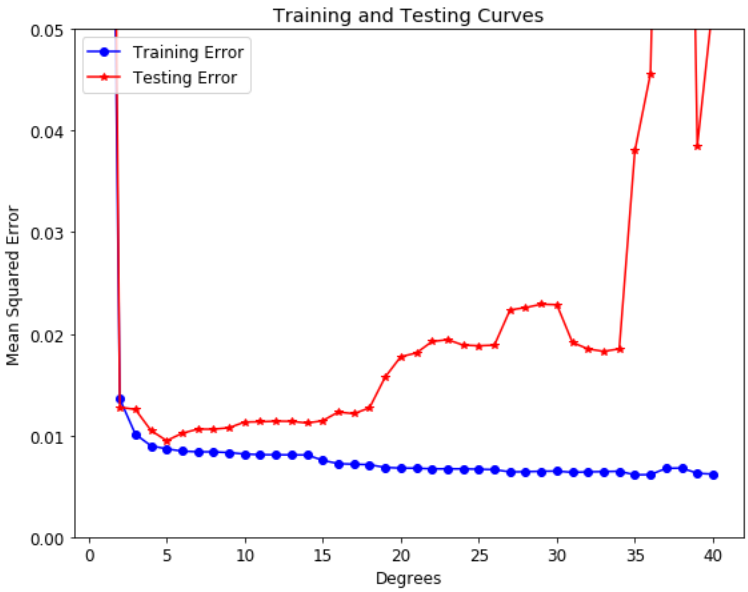
CSYE 7245–
Big-Data Systems and Intelligence Analytics
Sample Quiz One

This should all be a review from an Intro to Machine Learning Course

Q1 (15 Points) Each of n students gives their phone the Professor before a test. The Professor (being lazy) gives the phones back to the students in a random order. What is the expected number of students who get back their own phone?

Q2 (15 Points) Polynomial models of various order (degree) were fit to the same data. Are any of the models below underfitting or overfitting? Name which models are under, overfit or correctly fit.



Q3 (15 Points) Assume regression is being used to predict whether a student will graduate with honors or not. The dependent variable is **hon** which is yes as indicated by a 1 a no indicated by a 0. Assume the only independent variable called female and encoded with female as 1 and male as 0. The stats for the fit are shown in the table below.

hon	Coef.	Std. Err.	z
female	.5927822	.3414294	1.74
intercept	-1.470852	.2689555	-5.47

A. Write an equation that describes the model.
B. Is the coefficient female significant? How does one interpret the meaning of its value?
C. Is the coefficient intercept significant? How does one interpret the meaning of its value?

Q5 (15 Points) Calculate the increase in the odds of receiving honors if one is female given the table in Q4.

Q6 (25 Points) Assume you have 3 binary classifiers (A,B,C) each with a 70% accuracy. You can view these classifiers for now as pseudo-random number generators which output a "1" 70% of the time and a "0" 30% of the time. The output of these classifiers is independent.

Create a machine learning ensemble algorithm that combines these 3 binary classifiers in to one prediction.

Describe your algorithm in detail.

What is accuracy of your algorithm?