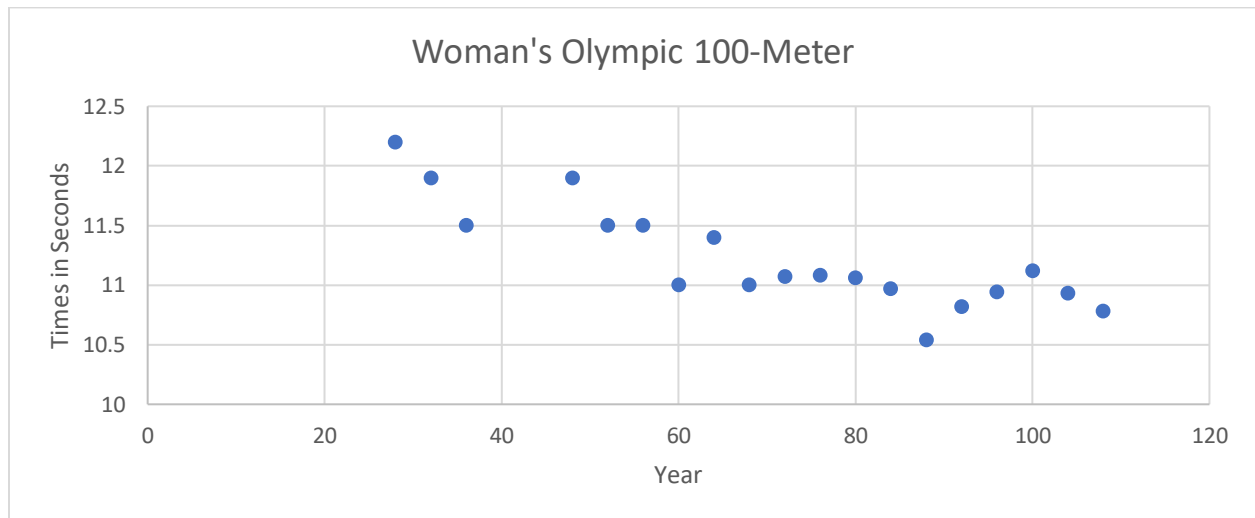


Project 2: Choosing a Model for Predicting on Unseen Data

Due: Before Midnight on February 8, 2020

For Project 2 you will use the regression program that you created in Project 1 to choose a model to predict future times for the women's Olympic 100-meter race. We will code the year of each race as we did in video lecture 2.6. A text file with the data is available on Canvas for the years 1928 through 2008 when the Olympics were actually held.



Your project assignment is to compare three different models, linear, quadratic, and cubic.

$$h_w(x) = w_0 + w_1x$$

$$h_w(x) = w_0 + w_1x + w_2x^2$$

$$h_w(x) = w_0 + w_1x + w_2x^2 + w_3x^3$$

using 5-fold cross validation.

You should turn in a pdf that contains a title for your project and your name. Then you should present a **chart**, similar to the one in the lecture (see below), of all your test results and a **plot** of your training and test J's with respect to the polynomial degree.

	Linear	Quadratic	Cubic
1234	J	J	J
5			
1235			
4			
1245			
3			
1345			
2			
2345			
1			
Mean for Training			
Mean for Testing			

Based on your data and plot, you should then argue which model (linear, quadratic, or cubic) you expect will best predict the times for the women's Olympic 100-meter race in the future.

Your pdf document should be named yourlastname_yourfirstname_P2.pdf.

Then, compute weights using **the complete data set** for the training set with your best model. Using those weights, write a Python program that takes a year as input, then predicts the winning women's Olympic 100-meter race time for that year.

Your Python program should be named yourlastname_yourfirstname_P2.py.

Zip the two files together and upload them. Make sure that the zip file title also contains your name.