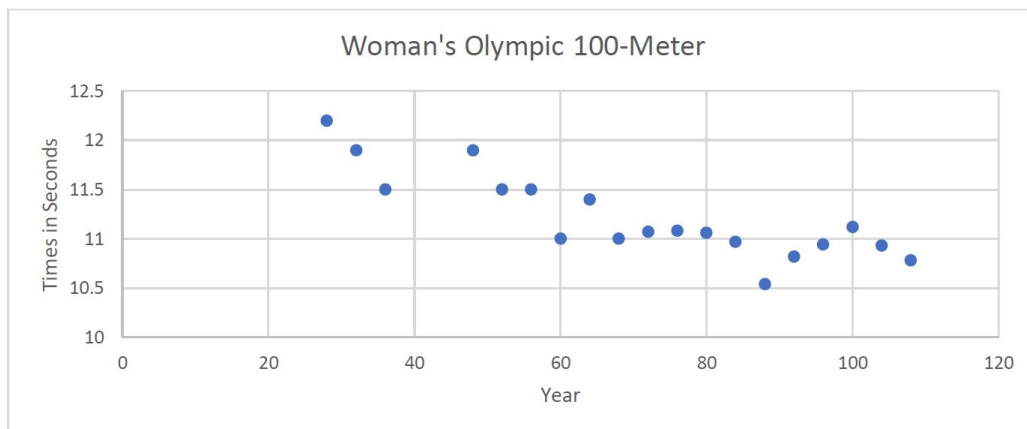


Project 2: Choosing a Model for Predicting on Unseen Data

Due: 11:59 pm on Feb. 9th, 2022

For Project 2 you will create a regression program and choose a model to predict the women's Olympic 100-meter race record time for year 2022. We will code the year of each race as we did in lecture 2.3. A text file with the data is available on Canvas for the years 1928 through 2008 when the Olympics were held. The first line of the text file indicating there're **m** lines of data and a **n** number of features (in this case, one).



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.

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

Important Note:

You **cannot** use python machine learning package that can have the k-fold validation algorithm as embedded function, for instance, sklearn package.

You are required to submit a project report, including:

- The J value **chart** as shown in the table above.
- A **plot** of your training and test J's with respect to the polynomial degree
- Argue which model (linear, quadratic, or cubic) you will choose
- The final hypothesis function $h_w(\mathbf{x})$
- Predict the women's Olympic 100-meter race record time for this winter Olympic (2022)
- Full screenshot of your python console.
- A copy of your code

Your report should be named yourlastname_yourfirstname_P2.docx or .doc or .pdf. Your Python program should be named yourlastname_yourfirstname_P2.py, then zipped together with your project report and uploaded to Canvas