Polynomial Regression

Name Surname: Beria Ayşenur Can
Student Id: 090200705
CRN: 11512
Instructor: Emre Onur Kahya

1. Homework :

1) Plot $J_0(x)$, Bessel function for $n = 0$, for a small interval (from 0 to 8)
2) Your target will be a bessel plus s small noise $J_0(x) + 1/10.N(0, 1)$.Namely one tenth of a small Gaussian noise (0 mean and unit variance).
3) You are asked to train an 8th order polynomial regression model with $a_0 + a_1x + ...a_8x^8$ to fit the data that you syntectically produce.
4) First collect 10 syntectic random data points from $J_0(x) + 1/10.N(0, 1)$. Take 8 of them as training set and take 2 of them for validation set and show that training error is zero but validation error is high.
5) Repeat the same thing for $20, 100, 1000, 10^4$ points and show that the normalized L_2 loss gets smaller for validation error as you increase the data points.
6) Repeat the same exercise with a Ridge Regularization/Regression term and show that overfitting is avoided.

2. Aim

The purpose of this assignment is to find the best coefficient of polynomial fit to avoid overfit for randomly selected data from the noisy Bessel Function that is sufficiently similar to the real Bessel Function.

3. Results :

4. Analysis :

For 8 data, an 8th order polynomial fits perfectly.If I have 10 pieces of data, 8 of them made perfect fit and 2 of them make overfit.We calculate the loss by optimizing.Instead of increasing the data, overfit is prevented when regularization term is used.In short, more data reducing overfit

References: