Problem-1

a) In this part, program calculates the direct solution of optimization problem and plots it. It is very basic concepts. I added a bias term to the data when calculating the weights.

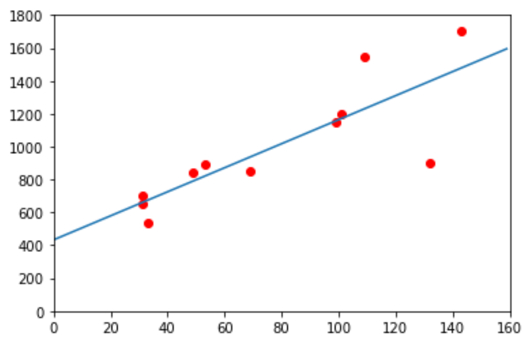

Figure 1: Direct_Linear_Reg

b) Second part of the first questions is not different than the first part. Only difference between them is the method. In this part, program uses gradient descent method to solve the optimization problem. Program prints empirical error and plots solution line for each iteration. When you run the code, you will see plots and empirical errors. I tried two different step size (learning rate or alpha). Bigger alpha converges faster because it takes bigger step size. But if we use bigger alpha this can be a cause of an oscillation problem. In that case, it will never converge to the local optima.
You can see the differences between empirical risk maps. High alpha congers faster.

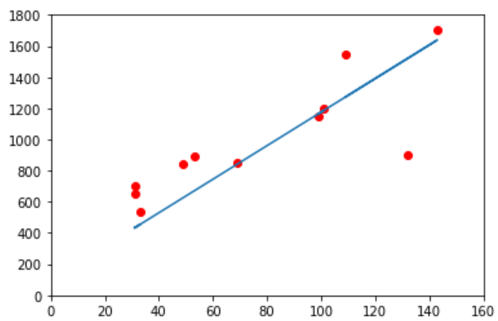

Figure 2: Last Plot of The Program

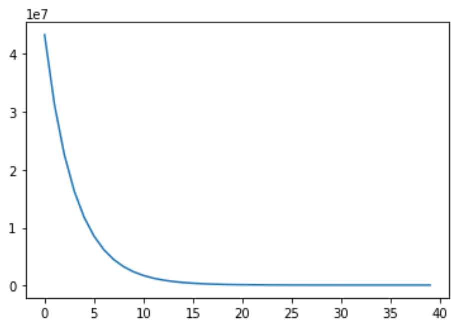

Figure 4: alpha = 10^(-5)
Empirical Risk Map

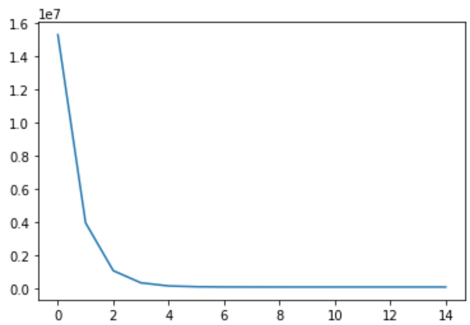

Figure 3: alpha=10^(-4)
Empirical Risk Map

Problem-2

First cell of the code uses direct method to find weight matrixes. I used four different degree for calculations. Best solution for this problem is fourth degree polynomial linear regression. The fourth-degree model has lowest empirical risk. Direct method gave pretty good results.

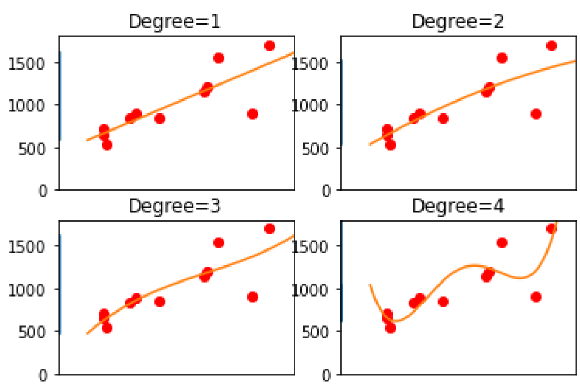

Figure 5: Different Models of Polynomial Regression

The second part of the code uses the gradient method with different degrees to find weight matrixes. All the models have different alpha values and different starting weights. My models did not give me the direct method's results. I think the most important reason for that is converging local optimums. The gradient is stuck in a local optimum and cannot escape there. Because of this problem, I initialize weights with different points. Thanks to these points, it converges better and faster. Although this problem, it gives good results. Maybe using normalization, I may get better results.

Model-1:

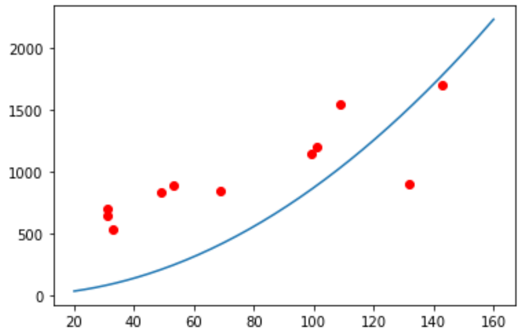

Figure 7: Model-1 Curve

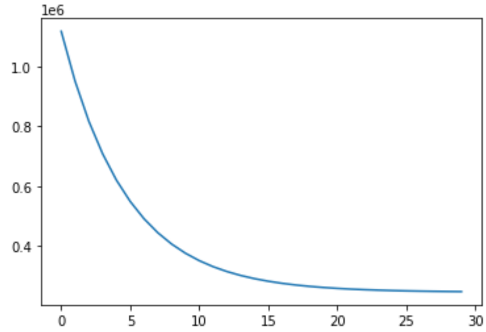

Figure 6: Model-1 Empirical Risk

Model-2

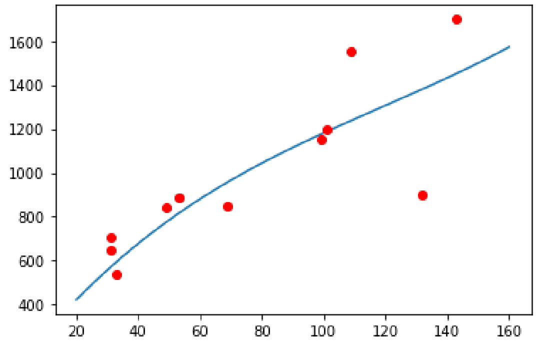

Figure 9: Model-2 Curve

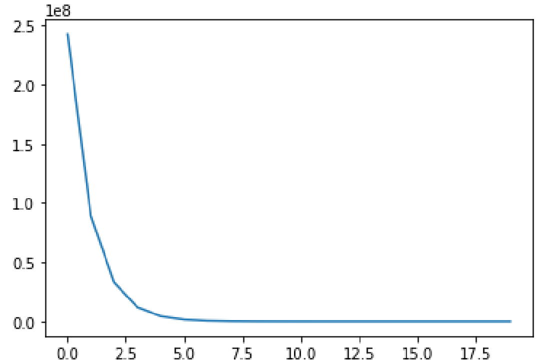

Figure 8: Model-2 Empirical Risk

Model-3

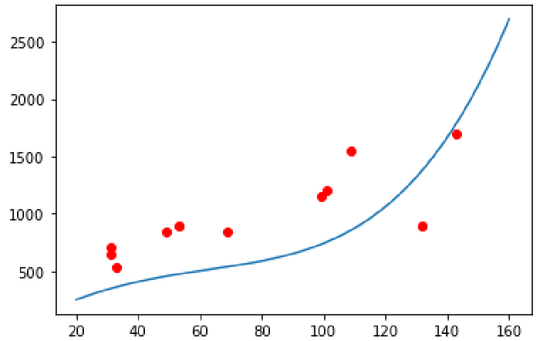

Figure 10: Model-3 Curve

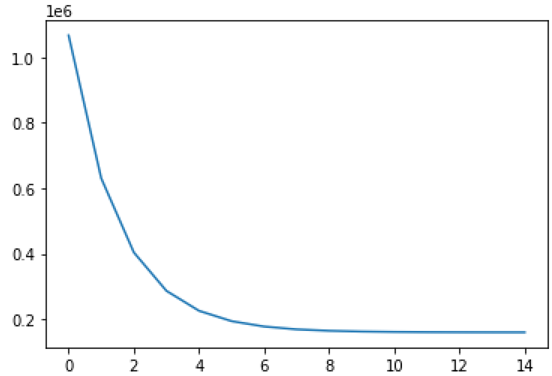

Figure 11: Model-3 Empirical Risk