Q1 :

Q2 :

When is redefined as [w;b] and x=[x,1], where this 1 represents a row of ones as large as x, then the formula becomes and since matrix multiplications work by multplieng the elements of the first matix’ row whit the seconds column it can be said that for each element in the new matrix the formula will be: where n and m are the size of . Doing this for each element corresponds to .

This means that now again the formula needs to be minimalized in order to get the lowest possible cost. Just like in the previous exercises and lecture 1 this corresponds to the Wiener filter:

Q3: implementing the previous answer on this data yields

Q4:In order to get a model that is more robust to noise regularization can be applied. L^2 smoothes out the weights making it not as vulnerable to noise. The learning algorithm perceives higher variance of X , causing it to shrink weights of features that have low covariance compared to the added variance

Q5: