Report - Programming Assignment 2 (CSE 574)
Group - 58

Team:
Roopali Vij (roopaliv)
Shreeju Jayesh Tanna (shreejuj)
Ranganatha Poola Narayana Swamy (rpoolana)

Report 1

Below are the accuracies obtained for Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) on the provided test data set : sample_test.pickle

- Linear Discriminant Analysis : 97 %
- Quadratic Discriminant Analysis : 96%

Reason for the difference between boundaries of LDA and QDA: LDA learns linear boundaries between different classes, but QDA learns quadratic ones. QDA has higher variance when the data size is relatively small due to the quadratic model. However, QDA is expected to perform better when the data size is larger or the boundaries are not necessarily linear.

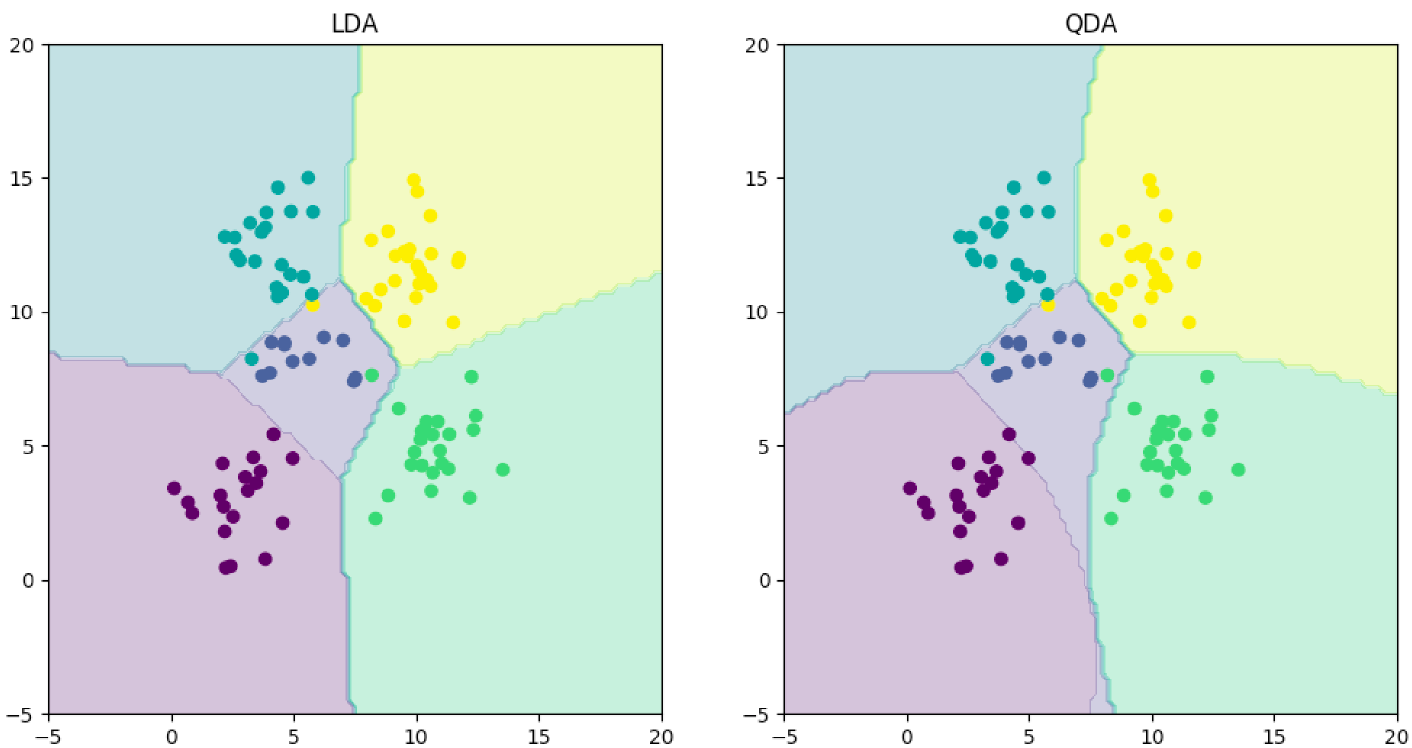


Figure: Discriminating boundaries for linear and quadratic discriminators

Report 2

MSE calculation with and without intercepts.

for training and test data for two cases: first, without using an intercept (or bias) term, and second with using an intercept. Which one is better?

MSE without intercept : 106775.36

MSE with intercept : 3707.84

Performance is better with intercept as the addition of bias term reduces MSE

Report 3

Calculate and report the MSE for training and test data using ridge regression parameters using the the testOLERegression function that you implemented in Problem 2. Use data with intercept. Plot the errors on train and test data for different values of λ. Vary λ from 0 (no regularization) to 1 in steps of 0.01. Compare the relative magnitudes of weights learnt using OLE (Problem 2) and weights learnt using ridge regression.

Weights learnt using OLE:

[[-4.12173303e+02], [-3.45940349e+02], [5.78814085e+02], [5.89243806e+01], [-1.35891612e+06]
[1.19462263e+06], [5.07036457e+05], [-1.34586972e+03], [4.47713283e+05], [4.77903799e+02]
[-1.40658376e+02], [-9.19340330e+02], [-3.95968904e+02], [-7.25692641e+04], [-8.95093666e+04]
[-3.23782610e+03], [1.40730022e+03], [3.91795187e+04], [2.65084262e+02], [5.12843509e+02]
[2.01158086e+02], [6.99140521e+01], [-4.24307012e+03], [3.44644877e+03], [2.22399630e+03]
[-1.77734526e+02], [1.58033897e+03], [1.08315565e+02], [2.47108215e+02], [-2.38458154e+01]
[7.93875956e+02], [2.96369205e+02], [-6.06952736e+02], [-5.63192652e+02], [-5.48068266e+02]
[9.69932252e+01], [5.90070172e+02], [-1.34356078e+03], [2.32884509e+03], [-1.59023051e+02]
[-7.56384887e+02], [4.36008445e+02], [-2.45555655e+02], [-8.74482672e+03], [7.02634079e+03]
[3.62258437e+03], [5.78813898e+02], [3.31947790e+03], [-2.39954611e+02], [1.58217349e+05]
[2.03869016e+04], [-3.86106074e+04], [2.70468290e+05], [5.91068449e+03], [-2.45407147e+04]
[3.71944333e+04], [-2.33074901e+05], [-5.09687335e+03], [1.29940874e+04], [-8.95897226e+04]
[-2.49967576e+03], [1.08307098e+04], [-5.54804850e+02], [-1.39904552e+03]]

Weights learnt using ridge regression:

[[1.48154876e+02], [1.27485219e+00], [-2.93383522e+02], [4.14725449e+02], [2.72089134e+02]
[-8.66394570e+04], [7.59144679e+04], [3.23416228e+04], [2.21101216e+02], [2.92995512e+04]
[1.25230360e+02], [9.44110833e+01], [-9.38628632e+01], [-3.37282800e+01], [3.35319772e+03]
[-6.21096300e+02], [7.91736534e+02], [1.76776039e+03], [4.19167405e+03], [1.19438121e+02]
[7.66103400e+01], [-1.52001293e+01], [8.22424594e+01], [-1.45666208e+03], [8.27386703e+02]
[8.69290952e+02], [5.86234495e+02], [4.27026727e+02], [9.02467690e+01], [-1.78876224e+01]
[1.41696774e+02], [5.82819384e+02], [-2.34037511e+02], [-2.56071452e+02], [-3.85177401e+02]
[-3.34176738e+01], [-1.07350066e+01], [2.57107189e+02], [5.99554592e+01], [3.83728042e+02]
[-4.04158390e+02], [-5.14286435e+02], [3.83636642e+01], [-4.46102889e+01], [-7.29643531e+02]
[3.77408337e+02], [4.39794290e+02], [3.08514373e+02], [1.89859679e+02], [-1.09773797e+02]
[-1.91965699e+03], [-1.92463378e+03], [-3.48979528e+03], [1.17969687e+04], [5.30674415e+02]
[5.43305906e+02], [1.82107518e+03], [-1.04639807e+04], [-5.16627611e+02], [2.06435917e+03]
[-4.19941335e+03], [-1.40495705e+02], [3.74157090e+02], [5.14757492e+01], [-4.64492730e+01]]

We observe that the weights learnt from Ridge regression are generally smaller in magnitude than the ones learnt from OLE regression

Error comparison between OLE regression and Ridge regression

OLE Regression:

MSE for training data without intercept: 19099.44

MSE for training data with intercept: 2187.16

MSE for test data without intercept: 106775.36

MSE for test data with intercept: 3707.84

Ridge Regression:

lamba = 0.0 | MSE_train = [2187.16029493] | MSE_test = [3707.84018167] | w_l = [[1.48154876e+02]

lamba = 0.01 | MSE_train = [2306.83221793] | MSE_test = [2982.44611971] | w_l = [[1.49604513e+02]

lamba = 0.02 | MSE_train = [2354.07134393] | MSE_test = [2900.97358708] | w_l = [[1.49932244e+02]

lamba = 0.03 | MSE_train = [2386.7801631] | MSE_test = [2870.94158888] | w_l = [[150.13407098]

lamba = 0.04 | MSE_train = [2412.119043] | MSE_test = [2858.00040957] | w_l = [[150.27450793]

lamba = 0.05 | MSE_train = [2433.1744367] | MSE_test = [2852.66573517] | w_l = [[150.3788659]

lamba = 0.06 | MSE_train = [2451.52849064] | MSE_test = [2851.33021344] | w_l = [[150.45959807]

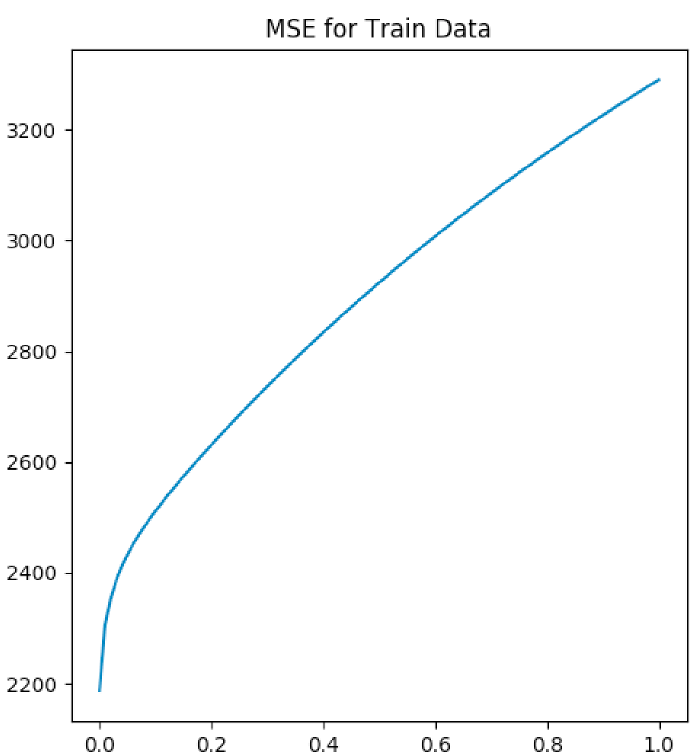
lamba = 0.07 | MSE_train = [2468.07755253] | MSE_test = [2852.34999406] | w_l = [[1.50523737e+02]

lamba = 0.08 | MSE_train = [2483.36564653] | MSE_test = [2854.87973918] | w_l = [[150.57563868]

lamba = 0.09 | MSE_train = [2497.74025857] | MSE_test = [2858.44442115] | w_l = [[150.61818516]

…...

Optimal value for λ = 0.06, because we see that MSE for test data reaches minimum value at lambda = 0.06 and grows again from there on.



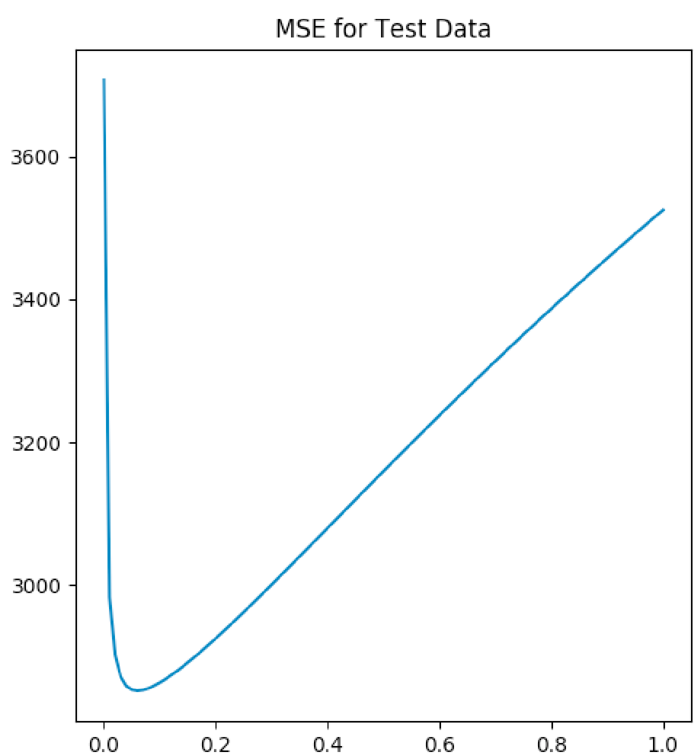


Figure: MSE comparisons for training and test data

Report 4

Observation: Errors on train and test data obtained by using the gradient descent based learning by varying the regularization parameter λ using Scipy’s minimize function closely follows the direct minimization, especially as the number of iterations increase to say 100

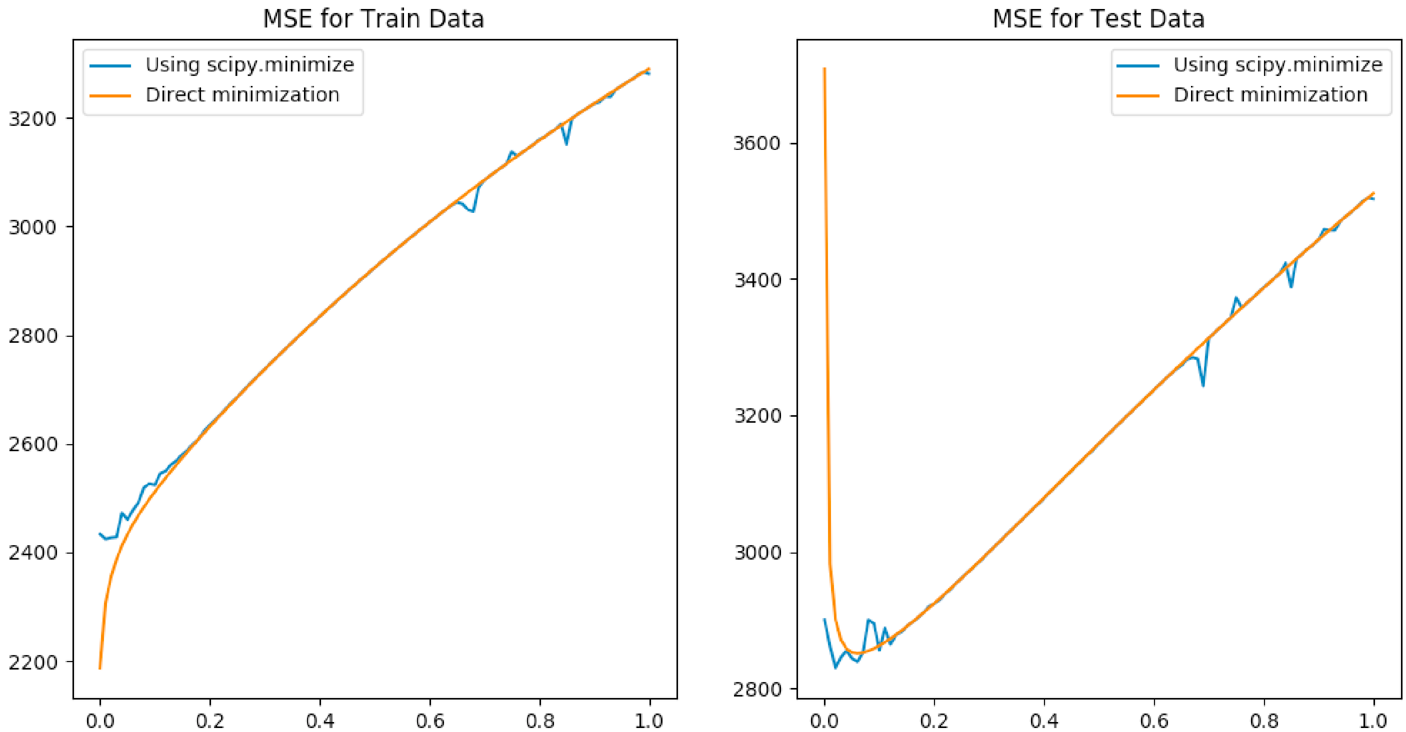


Figure: Errors by using the gradient descent by varying the regularization parameter λ (Scipy iter : 20)

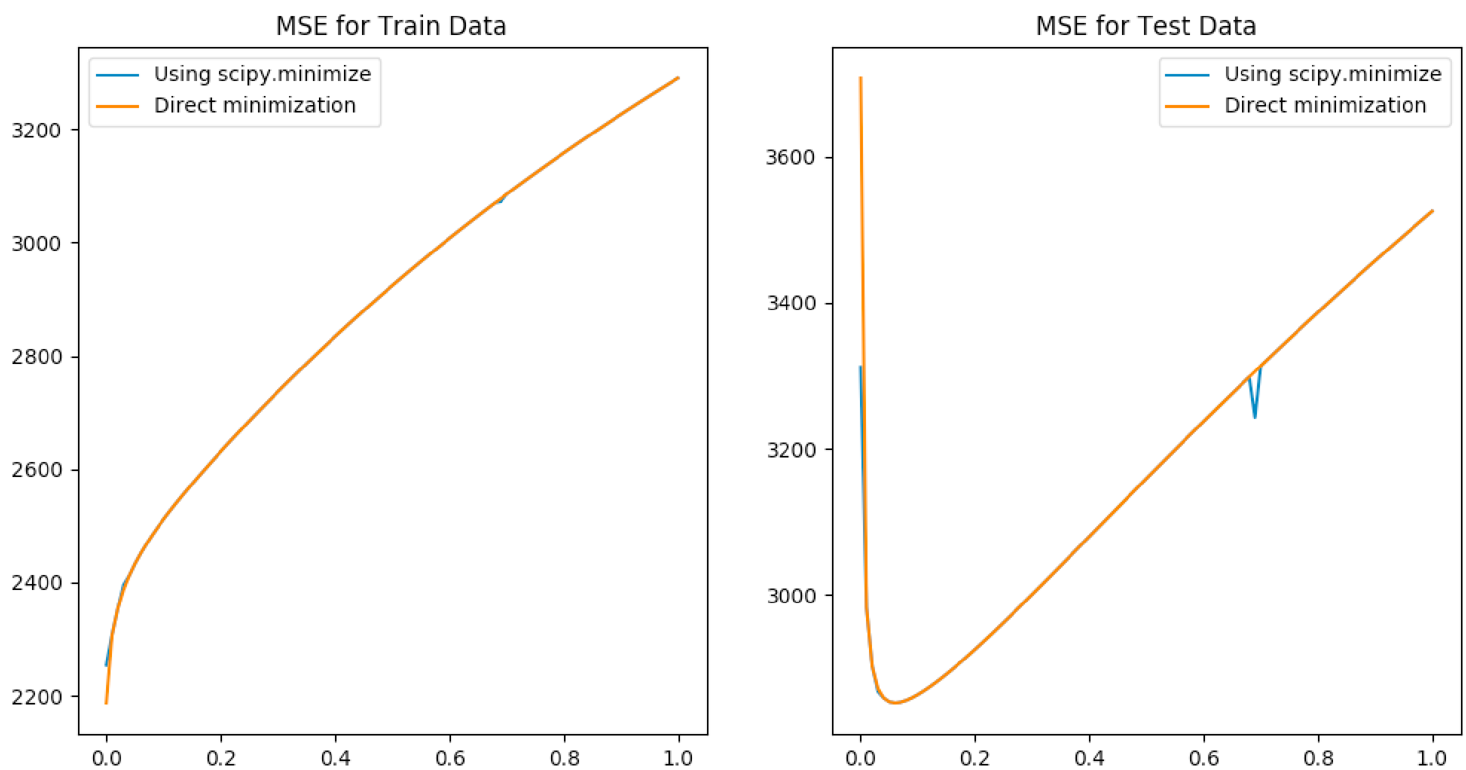


Figure: Errors by using the gradient descent by varying the regularization parameter λ (Scipy iter : 100)

Report 5

From Problem 3, we observed that the error was minimum at $\lambda = 0.06$.
As we continue increasing the value of p from 0 to 6, we see that the errors decrease linearly till p=1, stays minimum for p=2 and p=3. Till here, the regularization parameter does not have significant influence.

However, for p=4, p=5 and p=6, we see that, without regularization parameter λ, the errors start increasing significantly and regularization parameter $\lambda = 0.06$ is required to keep the errors minimal. The same can be observed in the plot below

We choose p = 3 to be optimal

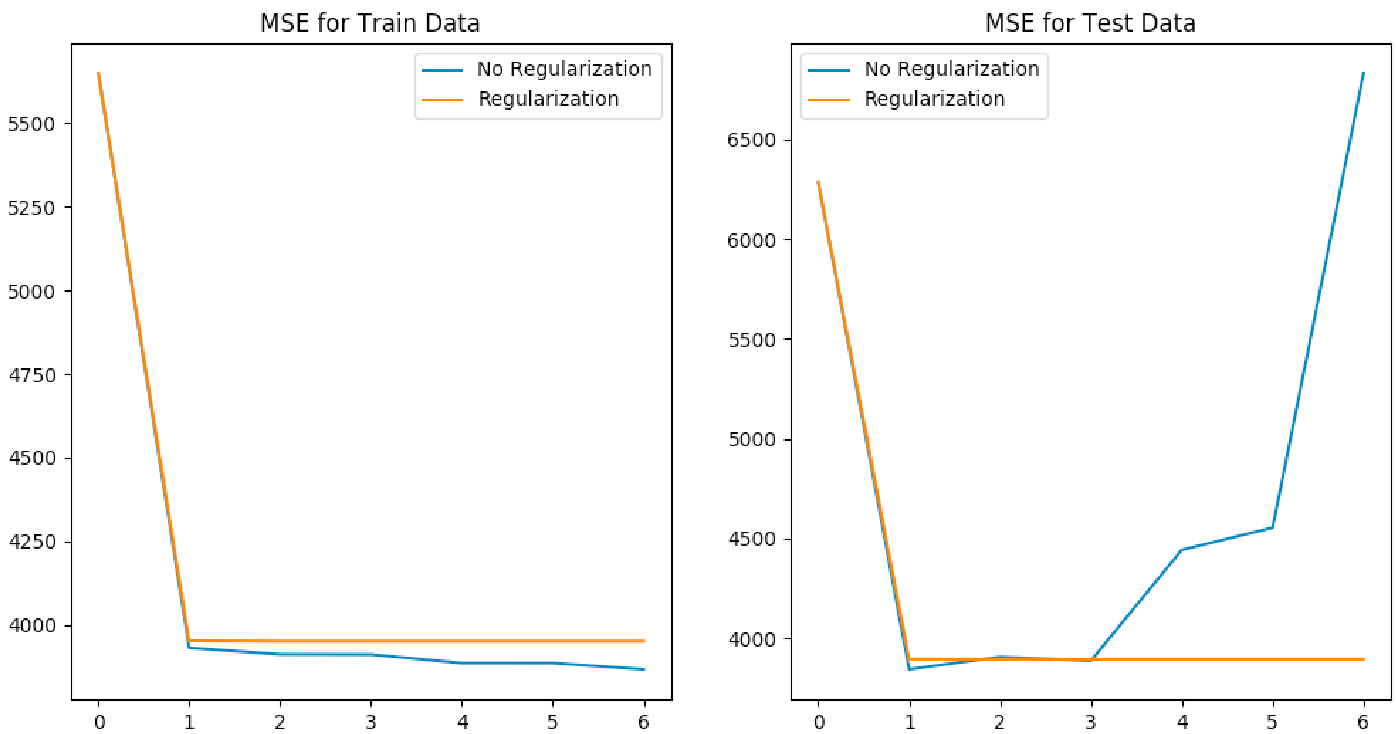


Figure: Errors in Ridge regression with and without regularization (λ=0 and λ=0.06)

Report 6

From the implementation of LDA and QDA we realize that LDA performs marginally better for the given data set. However, QDA is preferable to classify a sufficiently large data set.

With linear regression, we observe that there can be a large error in fitting when there are outliers in the dataset. This is because linear regression tries to minimize the mean square error between the fit line and all of the points, which would mean that the deviation of the outlier from the other points would largely influence the fit.

To minimize the effect of outliers, we adapt few modifications.

Regularization parameter used in Ridge regression reduces the effect of outliers. When some of the weights calculated by linear regression are significantly larger than some of the other weights, the smaller weights will have lesser influence on the fit than the larger ones. This can also be tackled by the regularization parameter.

Another approach is to use higher order parameters of the attributes of the input parameters. This results in a better fit. However, as the order of the parameters go up, it might lead to overfitting of training data and poor fit with the test data. Regularization parameter could help minimize the overfitting issues caused by the higher order parameters.