Q1.

1. APS Failure at Scania Trucks Data Set

Data from https://archive.ics.uci.edu/ml/datasets/APS+Failure+at+Scania+Trucks

I choose to use logistic regression for classification because fitting models that map continuous inputs to different 'classes' and I would do multi-class classification. (a positive class consists of component failures for a specific component of the APS system and a negative class consists of trucks with failures for components not related to the APS.)

There are 171 input variables and 2 outputs. The attribute names of the data have been anonymized for proprietary reasons.

2. Robot Execution Failures Data Set (failures in approach to grasp position)

Data from https://archive.ics.uci.edu/ml/datasets/Robot+Execution+Failures

I choose to use logistic regression for classification because fitting models that map continuous inputs to different 'classes' and I would do multi-class classification.

There are five inputs which are a force in the x-axis, force in the y-axis, force in the z-axis, torque in the y-axis, and torque in the z-axis. Four outputs are normal, collision, fr_collision, and obstruction. The predicted variable is failure type.

3. Accelerometer Data Set

Data from https://archive.ics.uci.edu/ml/datasets/Accelerometer#

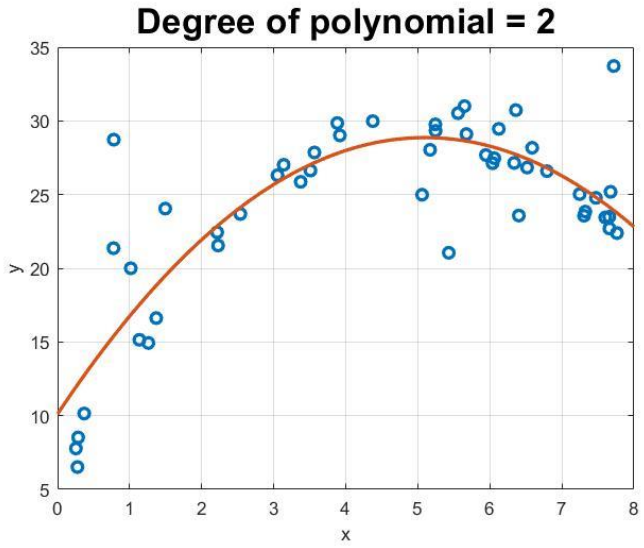
I choose to use logistic regression for classification because fitting models that map continuous inputs to different 'classes' and I would do multi-class classification.

There are four inputs: pctid: Cooler Fan RPM Speed Percentage ID (20 means 20%, and so on), x: Accelerometer x value, y: Accelerometer y value, and z: Accelerometer z value.

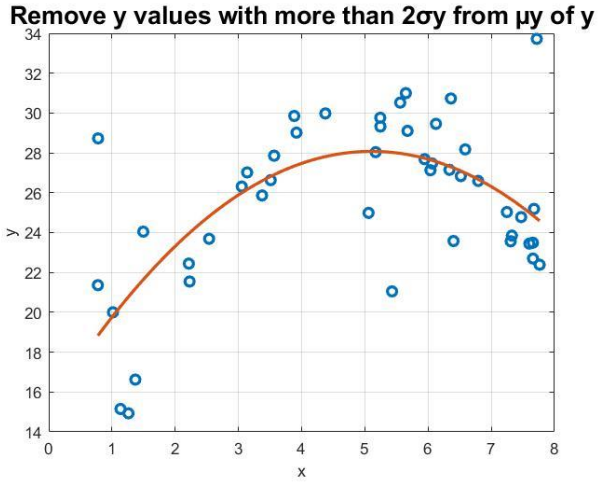
wconfid: Weight Configuration ID (1 - 'red' - normal configuration; 2 - 'blue' - perpendicular configuration; 3 - 'green' - opposite configuration) can be the predicted variable.

Q2first.

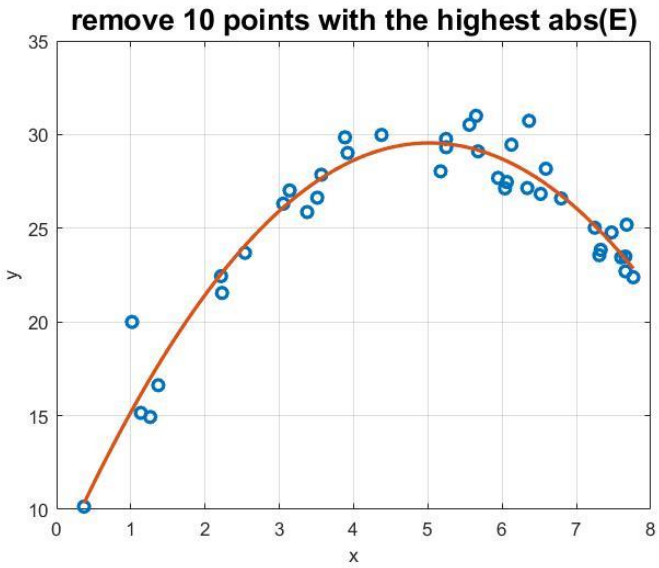
a)


Degree of polynomial = 2
y
x

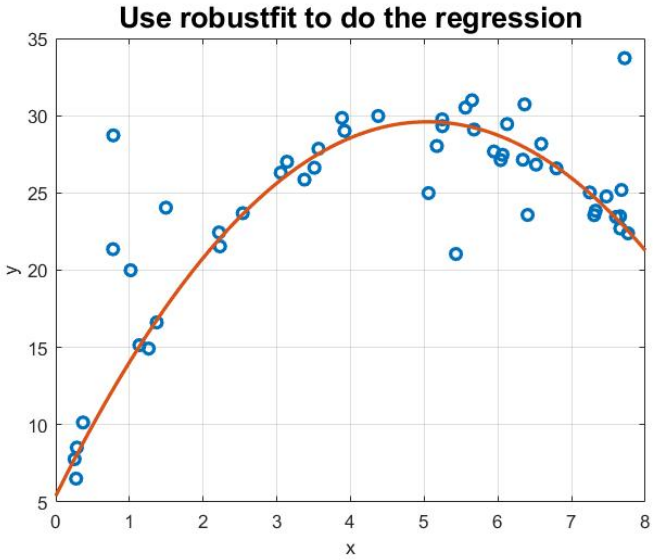
b)


Remove y values with more than 2σy from μy of y
y
x

C)



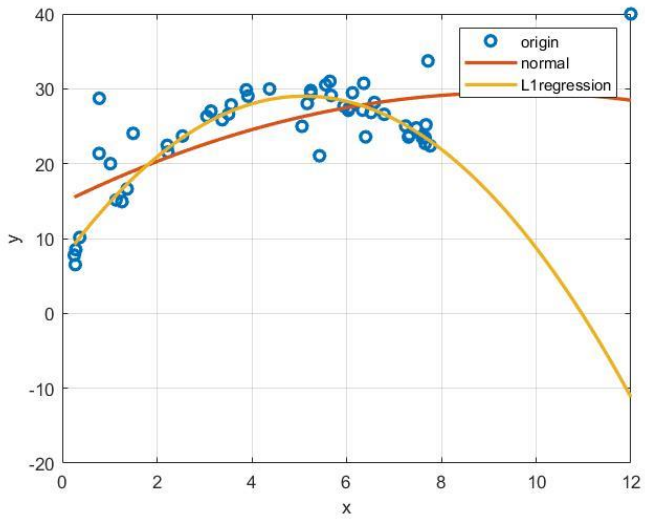
d) Robust regression uses iteratively reweighted least squares to assign a weight to each data point which can automatically and iteratively calculate the weights. This method is less sensitive to large changes in small parts of the data. As a result, robust linear regression is less sensitive to outliers than standard linear regression.



e) L1regression is minimize L1 norm which is find the median rather than mean. This making regression process less robust to outliers.

f)

This graph shows that adding the outlier does not very affect the obtained coefficients if use L1regression.



Q2second.

a)

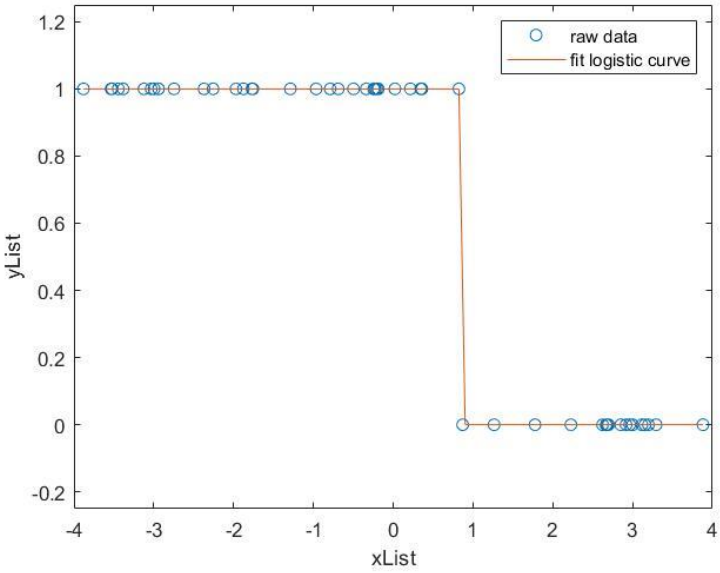
Bresult:

503.683980280280

-599.990008151574

The boundary condition is x = 0.8395

c) The accuracy is 1.



b)

Bresult:

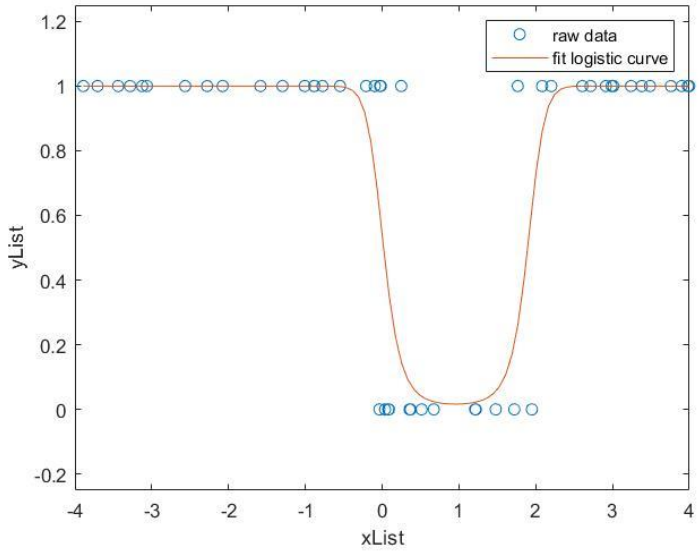
0.165983449855718

-8.89573996358648

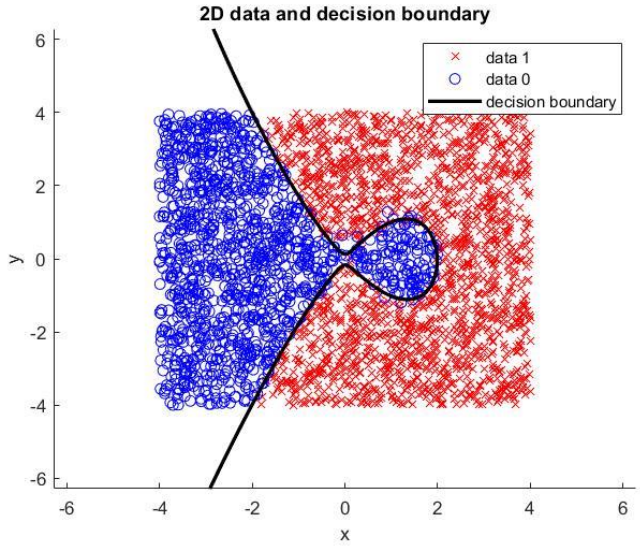
4.64802663343994

The boundary condition is x = 1.462 and x = 0.0145.

c) The accuracy is 0.92


raw data
fit logistic curve
yList
xList
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-4
-3
-2
-1
0
1
2
3
4

Q3.



The accuracy is 0.9897.

Bresult:

-0.202361768576166

0.304288153133607

0.397381401448749

-18.9968499783588

9.33799094298506

-0.390527927071283

9.48832072550174

0.289240485312953

-0.136254306626901

-0.150094660608715

Q5.

a) Determine whether the image contains a number is more than 5 or not.

What is the training set accuracy and the test set accuracy?

Accuracy_test = 0.8618

accuracy_train = 0.8917

errorRate_FalsePositive = 0.1167

errorRate_FalseNegative = 0.1708

b) Determine whether the image contains a number is even or odd.

What is the training set accuracy and the test set accuracy?

accuracy_test = 0.8758

accuracy_train = 0.8843

errorRate_FalsePositive = 0.1198

errorRate_FalseNegative = 0.1287

c) Determine whether the image contains a number is 7 or not.

What is the training set accuracy and the test set accuracy?

Accuracy_test = 0.9805

accuracy_train = 0.9892

errorRate_FalsePositive = 0.0115

errorRate_FalseNegative = 0.0895