ANSWER

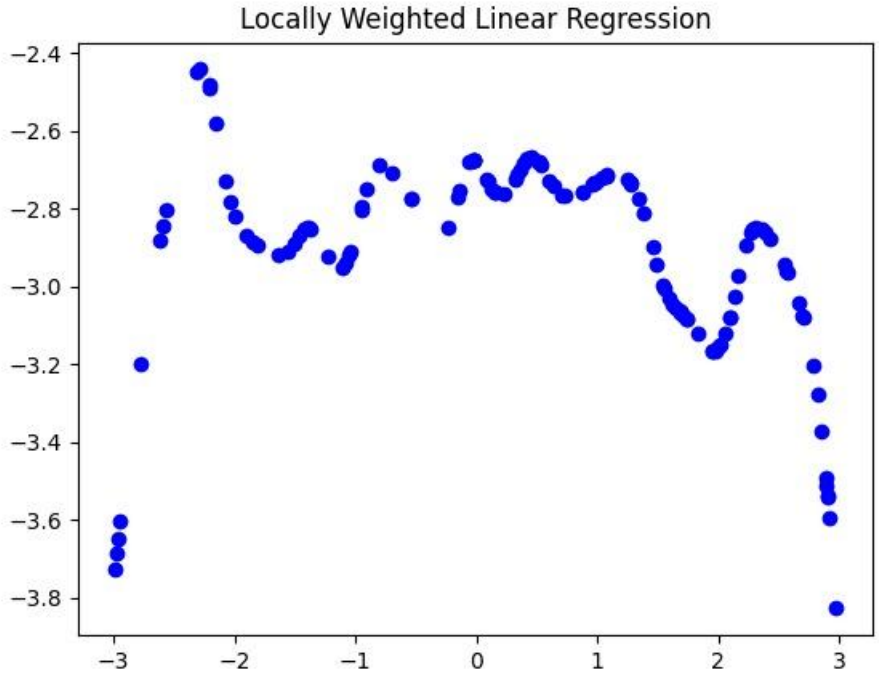
1c)Evaluate your regression functions by computing the error on the test data points that were generated for Question 1c). Compare the error results and try to determine for what "function depths" overfitting might be a problem. Which 'function depth' would you consider the best prediction function and why?

When we see the graphs low MSE can be found at depth 0 and increase MSE with the increase in the depth so I consider depth 0

1d)Repeat the experiment and evaluation of part b) and c) using only the first 20 elements of the training data set part b) and the Test set of part c). What differences do you see and why might they occur?

There is no noticeable difference when evaluating the first 20 elements.

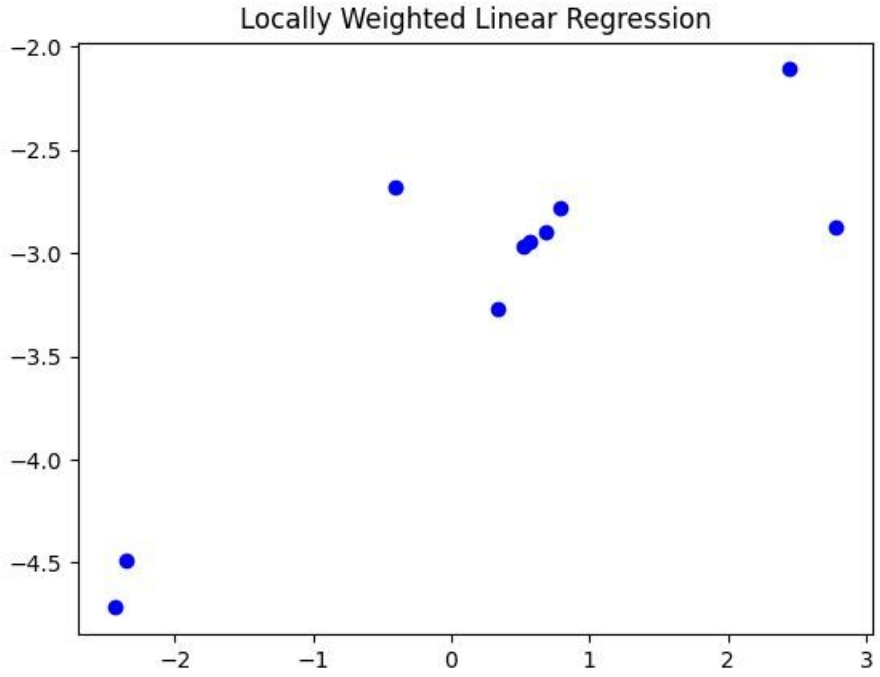
2a,2b)



2c) Evaluate the locally weighted linear regression on the Test data from Question 1 c). How does the performance compare to the one for the results from Question 1 c) ?

The error is 0.8784 when we compare this with results from 1c it performed very well with least error.

2d)Repeat the experiment and evaluation of part b) and c) using only the first 20 elements of the training data set. How does the performance compare to the one for the results from Question 1 d) ? Why might this be the case?



When differentiating the results of Questions 1 and 2, even though Question 2 performed better than Question 1 despite only having the first 20 items.

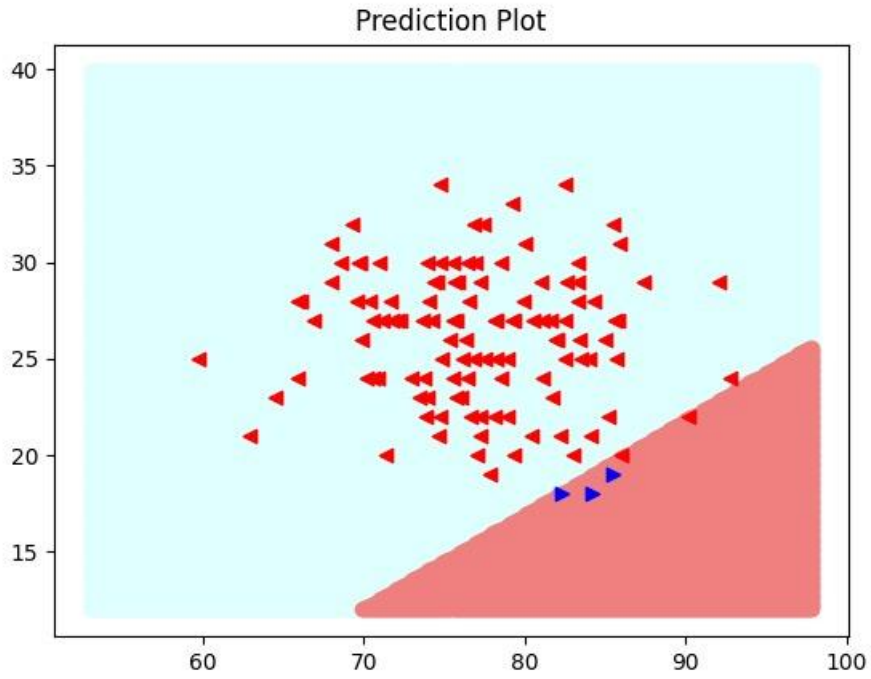
2e) Given the results form parts c) and d), do you believe the data set you used was actually derived from a function that is consistent with the function format in Question 1 ? Justify your answer.

No it is not consistent because question 1 didn't meet the requirements as it has high Mean Squared error compared to question 2 which has low Mean squared error.

Q3 Logistic Regression

3a,3b)Implement logistic regression to classify this data (use the individual data elements, i.e. height, weight, and age, as features). Your implementation should take different data sets as input for learning.

Plot the resulting separating surface together with the data. To do this plotting you need to project the data and function into one or more 2D space. The best visual results will be if projection is done along the separating hyperplane (i.e. into a space described by the normal of the hyperplane and one of the dimension within the hyperplane)



3c) Evaluate the performance of your logistic regression classifier in the same way as for Project 1 using leave-one out validation and compare the results with the ones for KNN and Naïve Bayes Discuss what differences exist and why one method might outperform the others for this problem.

By comparing KNN and Naive Bayes the accuracy is 47.5% .maximum accuracy is achieved through KNN it is best fit for our data even though it have a drawback which is not fast as expected

3d) Repeat the evaluation and comparison from part c) with the age feature removed. Again, discuss what differences exist and why one method might outperform the others in this case

After removing the Age the accuracy is 55.83%.Even with the different approaches with KNN and naive bias both ended up on the same stand.