ANSWER

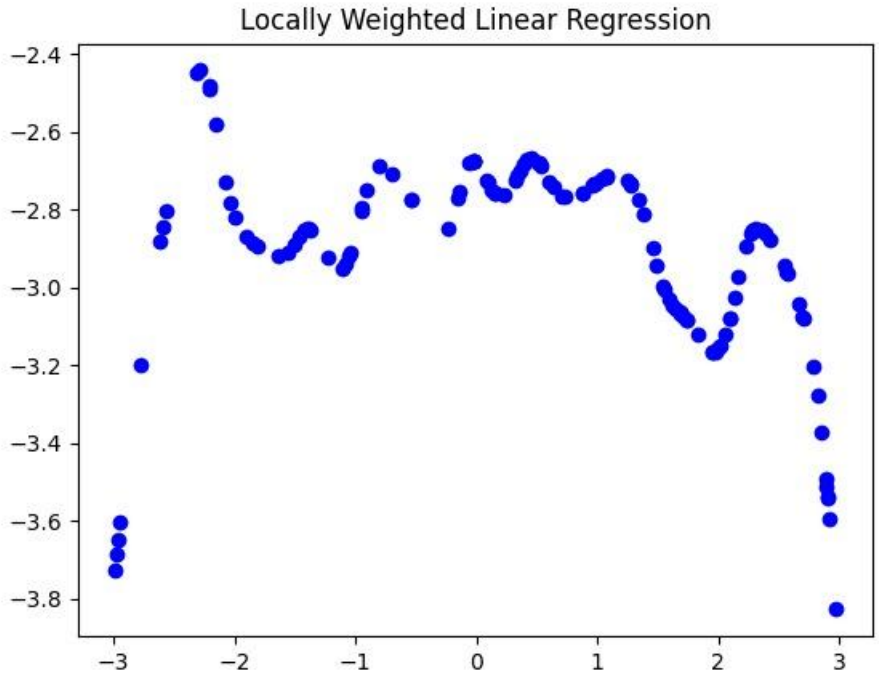
1c)Evaluate your regression functions by computing the error on the test data points that were generated for Question 1c). Compare the error results and try to determine for what "function depths" overfitting might be a problem. Which 'function depth' would you consider the best prediction function and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

By Observing the MSE from graphs from depth 0 to depth 6 the MSE increases from depth 0 to depth 6 the best fit is found at 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?

The difference while evaluating the first 20 elements is not much.

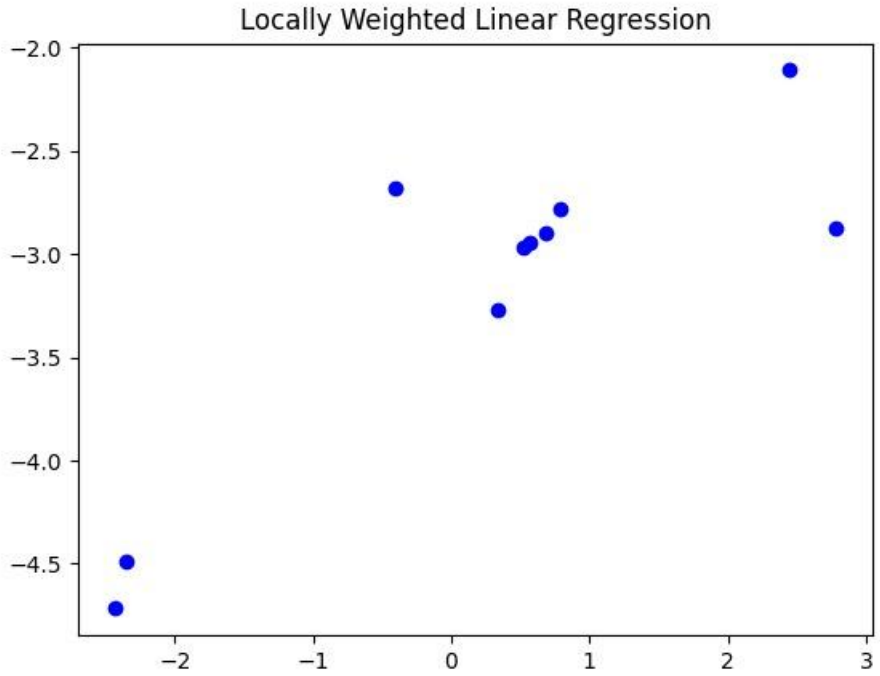
2a,2b)



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 locally weighted linear regression is better compared to question 1c

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?



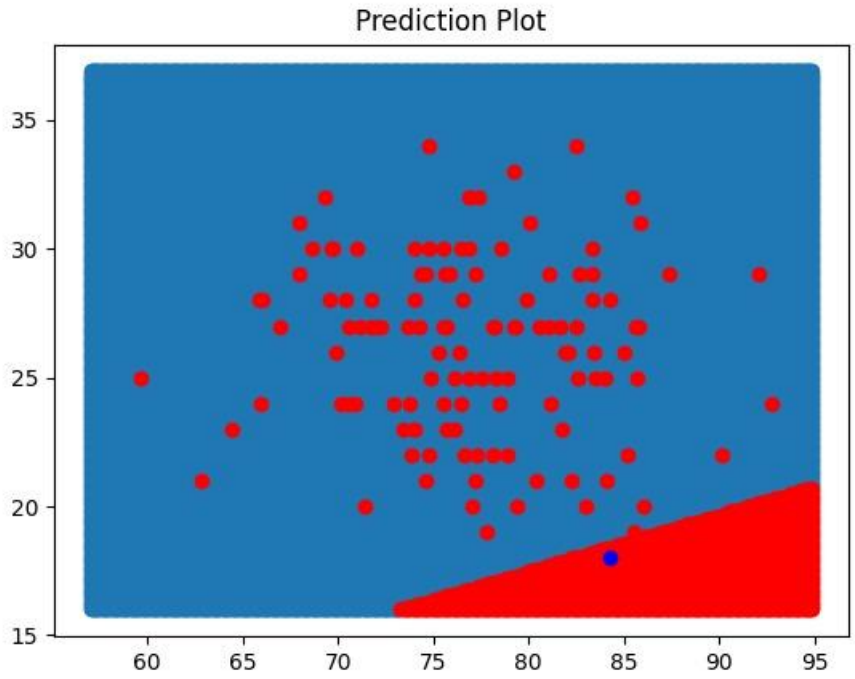
Question 2 performs better compared to question 1 by using only the first 20 elements of the training dataset.

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 to best fit with any model the MSE should be low in this case question 2 is having low MSE compared to the question 1.

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.

Among KNN and Naive Bayes. KNN gives highest accuracy . KNN is not much faster but it meets with our requirements.

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

When age is removed KNN and Naive Bayes gives same results but Logistics regression fails.