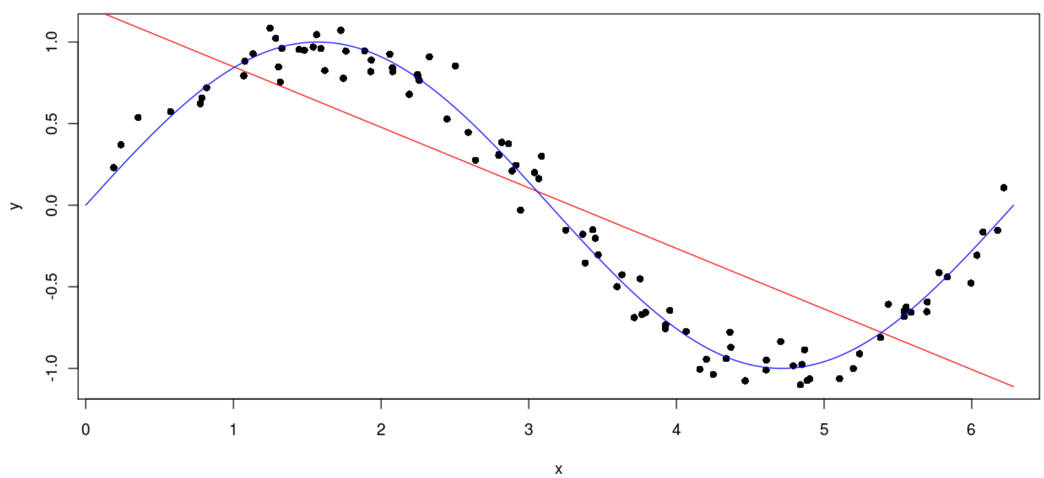
Statistics 745 **Ekrem Kocaguneli**

Group Assignment 3

1. Implement your own least squares classifier in R.

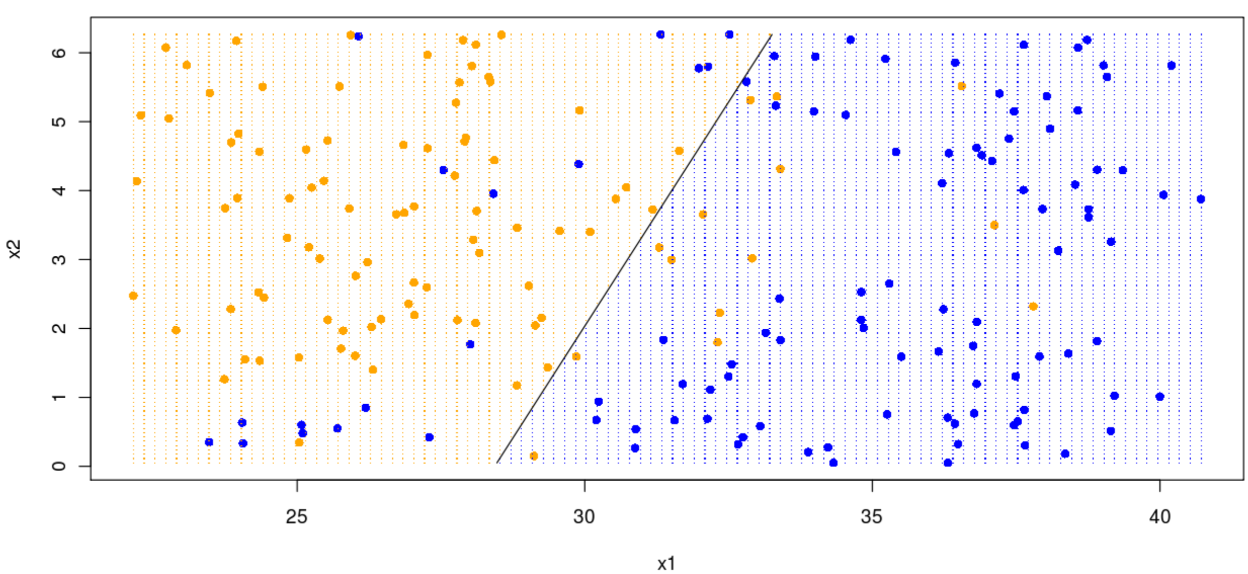
(a) Apply it to the sin(x) regression example provided as supplementary material.

Below is the graph when regression is applied to sin(x). Note that I also included the intercept calculation (i.e. apped 1's to x) when solving the regression.



(b) Apply it to the simulated classification example.

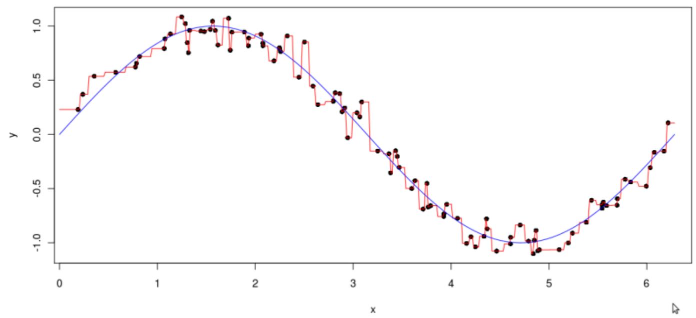
Below is the graph that is elicited when least square regression is applied to simulated classification example. Again in this example I used the intercept.

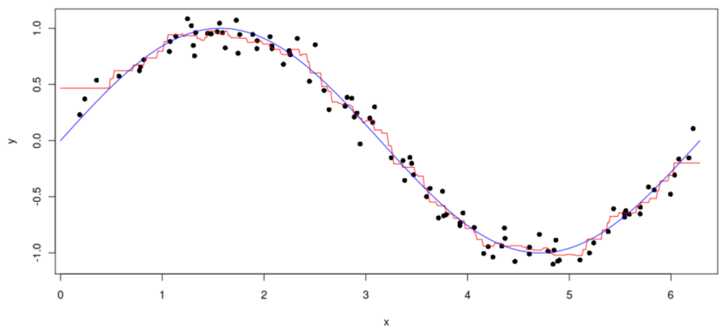


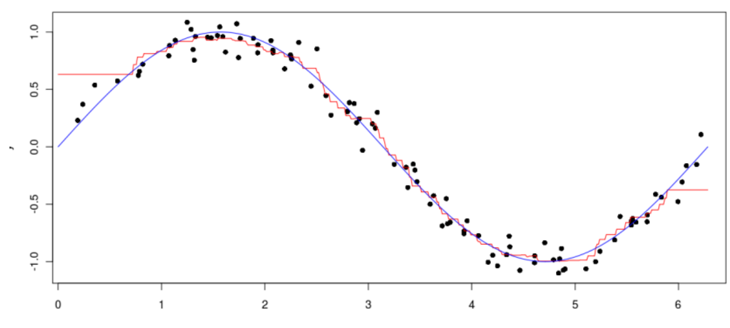
2. Implement your own k-NN function.

(a) Apply it to the sin(x) regression example provided as supplementary material. Investigate the effect of k.

Below are 3 graphs for 3 different k values: 1, 5 and 10 respectively.



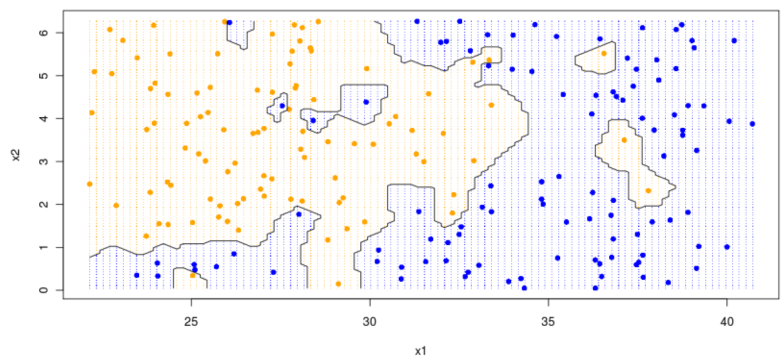


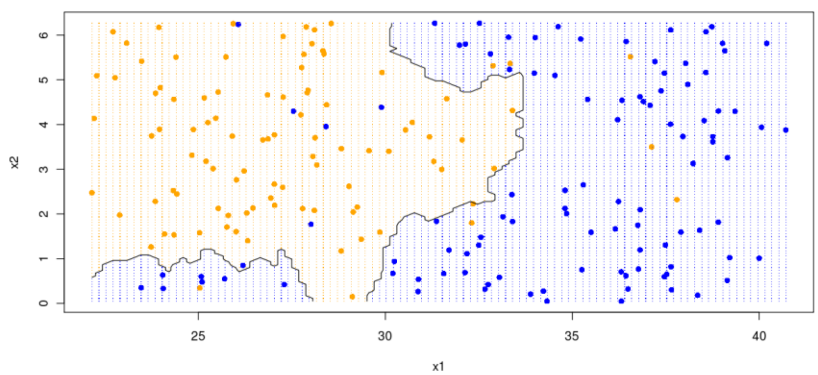


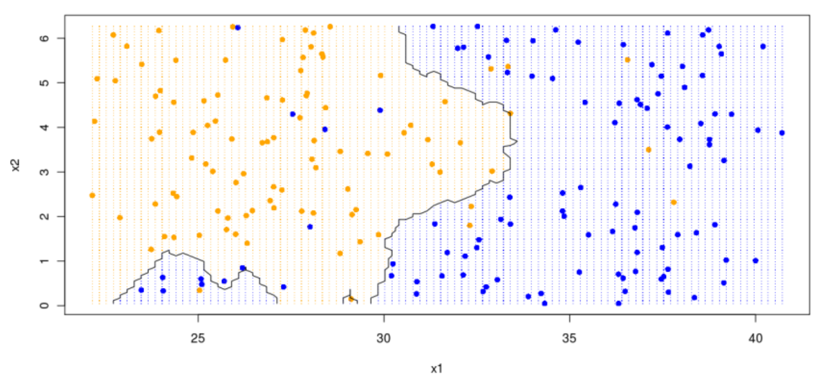
As can be seen from the above plots, as the k-value increases, the line becomes closer to the actual regression line and the variance decreases and bias increases (see the left and right tails becoming horizontal, meaning that estimated and actual values on the tails are biased, i.e. estimated values are closer to majority values rather than actual values). However, for low k-values, the variance is very high, the learner tries to memorize each individual point; on the other hand, the bias is very low (the estimated values on right and left tails are still close to actual values).

(b) Apply it to the simulated classification example. Investigate the effect of k.

Below are 3 graphs for 3 different k values: 1, 5 and 10 respectively.







As can be seen from above figures, as the k-value increases the variance decreases and bias increases (less and less jiggles and less and less islands). On the contrary, as the k-value decreases, the variance increases, meaning that single points have more impact than majority of points; similarly, the bias decreases, meaning that the impact of majority decreases.