

Tutorial 4 On-site Questions1. The  $K$ -nearest neighbor classifier

The table below provides a training data set containing six observations, three predictors, and one qualitative response variable,  $Y$ .

Obs	$X_1$	$X_2$	$X_3$	$Y$
1	0	3	0	Red
2	2	0	0	Red
3	0	1	3	Red
4	0	1	2	Green
5	-1	0	1	Green
6	1	1	1	Red

Suppose we wish to use this data set to make a prediction for a new observation with response  $Y$  given that it has  $X_1 = X_2 = X_3 = 0$  using  $K$ -nearest neighbors.

- (a) Compute the Euclidean distance between each observation and the test point,  $X_1 = X_2 = X_3 = 0$ .
- (b) What is our prediction with  $K = 1$ ? Why?
- (c) What is our prediction with  $K = 3$ ? Why?
- (d) If the Bayes decision boundary (the gold standard decision boundary) in this problem is highly non-linear, then would we expect the best value for  $K$  to be large or small? Why?