

Use MATLAB for the following exercises, and hand in the code along with your results.

1. Compare the classification performance of linear regression and k -nearest neighbor classification on the `zipcode` data. In particular, consider only the 2's and 3's, and $k = 1, 3, 5, 7$ and 15. Show both the training and test error for each choice. The `zipcode` data is available from the book website www-stat.stanford.edu/ElemStatLearn.

Let the labels for the 2's and 3's be $\{0, 1\}$. What happens to the regression results when the labels are changed to $\{-1, 1\}$? Explain your findings.

2. Generate 50 points uniformly at random on $[-1, 1]^d$. Plot the l_∞ nearest neighbor distance to the origin, for $d = 2 \dots 100$. What do you observe? How do your results compare with Formula 2.24 (page 23)? Repeat the experiment for 5 and 500 points, and comment on the differences.