

distribution, but only a single variable  $x_1$  is relevant to the output. Suppose further that this feature simply encodes the output directly, i.e. that  $y = x_1$  in all cases. Nearest neighbor regression will not be able to detect this simple pattern. The nearest neighbor of most points  $\mathbf{x}$  will be determined by the large number of features  $x_2$  through  $x_{100}$ , not by the lone feature  $x_1$ . Thus the output on small training sets will essentially be random.