1 Nearest Neighbor Methods

1.1 Nearest Neighbor for Classification

We are given a set of n ordered pairs (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) where $y_i \in C$ with $C = \{c_1, ..., c_M\}$ being a set of labels. The nearest neighbor classification $\hat{y}$ for a new point x is y_j where j is the following:

$$\arg\min_j ||x - x_j||$$

Let $x_1, x_2, ..., x_k$ be the k closest points to x. Let I be those indicies 1 to k. The k-nearest nearest neighbor classification for x is the label c_m where m is the following:

$$\arg\max_m \mathbb{1}_I(y_i = c_m)$$

1.2 Nearest Neighbor for Regression

We are given a set of n ordered pairs (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) where $y_i \in \mathbb{R}$. The nearest neighbor regression value $\hat{y}$ for a new point x is y_j where j is the following:

$$\arg\min_j ||x - x_j||$$

Let $x_1, x_2, ..., x_k$ be the k closest points to x. Let I be those indicies 1 to k. The k-nearest nearest neighbor regression for x is the mean of the y_i values:

$$\frac{1}{k}\sum_{i \in I} y_i$$

2 Bayes Classifier

Estimate probability of class $p(y = c)$ using a probability model.

2.1 Multivariate Gaussian

2.2 Naive Bayes