Curse of dimensionality

- The density of samples decreases with increase in dimension

Curse of dimensionality

- The density of samples decreases with increase in dimension
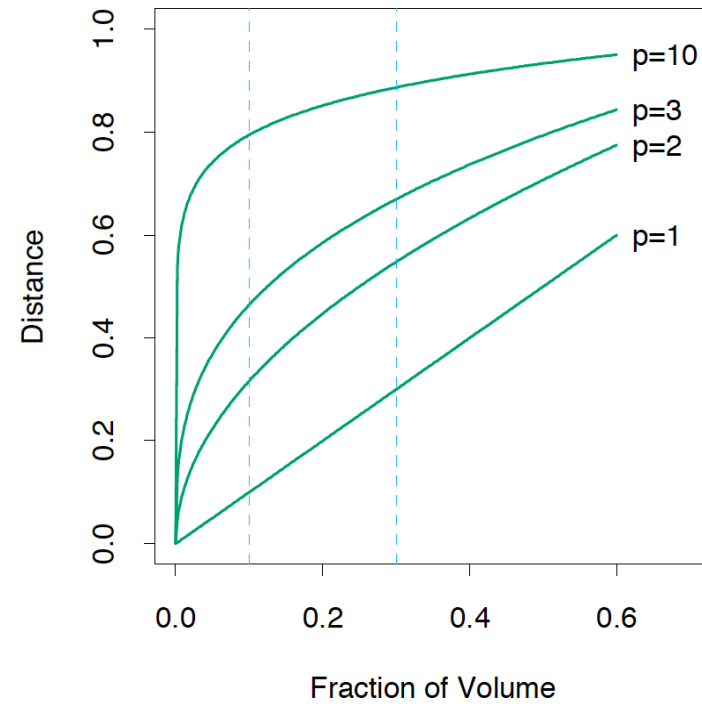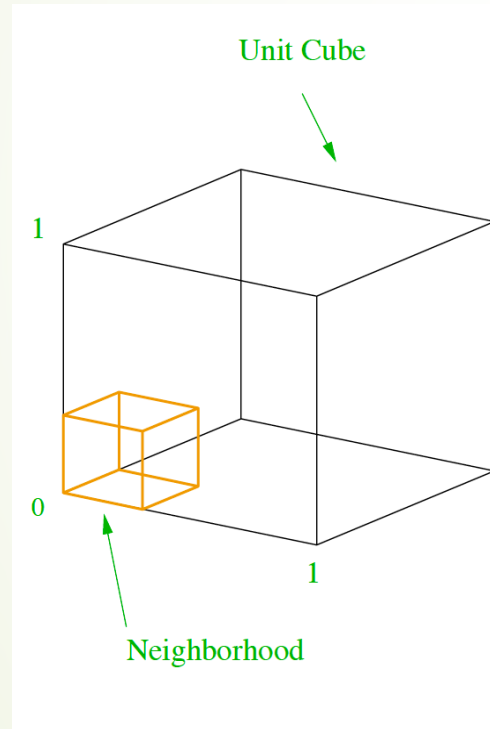- Lack of close neighbors causes increases in bias and variance

Curse of dimensionality

- The density of samples decreases with increase in dimension
- Lack of close neighbors causes increases in bias and variance
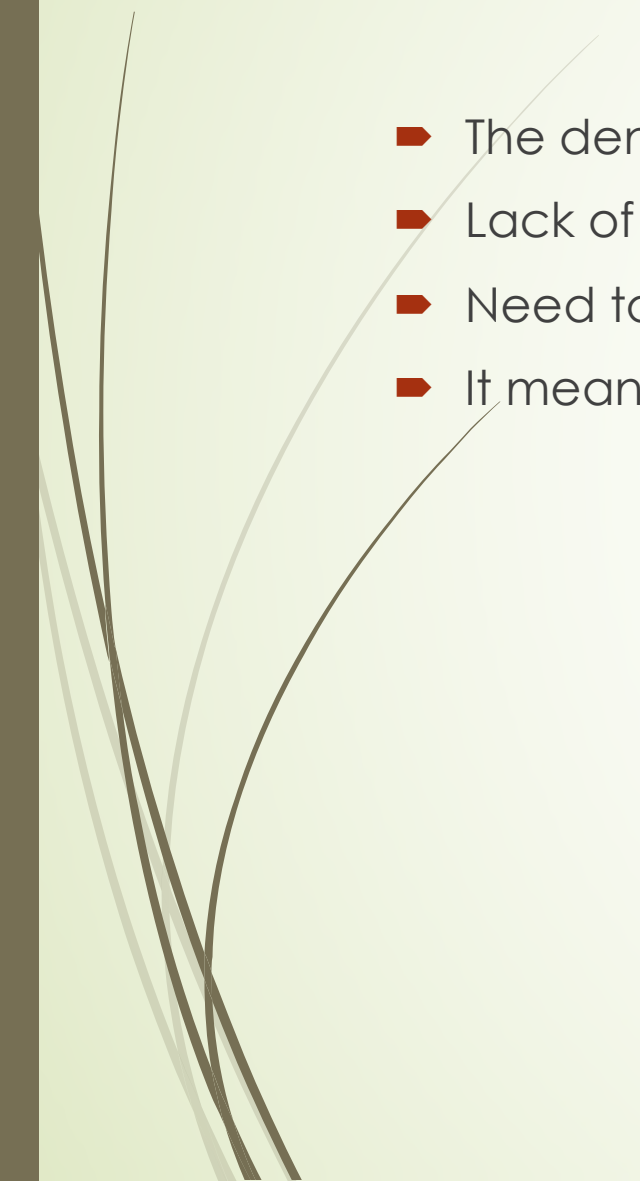- Need to look at almost the whole space to make a prediction

Curse of dimensionality

- The density of samples decreases with increase in dimension
- Lack of close neighbors causes increases in bias and variance
- Need to look at almost the whole space to make a prediction
- It means you average over points that are quite different

KNN Classification

KNN Classification

- Categorical variables belonging to some classes

KNN Classification

- Categorical variables belonging to some classes
- The probability that a location x gets assigned to class j is approximated by

$$\hat{p}_k(x) = \frac{1}{K} \sum_{i \in N_K(x)} \mathbb{I}(y_i = j)$$

KNN Classification

- Categorical variables belonging to some classes
- The probability that a location x gets assigned to class j is approximated by

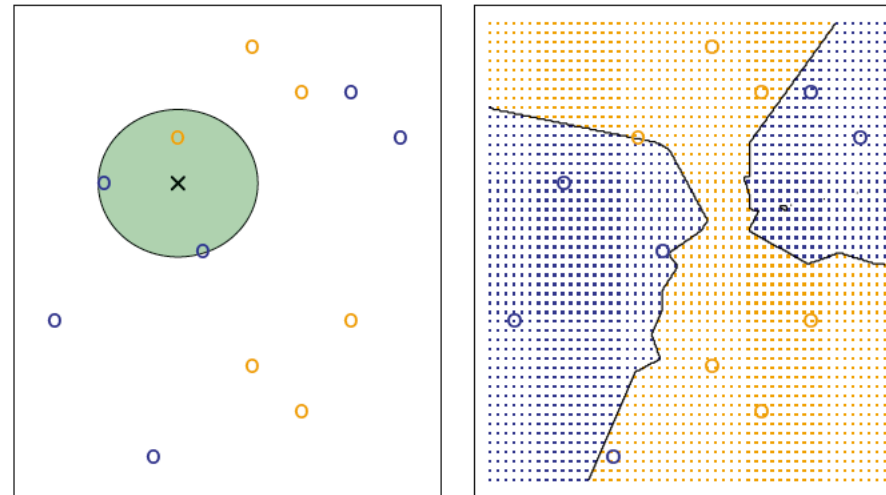$$\hat{p}_k(x) = \frac{1}{K} \sum_{i \in N_K(x)} \mathbb{I}(y_i = j)$$

- Assign the class with highest probability. This is like taking majority votes.

KNN Classification

- Categorical variables belonging to some classes
- The probability that a location x gets assigned to class j is approximated by

$$\hat{p}_k(x) = \frac{1}{K} \sum_{i \in N_K(x)} \mathbb{I}(y_i = j)$$

- Assign the class with highest probability. This is like taking majority votes.

Next up: Linear regression, logistic regression, decision trees

- Go to [this page](this%20page)