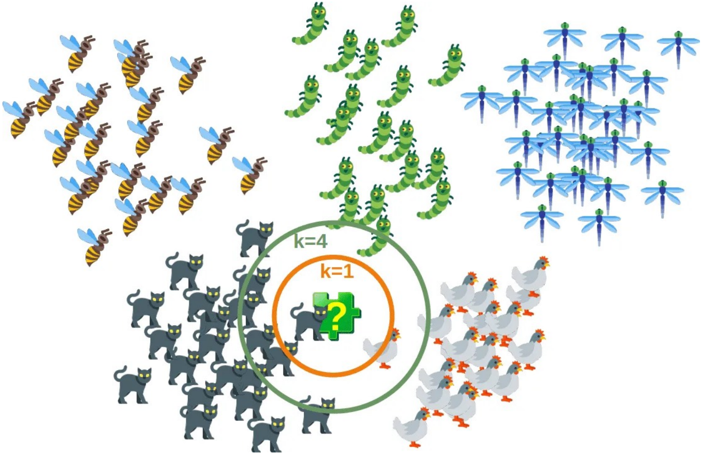
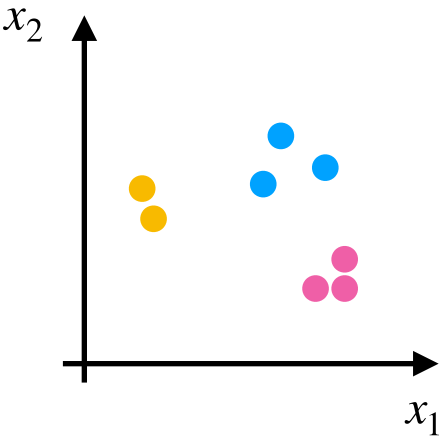
Nearest Neighbors



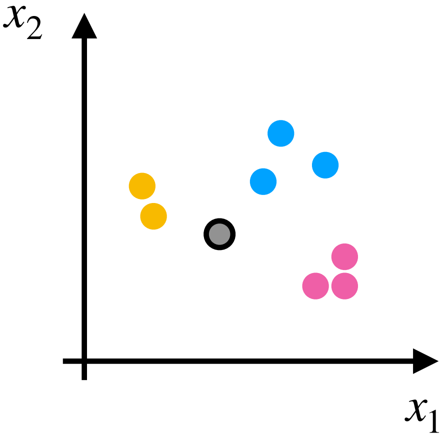
- K-nearest neighbors of a new testing point x: data points that have the k smallest distance to x.

Nearest Neighbors



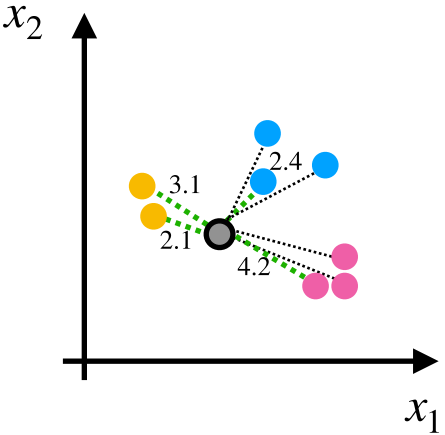
Collect training data.

Nearest Neighbors



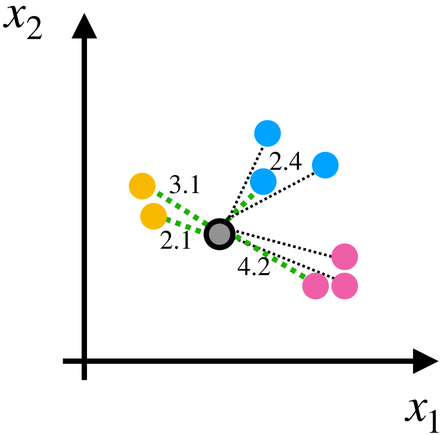
For a new testing gray point, you will want to classify the point to yellow, blue, or pink.

Nearest Neighbors



Calculate distance to find nearest neighbors:

Nearest Neighbors



Majority voting to decide class label:

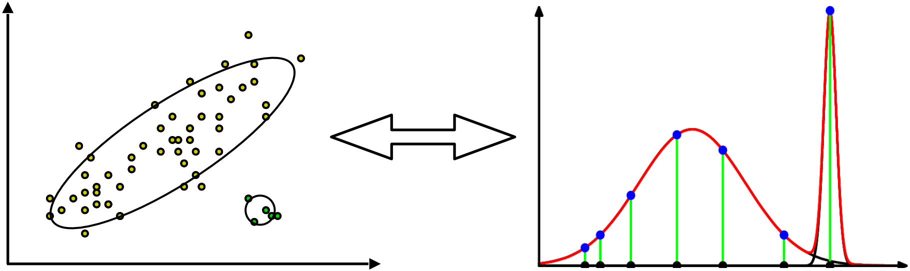
Maximum Likelihood Estimation

Foundations of Data Analysis

February 13, 2023

Why Maximum Likelihood?

Goal of MLE is to find the best distributions to fit your data.



Data

Parametric distribution

Likelihood as Joint Probability Function

θ is a parameter, for example, mean and variance of Gaussian.

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i; \theta)$$

Maximum a Likelihood

Maximize the likelihood function to estimate θ:

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i; \theta)$$

(See class notes for derivations of MLE of Gaussians).

MLE of Bernoulli distribution

$X \sim \text{Ber}(\theta)$, θ is the probability of x_i taking value one,
$1 - \theta$ is the probability of x_i taking value zero .

For *n* data points, we assume *k* points take the value one.

$$L(\theta \mid x_1, \ldots, x_n) = \theta^k (1-\theta)^{n-k}, \text{where } \mathrm{k} = \sum_{\mathrm{i}} \mathrm{x_i}$$

MLE of Bernoulli distribution

$$X \sim \mathrm{Ber}(\theta)$$

$$L(\theta \,|\, x_1, \ldots, x_n) = \theta^k (1-\theta)^{n-k}, \text{where k} = \sum_{\mathrm{i}} \mathrm{x_i}$$

$$\frac{dL}{d\theta} = k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1}$$

MLE of Bernoulli distribution

$$X \sim \mathrm{Ber}(\theta)$$

$$L(\theta \,|\, x_1, \ldots, x_n) = \theta^k (1-\theta)^{n-k}, \text{where } \mathrm{k} = \sum_{\mathrm{i}} \mathrm{x_i}$$

$$\frac{dL}{d\theta} = k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1}$$

$$= (k(1-\theta) - (n-k)\theta)\theta^{k-1}(1-\theta)^{n-k-1}$$

MLE of Bernoulli distribution

$$X \sim \mathrm{Ber}(\theta)$$

$$L(\theta \,|\, x_1, \ldots, x_n) = \theta^k (1-\theta)^{n-k}, \text{where } \mathrm{k} = \sum_{\mathrm{i}} \mathrm{x_i}$$

$$\frac{dL}{d\theta} = k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1}$$

$$= (k(1-\theta) - (n-k)\theta)\theta^{k-1}(1-\theta)^{n-k-1}$$

$$= (k - n\theta)\theta^{k-1}(1-\theta)^{n-k-1}$$

MLE of Bernoulli distribution

$$X \sim \mathrm{Ber}(\theta)$$

$$L(\theta \,|\, x_1, \ldots, x_n) = \theta^k (1-\theta)^{n-k}, \text{where } \mathrm{k} = \sum_{\mathrm{i}} \mathrm{x_i}$$

$$\frac{dL}{d\theta} = k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1}$$

$$= (k(1-\theta) - (n-k)\theta)\theta^{k-1}(1-\theta)^{n-k-1}$$

$$= (k - n\theta)\theta^{k-1}(1-\theta)^{n-k-1}$$

$$\frac{dL}{d\theta}(\hat{\theta}) = 0 \;\Rightarrow\; \hat{\theta} = \frac{k}{n}$$