Parameter estimation

Usually do not know distributions of classes

- Learn from training data!
- Two main approaches: Bayesian vs. maximum-likelihood

We will cover maximum-likelihood estimation first and defer Bayesian parameter estimation for later in the course.

Maximum-likelihood parameter estimation

Problem statement: Given a data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ (identically and independently distributed) and a parametric density function $p(\mathbf{x}|\theta)$ with parameters θ, find θ^* such that:

$$p(\mathbf{X}|\theta^*) > p(\mathbf{X}|\theta), \forall\theta.$$

Maximum-likelihood estimation

General approach:

$$p(\mathbf{X}|\theta) = \prod_{j=1}^{n} p(\mathbf{x}_j|\theta)$$

$$\ln p(\mathbf{X}|\theta) = \sum_{j=1}^{n} \ln p(\mathbf{x}_j|\theta)$$

Maximization — solve:

$$\nabla_\theta \ln p(\mathbf{X}|\theta) = 0$$

Maximum-likelihood estimation

Maximize:

$$\nabla_\theta \ln p(\mathbf{X}|\theta) = 0$$

Expand and solve:

$$\nabla_\theta \sum_{j=1}^{n} \ln p(\mathbf{x}_j|\theta) = 0$$

$$\sum_{j=1}^{n} \nabla_\theta \ln p(\mathbf{x}_j|\theta) = 0$$