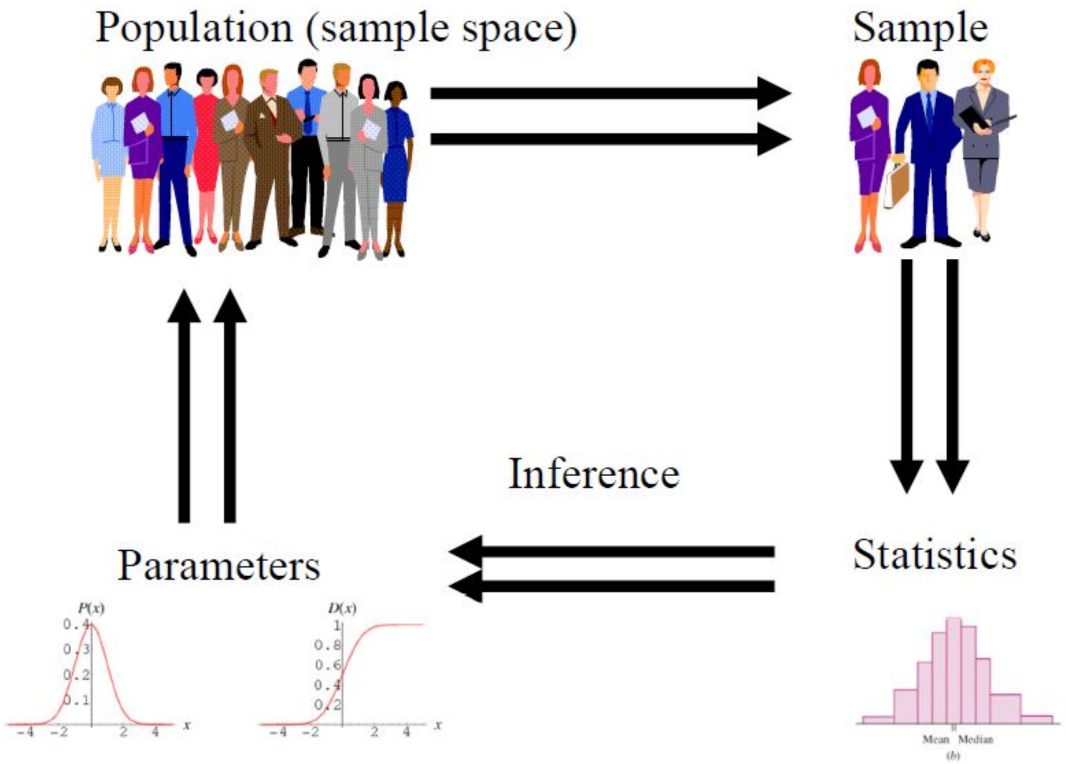
Chapter 4: Parametric Methods

Sample Statistics and Population Parameters

A Schematic Depiction



Parametric Methods

- Discussed how to make optimal decisions when the uncertainty is modeled using probabilities
- Now see how we can estimate these probabilities from a given training set.
- We start with the parametric approach for classification and regression.

- Statistical inference: make decision based on information provided by sample
- Parametric approach: Assume that the sample is drawn from a distribution that obeys a known model, e.g. Gaussian
- Advantage: Small number of parameters.
  - E.g. mean and variance - sufficient statistics of the distribution
- Estimate parameters of a distribution: Maximum likelihood method

Parametric estimation

$X = \{ x^t \}_t$ where $x^t \sim p(x)$

Parametric estimation:

Assume a form for $p(x \mid \theta)$ and estimate θ, its sufficient statistics, using X

e.g., $N(\mu, \sigma^2)$ where $\theta = \{\mu, \sigma^2\}$

Maximum likelihood method/estimation

Maximum Likelihood Estimation

$\mathcal{X} = \{x^t\}_{t=1}^N$ independent and identically distributed (iid) sample

$x^t \sim p(x|\theta)$ x^t are instances drawn from some known probability density family, $p(x|\theta)$,

find θ that makes sampling x^t from $p(x|\theta)$ as likely as possible

$l(\theta|\mathcal{X}) = p(\mathcal{X}|\theta) = \prod_{t=1}^{N} p(x^t|\theta)$ the *likelihood* of sample $\mathcal{X}$

In maximum likelihood estimation, we are interested in finding θ that makes X the most likely to be drawn. We thus search for θ that maximizes the likelihood of the sample, which we denote by l(θIX).

$l(\theta)$

Maximum Likelihood Estimation

Likelihood of θ given the sample X

$$l(\theta|\mathcal{X}) = p(\mathcal{X}|\theta) = \prod_{t=1}^{N} p(x^t|\theta)$$

Log-likelihood of θ given the sample X

$$\mathcal{L}(\theta|\mathcal{X}) \equiv \log l(\theta|\mathcal{X}) = \sum_{t=1}^{N} \log p(x^t|\theta)$$

Maximum likelihood estimator:

Maximum Likelihood Estimation

- Because logarithm will not change the value of θ when it take its maximum (monotonically increasing/decreasing)
  - Finding θ that maximizes the likelihood of the instances is equivalent to finding θ that maximizes the log likelihood of the samples

$$a \ge b$$
$$\Rightarrow \log a \ge \log b$$

$$L(\theta|\mathbf{x}) = \log l(\theta|\mathbf{x}) = \sum_{t=1}^{N} \log p(x^t|\theta)$$

  - As we shall see, logarithmic operation can further simplify the computation when estimating the parameters of those distributions that have exponents

Bernoulli Density

A random variable X takes either x=1 (with prob. p) or x=0.

The associated probability distribution: $P(x) = p^x(1-p)^{1-x}, x \in \{0,1\}$

Likelihood: $l(\theta|\mathcal{X}) = p(\mathcal{X}|\theta) = \prod_{t=1}^{N} p(x^t|\theta)$

Maximize log-likelihood

$$\begin{aligned} \mathcal{L}(p|\mathcal{X}) &= \log \prod_{t=1}^{N} p^{(x^t)}(1-p)^{(1-x^t)} \\ &= \sum_t x^t \log p + \left(N - \sum_t x^t\right) \log(1-p) \end{aligned}$$

$$\hat{p} = \frac{\sum_t x^t}{N}$$

Gaussian Density

X is Gaussian (normal) distributed with mean $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$, denoted as $\boldsymbol{N}(\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2})$, if its density function is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty$$

Given a sample $\mathcal{X} = \{x^t\}_{t=1}^N$ with $x^t \sim \mathcal{N}(\mu, \sigma^2)$, the log likelihood of a Gaussian sample is

$$\mathcal{L}(\mu, \sigma | \mathcal{X}) = -\frac{N}{2}\log(2\pi) - N\log\sigma - \frac{\sum_t (x^t - \mu)^2}{2\sigma^2}$$

$$m = \hat{\mu} = \frac{\sum_{t=1}^{N} x^t}{N} \qquad s^2 = \hat{\sigma}^2 = \frac{\sum_{t=1}^{N} \left(x^t - m\right)^2}{N}$$