

$$P(\omega_i | x) = \frac{P(x | \omega_i) P(\omega_i)}{P(x)}$$

$P(\omega_i)$  → prior  
 $P(x)$  → marginal probability

(posterior)

Decision rule,

$$P(\omega_1 | x) > P(\omega_2 | x) \Rightarrow \omega_1$$

$$P(\omega_1 | x) < P(\omega_2 | x) \Rightarrow \omega_2$$

$$P(x | \omega_1) P(\omega_1) > P(x | \omega_2) P(\omega_2) \Rightarrow \omega_1$$

Decision is made based on a priori probabilities and class conditional probabilities.

Bayesian Decision Theory helps to give us the minimum possible error.

if  $y = 1, 2, 3, 4$

then  $p(x, y)$

stands for all of the below

$$p(x, y=1), p(x, y=2), p(x, y=3), p(x, y=4)$$

2.25

## Parametric Methods

### 1. Maximum Likelihood Estimate

Generally when we are given the data, we intend to find and learn the distributions.

Using Bayes' Theorem, for finding likelihood

$$P(\theta|D) \propto P(D|\theta) \cdot P(\theta)$$

↓  
likelihood

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

$$\theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix} \quad \theta_1 = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \theta_2 = \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix}$$

Suppose we draw  $N$  data from a known distribution.  $P(x|\theta)$

$$x^t = P(x|\theta)$$

$$x = \{x^t\}_{t=1}^N$$

$$L(\theta|x) = \log l(\theta|x)$$

$$\left. \begin{array}{l} \text{log likelihood} \\ \text{function} \end{array} \right\} \sum_{t=1}^N \log p(x^t|\theta)$$

Bernoulli Distribution ( $p$ )

$$P(x) = p^x (1-p)^{1-x}; \quad x \in \{0,1\}$$

$$L(p|x) = \log \prod_{t=1}^N p(x^t) (1-p)^{(1-x^t)}$$

$$\Rightarrow \sum_t x^t \log p + (N - \sum_t x^t) \log (1-p)$$

Maximize

(Estimate)  $\hat{\theta}$

$$\hat{\theta} = \frac{\sum_{i=1}^N x^i}{N}$$

Bias And Variance

$x$  and  $\theta$

Estimator  $d = d(x)$

of  $\theta$

$$(d(x) - \theta)^2$$

$$r(d, \theta) = E \left[ (d(x) - \theta)^2 \right]$$

(Mean Square Error)

Bias of an estimator

$$b_{\theta}(d) = E[d(x)] - \theta$$

if bias is zero then it is an unbiased estimator.

if  $b_{\theta}(d) = 0$ , for all  $\theta$  values, then it is an unbiased estimator.

$$E[m] = E \left[ \frac{\sum_t x^t}{N} \right]$$

$$= \frac{1}{N} \sum_t \frac{x^t}{N} = \frac{N_p}{N} = \mu$$

$$\text{Var}(m) = \text{Var} \left[ \frac{\sum_t x^t}{N} \right] = \frac{\sigma^2}{N}$$

Formula,

variance

$$\gamma(d, \theta) = E[(d - E(d))^2]$$

$$\approx \underbrace{E(d) - \theta}_{\text{bias}}^2$$

$$\gamma(d, \theta) = \text{var}(d) + \text{bias}(d)^2$$

$$\text{error} = \text{variance} + \text{bias}$$