CHAPTER 4:

# Parametric Methods

# Introduction

- Use the Model is defined with small number of parameters

- Assume some distribution and estimate the parameters of the distribution

- Use the distribution to make decisions

# Parametric Estimation

- $\mathcal{X} = \{ x^t \}_t$ where $x^t \sim p\,(x)$
- Parametric estimation:

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

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

# Maximum Likelihood Estimation

***Bias rule:*** $p\ (\theta\,|\,\mathcal{X}) = p\ (\mathcal{X}\,|\,\theta)\ \mathrm{p}(\theta) / p\ (\mathcal{X})$

- ***Likelihood*** of $\theta$ given the sample $\mathcal{X}$

  $$l\ (\vartheta\,|\,\theta) = p\ (\mathcal{X}\,|\,\theta) = \prod_t p\ (x^t\,|\,\theta)$$

**Goal:** given $\mathcal{X}$ what is the best $\theta$ which describes data

- ***Log likelihood***

  $$\mathcal{L}(\theta\,|\,\mathcal{X}) = \log l\ (\theta\,|\,\mathcal{X}) = \sum_t \log p\ (x^t\,|\,\theta)$$

- ***Maximum likelihood estimator (MLE)***

  $$\theta^* = \mathrm{argmax}_\theta\, \mathcal{L}(\vartheta\,|\,\theta)$$

# Examples: Bernoulli/Multinomial

- *Bernoulli*: Two states, failure/success, $x$ in {0,1}

$P(x) = p_o^{x} (1 - p_o)^{(1 - x)}$

$$\mathcal{L}(p_o | \mathcal{X}) = \log \prod_t p_o^{x^t} (1 - p_o)^{(1 - x^t)}$$

MLE: $p_o = \sum_t x^t / N$

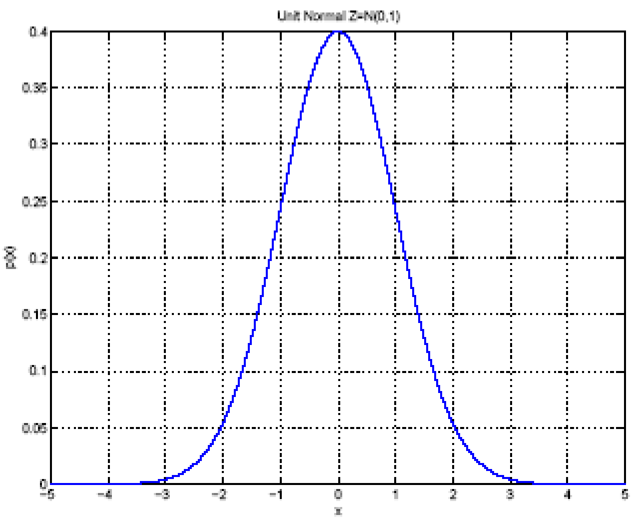
- *Multinomial*: $K$>2 states, $x_i$ in {0,1}

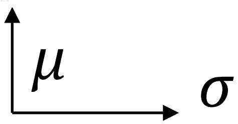
$P(x_1, x_2, \ldots, x_K) = \prod_i p_i^{x_i}$

$$\mathcal{L}(p_1, p_2, \ldots, p_K | \mathcal{X}) = \log \prod_t \prod_i p_i^{x_i^t}$$

MLE: $p_i = \sum_t x_i^t / N$

# Gaussian (Normal) Distribution




- $p(x) = \mathcal{N}(\mu, \sigma^2)$

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

- MLE for $\mu$ and $\sigma^2$:

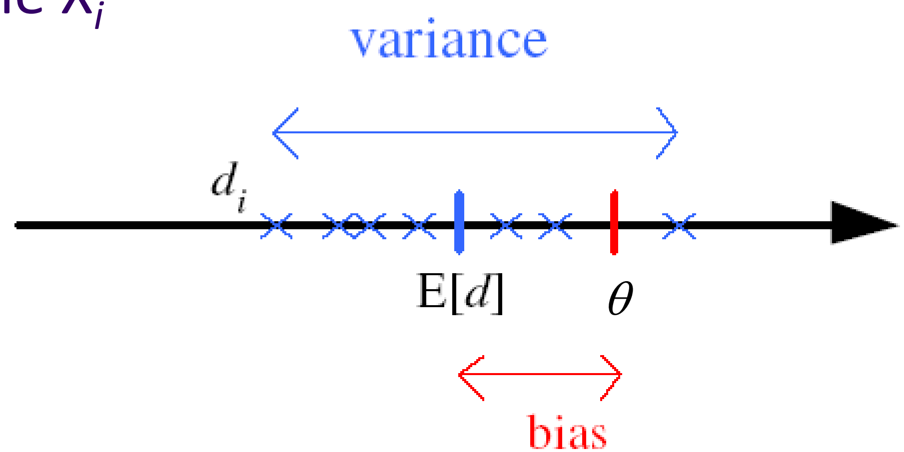
$$m = \frac{\sum_t x^t}{N}$$

$$s^2 = \frac{\sum_t (x^t - m)^2}{N}$$

# Bias and Variance

Unknown parameter $\theta$

Estimator $d_i = d\,(X_i)$ on sample $X_i$




Bias: $b_\theta(d) = E\,[d] - \theta$

Variance: $E\,[(d - E\,[d])^2]$

Mean square error:

$$r\,(d,\theta) = E\,[(d-\theta)^2]$$
$$= (E\,[d] - \theta)^2 + E\,[(d - E\,[d])^2]$$
$$= \text{Bias}^2 + \text{Variance}$$

# Bayes' Estimator

- Suppose we do know something about the parameter
- Treat $\theta$ as a random var with prior $p(\theta)$
- Bayes' rule: $p(\theta \mid \mathcal{X}) = p(\mathcal{X} \mid \theta)\, p(\theta) / p(\mathcal{X})$
- Full: $p(x \mid \mathcal{X}) = \int p(x \mid \theta)\, p(\theta \mid \mathcal{X})\, d\theta$
- Maximum a Posteriori (MAP): $\theta_{\text{MAP}} = \text{argmax}_{\theta}\, p(\theta \mid \mathcal{X})$
- Maximum Likelihood (ML): $\theta_{\text{ML}} = \text{argmax}_{\theta}\, p(\mathcal{X} \mid \theta)$
- Bayes': $\theta_{\text{Bayes'}} = E[\theta \mid \mathcal{X}] = \int \theta\, p(\theta \mid \mathcal{X})\, d\theta$

# Bayes' Estimator: Example

- $x^t \sim \mathcal{N}(\theta, \sigma_o^2)$ and $\theta \sim \mathcal{N}(\mu, \sigma^2)$
- $\theta_{ML} = m$
- $\theta_{MAP} = \theta_{Bayes'} =$

$$E[\theta \mid \mathcal{X}] = \frac{N/\sigma_0^2}{N/\sigma_0^2 + 1/\sigma^2} m + \frac{1/\sigma^2}{N/\sigma_0^2 + 1/\sigma^2} \mu$$

# Parametric Classification

$$g_i(x) = p(x \mid C_i)P(C_i)$$

or

$$g_i(x) = \log p(x \mid C_i) + \log P(C_i)$$

$$p(x \mid C_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right]$$

$$g_i(x) = -\frac{1}{2}\log 2\pi - \log \sigma_i - \frac{(x-\mu_i)^2}{2\sigma_i^2} + \log P(C_i)$$

- Given the sample $\mathcal{X} = \{x^t, r^t\}_{t=1}^{N}$

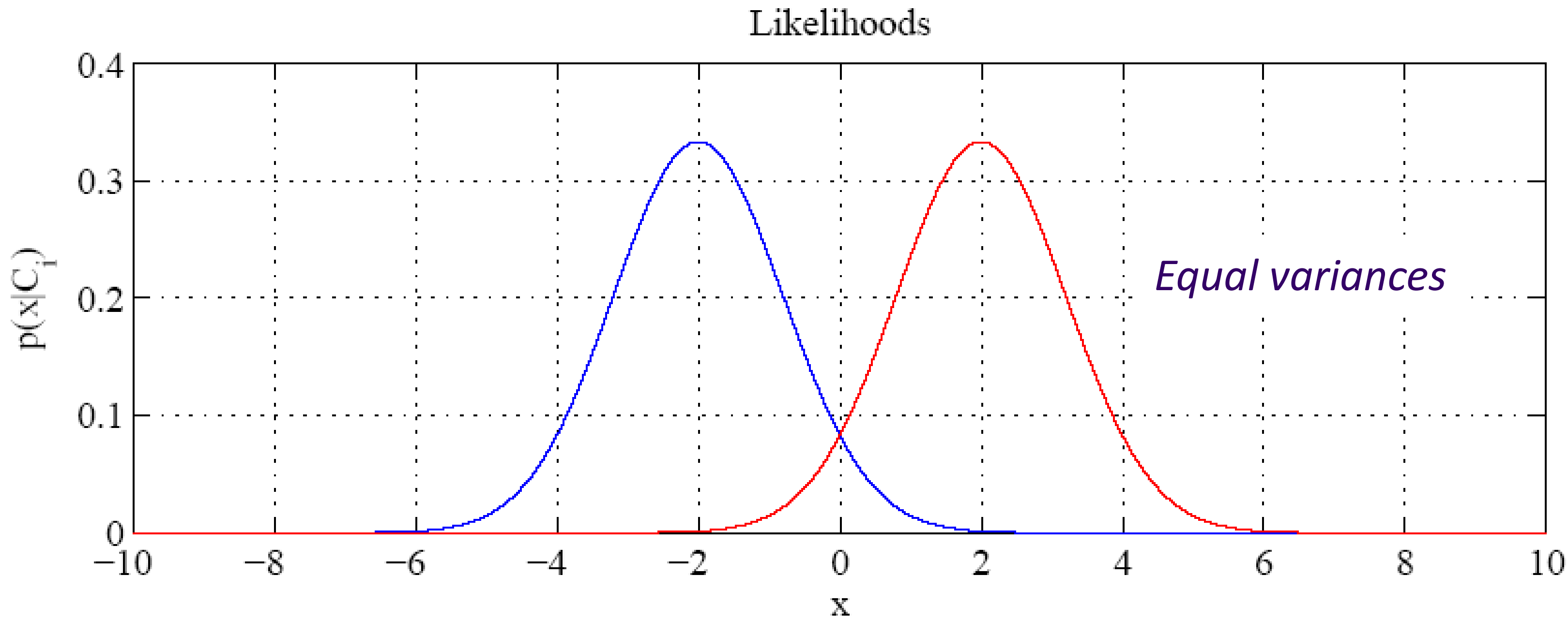
$$x \in \Re \qquad r_i^t = \begin{cases} 1 \text{ if } x^t \in C_i \\ 0 \text{ if } x^t \in C_j, j \neq i \end{cases}$$

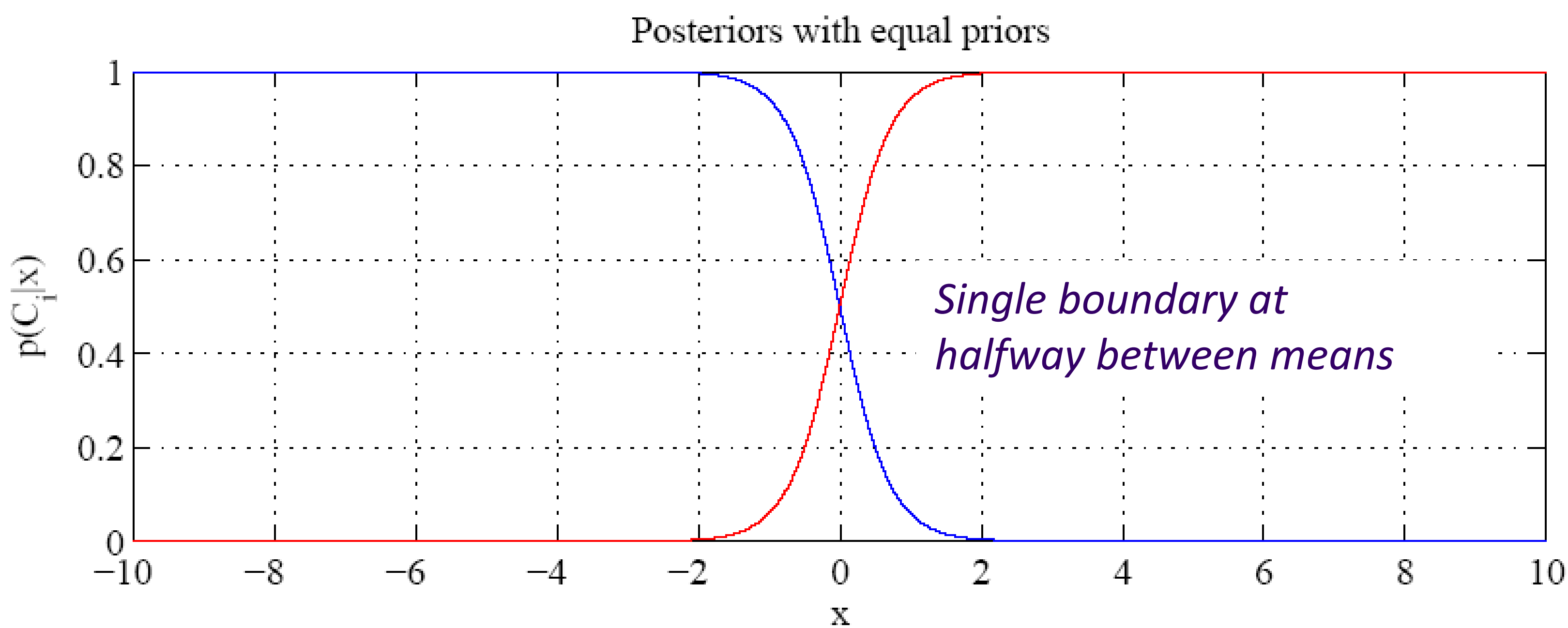
- ML estimates are

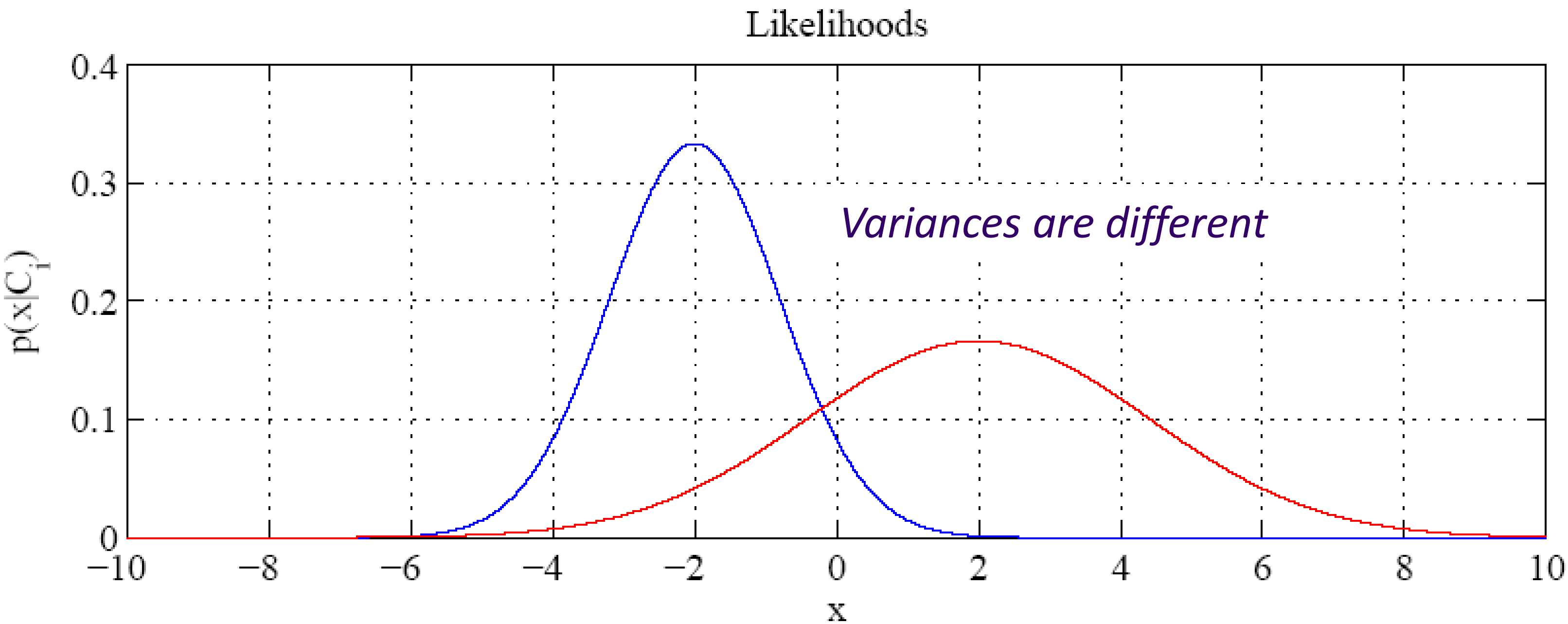
$$\hat{P}(C_i) = \frac{\sum_t r_i^t}{N} \quad m_i = \frac{\sum_t x^t r_i^t}{\sum_t r_i^t} \quad s_i^2 = \frac{\sum_t (x^t - m_i)^2 r_i^t}{\sum_t r_i^t}$$

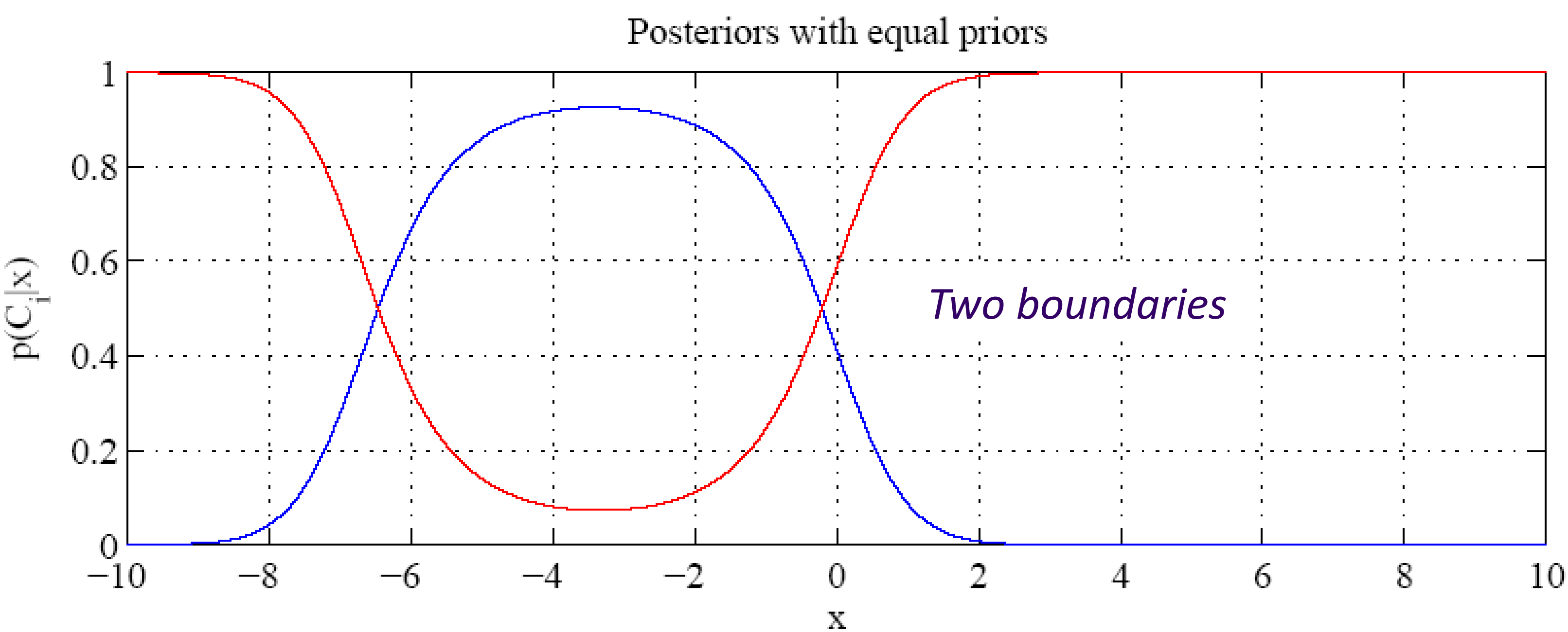
- Discriminant becomes

$$g_i(x) = -\frac{1}{2}\log 2\pi - \log s_i - \frac{(x - m_i)^2}{2 s_i^2} + \log \hat{P}(C_i)$$


Likelihoods
0.4
0.3
0.2
0.1
0
p(x|C_i)
−10
−8
−6
−4
−2
0
2
4
6
8
10
x
Equal variances


Posteriors with equal priors
1
0.8
0.6
0.4
0.2
0
p(C_i|x)
−10
−8
−6
−4
−2
0
2
4
6
8
10
x
Single boundary at
halfway between means


Likelihoods
0.4
0.3
0.2
0.1
0
p(x|C_i)
−10 −8 −6 −4 −2 0 2 4 6 8 10
x
Variances are different


Posteriors with equal priors
1
0.8
0.6
0.4
0.2
0
p(C_i|x)
−10 −8 −6 −4 −2 0 2 4 6 8 10
x
Two boundaries