BLG 454E Learning From Data

FALL 2022-2023

Assoc. Prof. Yusuf Yaslan

Parametric Methods

Parametric Estimation

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

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

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

Maximum Likelihood Estimation

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

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

- Log likelihood

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

- Maximum likelihood estimator (MLE)

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

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

Examples: Bernoulli (Derivation)

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

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

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

$$\frac{dL(p_0 \mid X)}{dp_0} = \sum_{t=1}^{N} x^t \frac{d}{dp_0}\log(p_0) + \sum_{t=1}^{N}(1 - x^t)\frac{d}{dp_0}\log(1 - p_0)$$

$$= \frac{1}{p_0}\sum_{t=1}^{N} x^t - \sum_{t=1}^{N}(1 - x^t)\frac{1}{1 - p_0} = 0$$

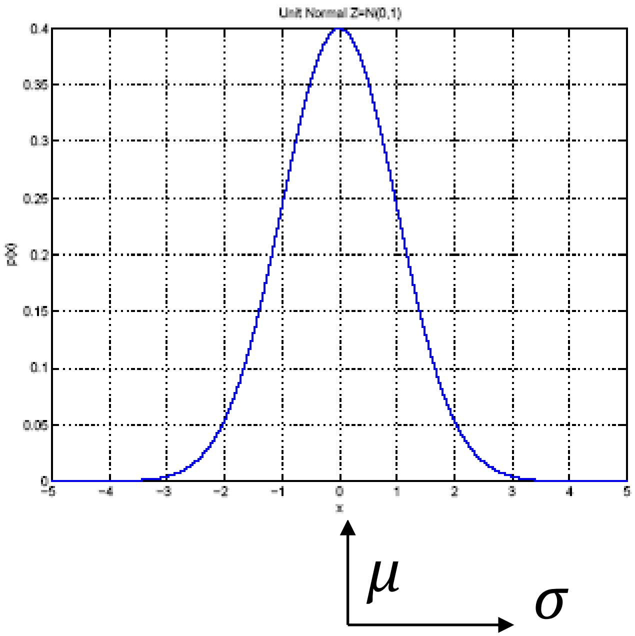
Bernoulli (Derivation)

$$= (1 - p_0)\sum_{t=1}^{N} x^t - p_0 \sum_{t=1}^{N} 1 + p_0 \sum_{t=1}^{N} x^t = 0$$

$$= \sum_{t=1}^{N} x^t - p_0 N = 0 \Rightarrow p_0 = \frac{1}{N}\sum_{t=1}^{N} x^t$$

MLE: $p_o = \Sigma_t \, x^t / N$

Gaussian (Normal) Distribution



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

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

Gaussian (Normal) Distribution

- Given that $\mathcal{X} = \{ x^t \}_t$ with $x^t \sim \mathcal{N}(\mu, \sigma^2)$

$$L(\mu, \sigma \mid X) = -\frac{N}{2}\log(2\pi) - N\log(\sigma) - \frac{\sum_{n=1}^{N}(x^t - \mu)^2}{2\sigma^2}$$

MLE for μ and σ^2:

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

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

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 $\mathcal{X} = \{x^t, r^t\}_{t=1}^{N}$

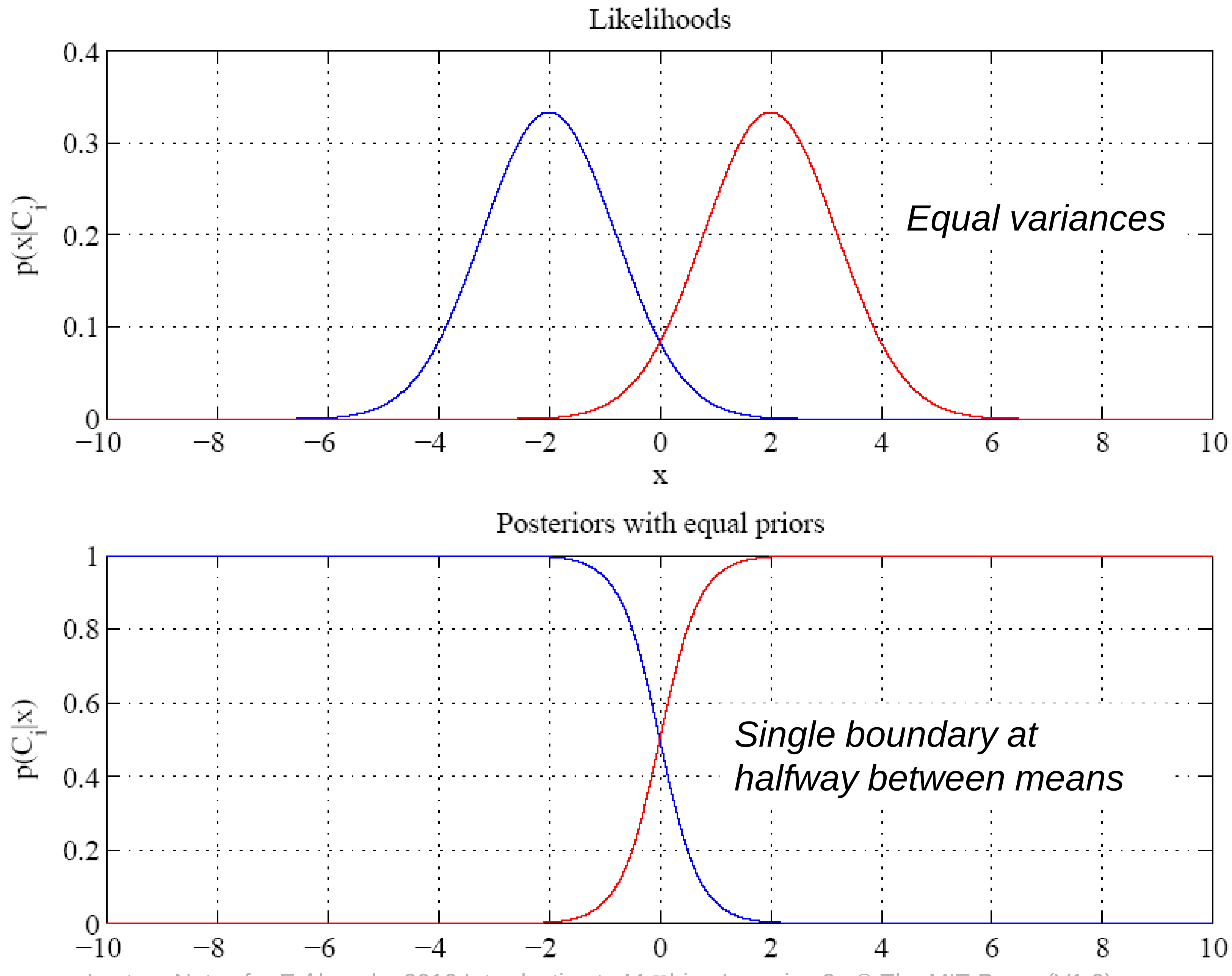
$$x \in \Re \qquad r_i^t = \begin{cases} 1 \text{ if } x^t \in C_i \\ 0 \text{ if } x^t \in C_j, j \neq i \end{cases}$$

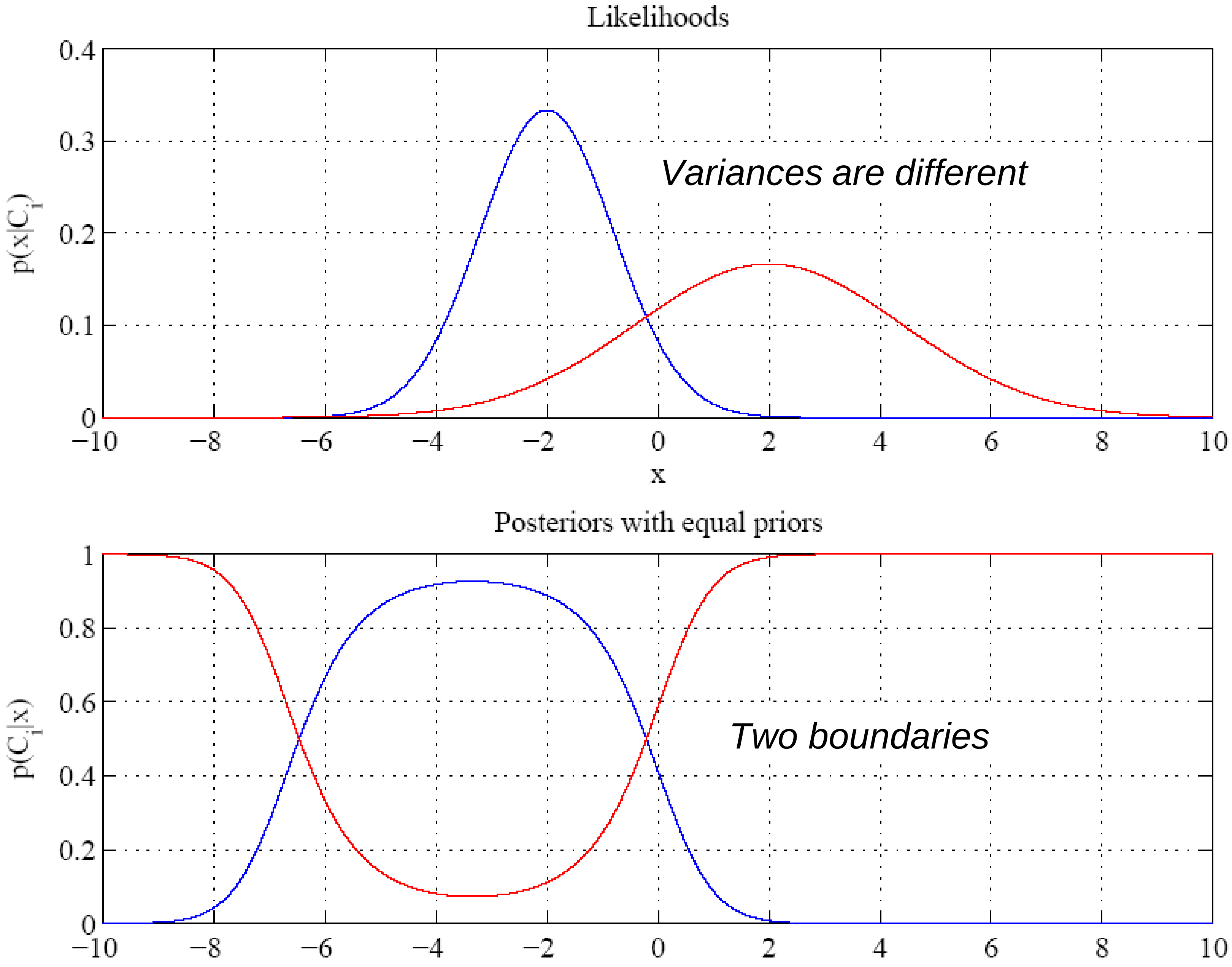
- ML estimates are

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

- 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
p(x|C_i)
x
Equal variances
Posteriors with equal priors
p(C_i|x)
Single boundary at
halfway between means
x


Likelihoods
Variances are different
p(x|C_i)
x
Posteriors with equal priors
Two boundaries
p(C_i|x)
x

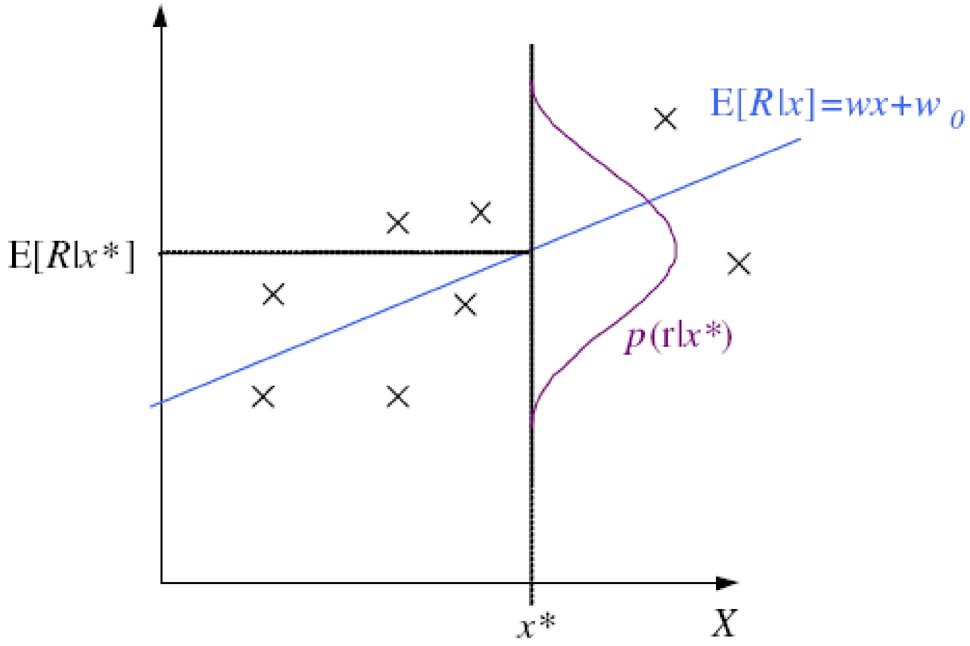
Probabilistic Interpretation of Linear Regression

$$r = f(x) + \varepsilon$$

$$\text{estimator}: g(x \mid \theta)$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$p(r \mid x) \sim \mathcal{N}(g(x \mid \theta), \sigma^2)$$



$$\mathcal{L}(\theta \mid \mathcal{X}) = \log \prod_{t=1}^{N} p(x^t, r^t)$$

$$= \log \prod_{t=1}^{N} p(r^t \mid x^t) + \log \prod_{t=1}^{N} p(x^t)$$

Regression: From LogL to Error

$$\mathcal{L}(\theta|\mathcal{X}) = \log \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left[r^t - g\left(x^t|\theta\right)\right]^2}{2\sigma^2} \right]$$

$$= -N \log\sqrt{2\pi}\sigma - \frac{1}{2\sigma^2} \sum_{t=1}^{N} \left[r^t - g\left(x^t|\theta\right)\right]^2$$

$$E(\theta|\mathcal{X}) = \frac{1}{2} \sum_{t=1}^{N} \left[r^t - g\left(x^t|\theta\right)\right]^2$$