M L
D M

Chapter 3
Parameter Estimation

Contents

MIMA

- Introduction
- Maximum-Likelihood Estimation
- Bayesian Estimation

Bayesian Theorem

MIMA

$$P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{p(\mathbf{x})}$$

$$p(\mathbf{x}) = \sum_{j=1}^{c} p(\mathbf{x} \mid \omega_i)P(\omega_i)$$

- To compute posterior probability $P(\omega_i \mid \mathbf{x})$, we need to know:

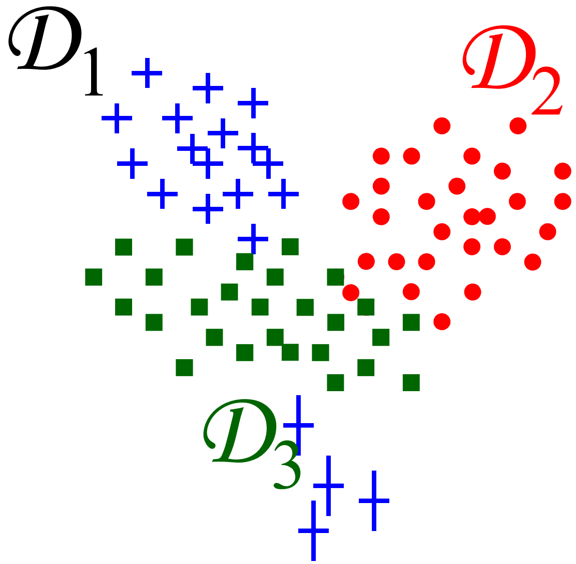
$$p(\mathbf{x} \mid \omega_i) \qquad P(\omega_i)$$

How can we get these values?

Samples

$$\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_c\}$$

The samples in D_j are drawn independently according to the probability law $p(x|\omega_j)$. That is, examples in D_j are i.i.d. random variables, i.e., independent and identically distributed.



It is easy to compute the prior probability:

$$P(\omega_i) = \frac{|D_j|}{\sum_{i=1}^{c} |D_i|}$$

Samples

MIMA

- For class-conditional pdf:
  - Case I: $p(x|\omega_j)$ has certain parametric form
    - e.g.

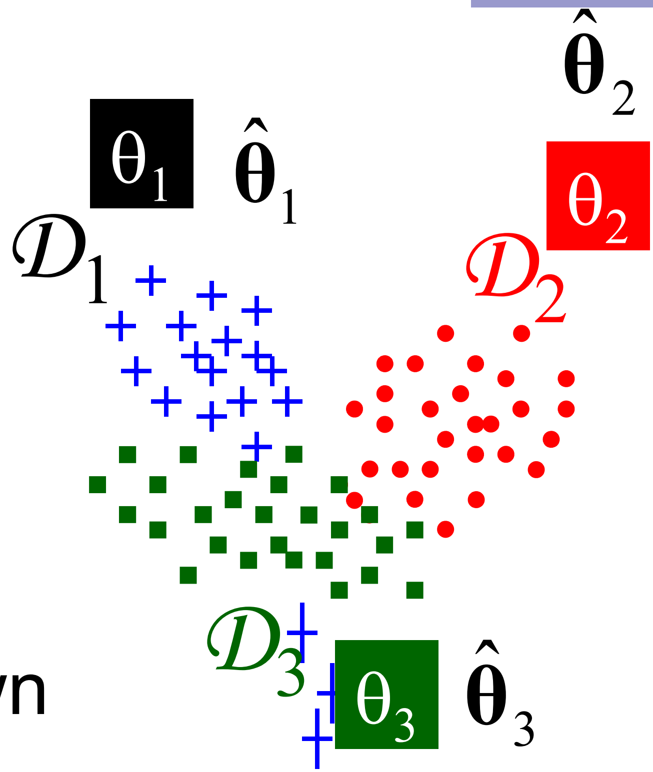
$$p(\mathbf{x}|\omega_j) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

$$\underbrace{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j}_{\theta_j} \Rightarrow \boldsymbol{\theta}_j = (\theta_1, \theta_2, \ldots, \theta_m)^T$$

    - If $X \in R^d$ θ_j contains "$d+d(d+1)/2$" free parameters.
  - Case II: $p(x|\omega_j)$ doesn't have parametric form
    - Next chapter.

Goal

$$\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_c\}$$



$$p(\mathbf{x} \mid \omega_j) \equiv p(\mathbf{x} \mid \boldsymbol{\theta}_j)$$

Use $\mathcal{D}_j$ to estimate the unknown parameter vector θ_j

$$\boldsymbol{\theta}_j = (\theta_1, \theta_2, \ldots, \theta_m)^T$$

Estimation Under Parametric Form

MIMA

- Maximum-Likelihood Estimation

View parameters as quantities whose values are fixed but unknown → Estimate parameter values by maximizing the likelihood (probability) of observing the actual examples.

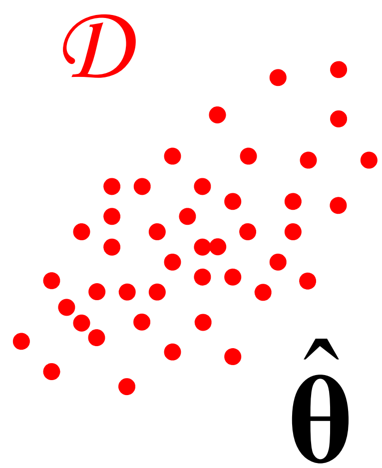
- Bayesian Estimation

View parameters as random variables having some known prior distribution → Observation of the actual training examples transforms parameters' prior into posterior distribution. (via Bayes rule)

Maximum-Likelihood Estimation

- Because each class is considered individually, the subscript used before will be dropped.
- Now the problem becomes:

Given a sample set $\mathcal{D}$, whose elements are drawn independently from a population possessing a known parameter form, say $p(x|\theta)$, we want to choose a $\hat{\theta}$ that will make $\mathcal{D}$ to occur most likely.



Maximum-Likelihood Estimation (Cont.)

- Criterion of ML

$$\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$$

- By the independence assumption, we have

$$p(\mathcal{D} \mid \boldsymbol{\theta}) = p(\mathbf{x}_1 \mid \boldsymbol{\theta})\, p(\mathbf{x}_2 \mid \boldsymbol{\theta}) \cdots p(\mathbf{x}_n \mid \boldsymbol{\theta}) = \prod_{k=1}^{n} p(\mathbf{x}_k \mid \boldsymbol{\theta})$$

- The Likelihood Function

$$L(\boldsymbol{\theta} \mid \mathcal{D}) = p(\mathcal{D} \mid \boldsymbol{\theta}) = \prod_{k=1}^{n} p(\mathbf{x}_k \mid \boldsymbol{\theta})$$

- The maximum-likelihood estimation:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} L(\theta \mid D)$$

Maximum-Likelihood Estimation (Cont.)

- Often, we resort to maximize the log-likelihood function

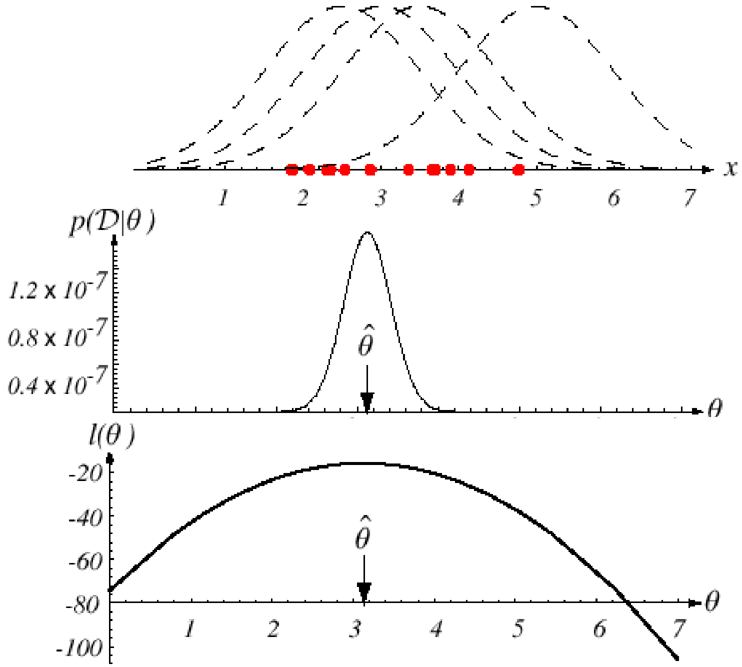
$$l(\boldsymbol{\theta} \mid \mathcal{D}) = \ln L(\boldsymbol{\theta} \mid \mathcal{D}) \quad = \sum_{k=1}^{n} \ln p(\mathbf{x}_k \mid \boldsymbol{\theta})$$

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta} \mid \mathcal{D})$$

$$\Updownarrow$$

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta} \mid \mathcal{D})$$

why?



Maximum-Likelihood Estimation (Cont.)

- Find the extreme values using the method in differential calculus.
- Gradient Operator
  - Let $f(\theta)$ be a continuous function, where $\theta=(\theta_1, \theta_2,\ldots, \theta_n)^T$.

Gradient Operator

$$\nabla_{\boldsymbol{\theta}} = \left(\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \Lambda , \frac{\partial}{\partial \theta_n} \right)^T$$

- Find the extreme values by solving

$$\nabla_{\boldsymbol{\theta}} f = 0$$

The Gaussian Case I

MIMA

- Case I: unknown μ, and Σ is known

$$p(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}\,|\,\boldsymbol{\Sigma}\,|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

$$L(\boldsymbol{\mu}\,|\,\mathcal{D}) = p(\mathcal{D}\,|\,\boldsymbol{\mu}) = \prod_{k=1}^{n} p(\mathbf{x}_k\,|\,\boldsymbol{\theta})$$

$$= \frac{1}{(2\pi)^{nd/2}\,|\,\boldsymbol{\Sigma}\,|^{n/2}} \prod_{k=1}^{n} \exp\left[-\frac{1}{2}(\mathbf{x}_k-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k-\boldsymbol{\mu})\right]$$

$$l(\boldsymbol{\mu}\,|\,\mathcal{D}) = \ln L(\boldsymbol{\mu}\,|\,\mathcal{D})$$

$$= -\ln(2\pi)^{nd/2}\,|\,\boldsymbol{\Sigma}\,|^{n/2} - \frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k-\boldsymbol{\mu})$$

The Gaussian Case I

MIMA

$$l(\boldsymbol{\mu} \mid \mathcal{D}) = \ln L(\boldsymbol{\mu} \mid \mathcal{D})$$

$$= -\ln(2\pi)^{nd/2} \mid \boldsymbol{\Sigma} \mid^{n/2} - \frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k - \boldsymbol{\mu})$$

$$\nabla_{\boldsymbol{\mu}} l(\boldsymbol{\mu} \mid \mathcal{D}) = \sum_{k=1}^{n} \boldsymbol{\Sigma}^{-1}(\mathbf{x}_k - \boldsymbol{\mu}) = 0$$

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k$$

Sample Mean!

Intuitive Result: Maximum estimate for the unknown μ is just the arithmetic average of training samples---sample mean.

The Gaussian Case II

- Case II: both μ and Σ are unknown
- Consider univariate case

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

$$\boldsymbol{\theta} = (\theta_1, \theta_2)^T = (\mu, \sigma^2)^T$$

$$L(\boldsymbol{\theta} \mid \mathcal{D}) = p(\mathcal{D} \mid \boldsymbol{\theta}) = \prod_{k=1}^{n} p(x_k \mid \boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}\sigma^n} \prod_{k=1}^{n} \exp\left[-\frac{(x_k-\mu)^2}{2\sigma^2} \right]$$

$$l(\boldsymbol{\theta} \mid \mathcal{D}) = \ln L(\boldsymbol{\theta} \mid \mathcal{D}) = -\ln(2\pi)^{n/2}\sigma^n - \frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2$$

$$= -\ln(2\pi)^{n/2}\theta_2^{\ n/2} - \frac{1}{2\theta_2} \sum_{k=1}^{n} (x_k - \theta_1)^2$$

The Gaussian Case II

MIMA

$$l(\boldsymbol{\theta} \mid \mathcal{D}) = -\ln(2\pi)^{n/2}\theta_2^{\,n/2} - \frac{1}{2\theta_2}\sum_{k=1}^{n}(x_k - \theta_1)^2$$

$$\nabla_{\boldsymbol{\theta}} l(\boldsymbol{\theta} \mid \mathcal{D}) = \begin{bmatrix} \frac{1}{\theta_2}\sum_{k=1}^{n}(x_k - \theta_1) \\ -\frac{n}{2\theta_2} + \sum_{k=1}^{n}\frac{(x_k - \theta_1)^2}{2\theta_2^2} \end{bmatrix} = \mathbf{0}$$

Unbiased Estimator:

$$E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$$

Consistent Estimator:

$$\lim_{n\to\infty} E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$$

unbiased

$$\hat{\mu} = \hat{\theta}_1 = \frac{1}{n}\sum_{k=1}^{n} x_k$$

Arithmetic average of n vectors

$$\hat{\sigma}^2 = \hat{\theta}_2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu})^2$$

biased

Arithmetic average of n matrices

$$(\mathbf{x}_k - \hat{\boldsymbol{\mu}})(\mathbf{x}_k - \hat{\boldsymbol{\mu}})^T$$

MLE for Normal Population

MIMA

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k$$

Sample Mean

$$E[\hat{\boldsymbol{\mu}}] = \boldsymbol{\mu}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \hat{\boldsymbol{\mu}})(\mathbf{x}_k - \hat{\boldsymbol{\mu}})^T$$

$$E[\hat{\boldsymbol{\Sigma}}] = \frac{n-1}{n}\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}$$

$$\mathbf{C} = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{x}_k - \hat{\boldsymbol{\mu}})(\mathbf{x}_k - \hat{\boldsymbol{\mu}})^T$$

Sample Covariance Matrix

$$E[\mathbf{C}] = \boldsymbol{\Sigma}$$

$$E(\sigma_{ML}^2) = E(\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2) = E[\frac{1}{N}\sum_{n=1}^{N}(x_n^2 - 2x_n\mu_{ML} + \mu_{ML}^2)]$$

$$= E[\frac{1}{N}\sum_{n=1}^{N}(x_n^2) - 2\mu_{ML} \bullet \frac{1}{N}\sum_{n=1}^{N}x_n + \mu_{ML}^2] = E[\frac{1}{N}\sum_{n=1}^{N}(x_n^2) - 2\mu_{ML} \bullet \mu_{ML} + \mu_{ML}^2]$$

$$= E[\frac{1}{N}\sum_{n=1}^{N}(x_n^2) - \mu_{ML}^2] = \frac{1}{N}\sum_{n=1}^{N}E(x_n^2) - E(\mu_{ML}^2)$$

$$E(x_n^2) = \sigma^2 + \mu^2$$

$$E(\mu_{ML}^2) = D(\mu_{ML}) + [E(\mu_{ML})]^2 = D(\frac{1}{N}\sum_{n=1}^{N}x_n) + [E(\mu_{ML})]^2 = \frac{1}{N^2}\sum_{n=1}^{N}D(x_n) + \mu^2$$

Bayesian Estimation

MIMA

- Settings
  - The parametric form of the likelihood function for each category is known
  - However, θ_j is considered to be random variables instead of being fixed (but unknown) values.

In this case, we can no longer make a single ML estimate $\hat{\boldsymbol{\theta}}$ *and then infer* $P(\omega_i \mid \mathbf{x})$ *based on* $P(\omega_i)$ *and* $p(\mathbf{x} \mid \omega_i)$

How can we proceed? → Fully exploit training examples!

Posterior Probabilities from sample

MIMA

$$\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_c\}$$

$$P(\omega_i \mid \mathbf{x}, \mathcal{D}) = \frac{P(\omega_i, \mathrm{x}, \mathcal{D})}{P(\mathrm{x}, \mathcal{D})} = \frac{P(\omega_i, \mathrm{x}, \mathcal{D})}{\sum_{j=1}^{c} P(\omega_j, \mathrm{x}, \mathcal{D})}$$

$$P(\omega_i, \mathbf{x}, \mathcal{D}) = P(D) \cdot P(\omega_i, \mathrm{x} \mid \mathcal{D}) = P(D) \cdot P(\omega_i \mid \mathcal{D}) \cdot P(\mathbf{x} \mid \omega_i, \mathcal{D})$$

Assumptions:

$$P(\omega_i \mid \mathcal{D}) = P(\omega_i)$$

$$P(\mathbf{x} \mid \omega_i, \mathcal{D}) = P(\mathbf{x} \mid \omega_i, \mathcal{D}_i)$$

$$P(\omega_i \mid \mathbf{x}, \mathcal{D}) = \frac{P(\mathbf{x} \mid \omega_i, \mathcal{D}_i) P(\omega_i)}{\sum_{j=1}^{c} P(\mathbf{x} \mid \omega_j, \mathcal{D}_j) P(\omega_j)}$$

Each class can be considered independently

Problem Formulation

MIMA

$$P(\omega_i \mid \mathbf{x}, \mathcal{D}) = \frac{P(\mathbf{x} \mid \omega_i, \mathcal{D}_i) P(\omega_i)}{\sum_{j=1}^{c} P(\mathbf{x} \mid \omega_j, \mathcal{D}_j) P(\omega_j)}$$

The key problem is to determine, $P(\mathbf{x} \mid \omega_i, \mathcal{D}_i)$ *,treat each class independently, the problem becomes* $P(\mathbf{x} \mid \mathcal{D})$

This is always the central problem of Bayesian Learning.

Class-Conditional Density Estimation

MIMA

Assume $p(x)$ is unknown but knowing it has a fixed form with parameter vector θ.

$$p(\mathbf{x} \mid \mathcal{D}) = \int p(\mathbf{x}, \boldsymbol{\theta} \mid D) d\boldsymbol{\theta}$$

θ :Random variable w.r.t. parametric form

$$= \int p(\mathbf{x} \mid \boldsymbol{\theta}, D)\, p(\boldsymbol{\theta} \mid \mathcal{D}) d\boldsymbol{\theta}$$

$$= \int p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathcal{D}) d\boldsymbol{\theta}$$

x is independent of D given θ

The form of distribution is assumed known

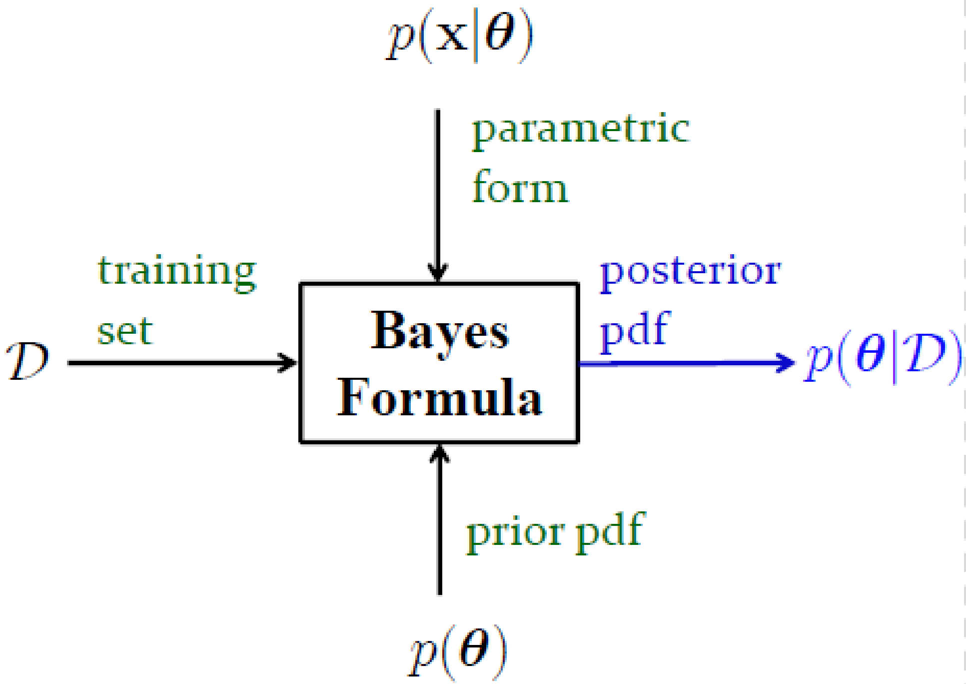
The posterior density we want to estimate

Bayesian Estimation: General Procedure

MIMA

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = ?$$

Phase I:



$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\boldsymbol{\theta}, \mathcal{D})}{p(\mathcal{D})}$$

$$= \frac{p(\boldsymbol{\theta})p(\mathcal{D}|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta}, \mathcal{D})d\boldsymbol{\theta}}$$

$$= \frac{p(\boldsymbol{\theta})p(\mathcal{D}|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})p(\mathcal{D}|\boldsymbol{\theta})d\boldsymbol{\theta}}$$

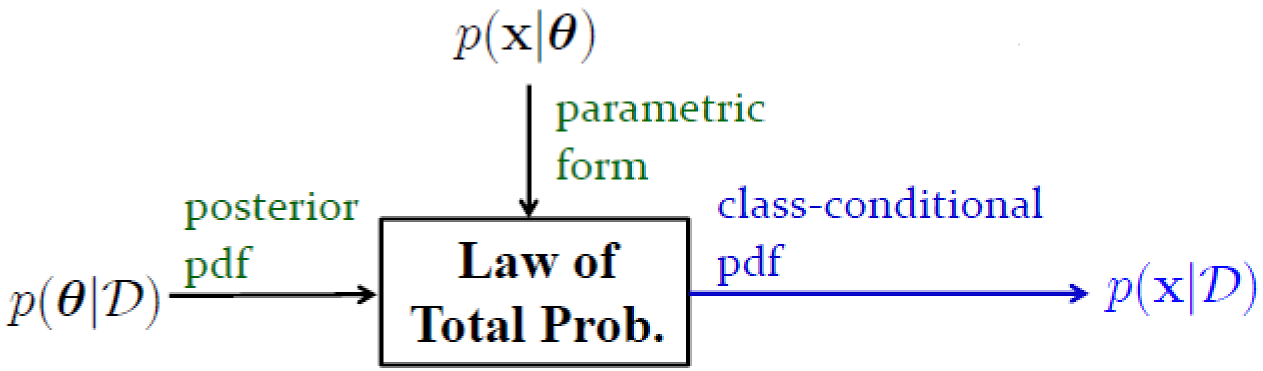
$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{n} p(\mathbf{x}_k|\boldsymbol{\theta})$$

Bayesian Estimation: General Procedure

MIMA

Phase II:

$$p(\mathbf{x} \mid \mathcal{D}) = \int p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathcal{D})\, d\boldsymbol{\theta}$$



Phase III:

$$P(\omega_i \mid \mathbf{x}, \mathcal{D}) = \frac{P(\mathbf{x} \mid \omega_i, \mathcal{D}_i) P(\omega_i)}{\sum_{j=1}^{c} P(\mathbf{x} \mid \omega_j, \mathcal{D}_j) P(\omega_j)}$$

The Gaussian Case

MIMA

- The univariate Gaussian: unknown μ

Phase I:

$$p(\theta \mid \mathcal{D}) = \alpha \prod_{k=1}^{n} p(\mathbf{x}_k \mid \theta) p(\theta)$$

$$p(\mu) + p(x \mid \mu) + D \Rightarrow p(\mu \mid D)$$

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

$$p(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2} \left(\frac{\mu-\mu_0}{\sigma_0} \right)^2 \right]$$

Other form of prior pdf could be assumed as well

The Gaussian Case

MIMA

$$p(\mu)=\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right] \qquad p(x\mid\mu)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$p(\boldsymbol{\theta}\mid\mathcal{D})=\alpha\prod_{k=1}^{n}p(\mathbf{x}_k\mid\boldsymbol{\theta})p(\boldsymbol{\theta})$$

$$p(\mu\mid\mathcal{D})=\alpha\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x_k-\mu}{\sigma}\right)^2\right]\cdot\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right]$$

$$=\alpha'\exp\left[-\frac{1}{2}\left(\sum_{k=1}^{n}\left(\frac{x_k-\mu}{\sigma}\right)^2+\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right)\right]$$

$$=\alpha''\exp\left[-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2-2\left(\frac{1}{\sigma^2}\sum_{k=1}^{n}x_k+\frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right]$$

The Gaussian Case

MIMA

$p(\mu \mid \mathcal{D})$ *is an exponential function of a quadratic function of* μ*; thus* $p(\mu \mid \mathcal{D})$ *is also a normal.*

$$p(\mu \mid \mathcal{D}) \sim N(\mu_n, \sigma_n^2)$$

$$\Rightarrow \quad p(\mu \mid \mathcal{D}) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_n}{\sigma_n}\right)^2\right]$$

$$= \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2\sigma_n^2}\left(\mu^2 - 2\mu_n\mu + \mu_n^2\right)\right]$$

$$p(\mu \mid \mathcal{D}) = \alpha'' \exp\left[-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{1}{\sigma^2}\sum_{k=1}^{n} x_k + \frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right]$$

Comparison

The Gaussian Case

MIMA

- Equating the coefficients in both form; then, we have

$$\mu_n = \left(\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \right) \hat{\mu}_n + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mu_0 \qquad \hat{\mu}_n = \frac{1}{n} \sum_{k=1}^{n} x_k$$

$$\sigma_n^2 = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2}$$

The Gaussian Case

MIMA

Phase II: $$p(\mathbf{x} \mid \mathcal{D}) = \int p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathcal{D})\, d\boldsymbol{\theta}$$

$$p(\mu \mid D) + p(x \mid \mu) \Rightarrow p(x \mid D)$$

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

$$p(\mu \mid \mathcal{D}) \sim N(\mu_n, \sigma_n^2)$$

How would $p(x \mid D)$ look like in this case?

The Gaussian Case

MIMA

$$p(\mathbf{x}\mid\mathcal{D})=\int p(\mathbf{x}\mid u)\,p(u\mid\mathcal{D})\,d\boldsymbol{\theta}$$

$$\begin{cases} p(x\mid\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right] \\ p(\mu\mid\mathcal{D})\sim N(\mu_n,\sigma_n^2) \end{cases}$$

$$p(x\mid\mathcal{D})=\frac{1}{2\pi\sigma\sigma_n}\int\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_n}{\sigma_n}\right)^2\right]d\mu$$

$$=\frac{1}{2\pi\sigma\sigma_n}\exp\left[-\frac{1}{2}\frac{(x-\mu_n)^2}{\sigma^2+\sigma_n^2}\right]\underbrace{\int\exp\left[-\frac{1}{2}\frac{\sigma^2+\sigma_n^2}{\sigma^2\sigma_n^2}\left(\mu-\frac{\sigma_n^2x+\sigma^2\mu_n}{\sigma^2+\sigma_n^2}\right)^2\right]d\mu}_{=?}$$

$p(x\mid\mathcal{D})$ is an exponential function of a quadratic function of x; thus, it is also a normal pdf.

The Gaussian Case

MIMA

$$p(\mathbf{x}|\mathcal{D}) = \int p(\mathbf{x}|u)\,p(u|\mathcal{D})\,d\boldsymbol{\theta}$$

$$\begin{cases} p(x|\mu) = \dfrac{1}{\sqrt{2\pi}\sigma}\exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right] \\ p(\mu|\mathcal{D}) \sim N(\mu_n, \sigma_n^2) \end{cases}$$

$p(x|\mathcal{D}$

$$p(x|\mathcal{D}) \sim N(\mu_n, \sigma^2 + \sigma_n^2)$$

$$= \frac{1}{2\pi\sigma\sigma_n}\exp\left[-\frac{1}{2}\frac{(x-\mu_n)^2}{\sigma^2+\sigma_n^2}\right]\int \exp\left[-\frac{1}{2}\frac{\sigma^2+\sigma_n^2}{\sigma^2\sigma_n^2}\left(\mu - \frac{\sigma_n^2 x + \sigma^2\mu_n}{\sigma^2+\sigma_n^2}\right)^2\right]d\mu$$

$p(x|\mathcal{D})$ is an exponential function of a quadratic function of x; thus, it is also a normal pdf.

$=?$

The Gaussian Case

MIMA

Phase III:

$$P(\omega_i \mid \mathbf{x}, \mathcal{D}) = \frac{P(\mathbf{x} \mid \omega_i, \mathcal{D}_i) P(\omega_i)}{\sum_{j=1}^{c} P(\mathbf{x} \mid \omega_j, \mathcal{D}_j) P(\omega_j)}$$

Summary

MIMA

- Key issue
  - Estimate prior and class-conditional pdf from training set
  - Basic assumption on training examples: i.i.d.
- Two strategies to key issue
  - Parametric form for class-conditional pdf
    - Maximum likelihood estimation
    - Bayesian estimation
  - No parametric form for class-conditional pdf

Summary

MIMA

- Maximum likelihood estimation
  - Settings: parameters as fixed but unknown values
  - The objective function: log-likelihood function
  - The gradient for the objective function should be zero
  - Gaussian
- Bayesian estimation
  - Settings: parameters as random variables
  - General procedure: I, II, III
  - Gaussian case

  Project 3.2

[Thank You !]

Any Question?