Tutorial 2: Maximum-Likelihood and Bayesian Parameter Estimation

Rui Zhao

rzhao@ee.cuhk.edu.hk

Jan. 23, 2014

Outline

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
3. Bayesian estimate: Univariate Gaussian
   - brief review
   - posteriori $p(\mu|\mathcal{D})$
   - conditional probability density $p(x|w_i, \mathcal{D}_i)$

Outline

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
3. Bayesian estimate: Univariate Gaussian
   - brief review
   - posteriori $p(\mu|\mathcal{D})$
   - conditional probability density $p(x|w_i, \mathcal{D}_i)$

Outline

1. ML estimate: $\hat{\boldsymbol{\Sigma}}$ of multivariate Gaussian
2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
3. Bayesian estimate: Univariate Gaussian
   - brief review
   - posteriori $p(\mu|\mathcal{D})$
   - conditional probability density $p(x|w_i, \mathcal{D}_i)$

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian

Multivariate Gassian Case: unknown μ and Σ

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

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^t \tag{2}$$

- $\hat{\mu}$ is the sample mean.
- $\hat{\mathbf{\Sigma}}$ is the arithmetic average of the n matrices $(\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^t$.

Multivariate normal density

$$p(\mathbf{x} \mid \mu,\ \mathbf{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^t\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right] \quad (3)$$

Draw $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ independently from $p(\mathbf{x} \mid \mu,\ \mathbf{\Sigma})$, and the joint density (likelihood) is:

$$p(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n \mid \mu, \mathbf{\Sigma}) = \quad (4)$$

$$\frac{1}{(2\pi)^{nd/2}|\mathbf{\Sigma}|^{n/2}} \exp\left[-\frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k-\mu)^t\mathbf{\Sigma}^{-1}(\mathbf{x}_k-\mu)\right] \quad (5)$$

Log-likelihood $l(\mu,\ \mathbf{\Sigma})$ is

$$l(\mu, \mathbf{\Sigma}) = -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\ln|\mathbf{\Sigma}| - \frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k-\mu)^t\mathbf{\Sigma}^{-1}(\mathbf{x}_k-\mu) \quad (6)$$

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian

Multivariate normal density

Let $A = \mathbf{\Sigma}^{-1}$

$$l(\mu, \mathbf{\Sigma}) = -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\ln|\mathbf{\Sigma}| - \frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \mu)^t \mathbf{\Sigma}^{-1}(\mathbf{x}_k - \mu) \quad (7)$$

$$l(\mu, \mathbf{\Sigma}) = -\frac{nd}{2}\ln(2\pi) + \frac{n}{2}\ln \mathbf{A} - \frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \mu)^t \mathbf{A}(\mathbf{x}_k - \mu) \quad (8)$$

$$\frac{\partial l(\mu, \mathbf{\Sigma})}{\partial \mathbf{A}} = \frac{n}{2}\mathbf{A}^{-1} - \frac{1}{2}\sum_{k=1}^{n}(\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^t = 0 \quad (9)$$

Replace $\mathbf{A}$ by $\mathbf{\Sigma}^{-1}$

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^t, \quad (\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k) \quad (10)$$

Outline

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
3. Bayesian estimate: Univariate Gaussian
   - brief review
   - posteriori $p(\mu|\mathcal{D})$
   - conditional probability density $p(x|w_i, \mathcal{D}_i)$

Outline

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
3. Bayesian estimate: Univariate Gaussian
   - brief review
   - posteriori $p(\mu|\mathcal{D})$
   - conditional probability density $p(x|w_i, \mathcal{D}_i)$

2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian

Univariate Gassian Case

ML estimator $\hat{\sigma}^2$ is biased.

$$E[\hat{\sigma}^2] = E[\frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu})^2] = \frac{n-1}{n}\sigma^2 \tag{11}$$

An elementary unbiased estimator for σ^2 is given by $\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \hat{\mu})$.

2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian

Univariate Gassian Case

$$
\begin{aligned}
E[\frac{1}{n}\sum_{k=1}^{n}(x_k-\hat{\mu})^2] &= E\Big[\frac{1}{n}\sum_{k=1}^{n}(x_k-\frac{1}{n}\sum_{j=1}^{n}x_j)^2\Big] &(12)\\
&= E\Big[\frac{1}{n}\sum_{k=1}^{n}(x_k^2-\frac{2}{n}x_k\sum_{j=1}^{n}x_j+\frac{1}{n^2}(\sum_{j=1}^{n}x_j)^2)\Big] &(13)\\
&= E\Big[\frac{1}{n}\Big(\sum_{k=1}^{n}x_k^2-\frac{2}{n}(\sum_{k=1}^{n}x_k)^2+\frac{n}{n^2}(\sum_{k=1}^{n}x_k)^2\Big)\Big] &(14)\\
&= E\Big[\frac{1}{n}\Big(\sum_{k=1}^{n}x_k^2-\frac{1}{n}(\sum_{k=1}^{n}x_k)^2\Big)\Big] &(15)\\
&= \frac{1}{n}E[\sum_{k=1}^{n}x_k^2]-\frac{1}{n^2}E[(\sum_{k=1}^{n}x_k)^2] &(16)\\
&= E[x^2]-\frac{1}{n^2}E\Big[\sum_{k=1}^{n}x_k^2+\sum_{i\neq j}x_ix_j\Big] &(17)\\
&= E[x^2]-\frac{1}{n^2}E[\sum_{k=1}^{n}x_k^2]-\frac{1}{n^2}E[\sum_{i\neq j}x_ix_j] &(18)\\
&= E[x^2]-\frac{1}{n}E[x^2]-\frac{n^2-n}{n^2}E[x_ix_j]=\frac{n-1}{n}\Big(E[x^2]-(E[x])^2\Big) &(19)\\
&= \frac{n-1}{n}\sigma^2 &(20)
\end{aligned}
$$

Outline

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
3. Bayesian estimate: Univariate Gaussian
   - brief review
   - posteriori $p(\mu|\mathcal{D})$
   - conditional probability density $p(x|w_i, \mathcal{D}_i)$

Outline

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
3. Bayesian estimate: Univariate Gaussian
   - brief review
   - posteriori $p(\mu|\mathcal{D})$
   - conditional probability density $p(x|w_i, \mathcal{D}_i)$

Outline

1. ML estimate: $\hat{\Sigma}$ of multivariate Gaussian
2. ML estimate bias: $\hat{\sigma}^2$ of univariate Gaussian
3. Bayesian estimate: Univariate Gaussian
   - brief review
   - posteriori $p(\mu|\mathcal{D})$
   - conditional probability density $p(x|w_i, \mathcal{D}_i)$

3.1 Bayesian estimate: brief review

Bayesian estimate

Given sample set $\mathcal{D}$, then posteriori for estimation is

$$P(w_i|\mathbf{x}, \mathcal{D}) = \frac{p(\mathbf{x}|w_i, \mathcal{D})P(w_i|\mathcal{D})}{\sum_{j=1}^{c} p(\mathbf{x}|w_i, \mathcal{D})P(w_i|\mathcal{D})} \tag{21}$$

- consider each class individually: $p(\mathbf{x}|w_i, \mathcal{D}) \to p(\mathbf{x}|\mathcal{D})$
- prior is known $P(w_i|\mathcal{D})$

Target: estimate posteriori $p(\mathbf{x}|\mathcal{D}) \to p(\mathbf{x})$

$$p(\mathbf{x}|\mathcal{D}) = \int p(\mathbf{x}, \theta|\mathcal{D})d\theta \tag{22}$$

$$= \int p(\mathbf{x}|\theta, \mathcal{D})p(\theta|\mathcal{D})d\theta = \int p(\mathbf{x}|\theta)p(\theta|\mathcal{D})d\theta \tag{23}$$

3.2 Bayesian estimate of univariate Gaussian: posteriori $p(\mu|\mathcal{D})$

Bayesian estimate

$P(w_i|\mathbf{x},\mathcal{D}) \Rightarrow p(\mathbf{x}|w_i,\mathcal{D}) \Rightarrow p(\mathbf{x}|\mathcal{D}) \Rightarrow p(\mathbf{x}|\theta)\ \&\ p(\theta|\mathcal{D})$

- $p(\mathbf{x}|\theta)$ is pre-assumed in form: $p(x|\mu) \sim \mathcal{N}(\mu,\sigma^2)$
- $p(\theta|\mathcal{D})$ is the posteriori: $p(\mu|\mathcal{D})$

Posteriori $p(\mu|\mathcal{D})$

$$p(\mu|\mathcal{D}) = \frac{p(\mathcal{D}|\mu)p(\mu)}{\int p(\mathcal{D}|\mu)p(\mu)d\mu} \tag{24}$$

$$= \alpha \prod_{k=1}^{n} p(x_k|\mu)p(\mu) \tag{25}$$

where $p(x_k|\mu) \sim \mathcal{N}(\mu,\sigma^2)$, and $p(\mu) \sim \mathcal{N}(\mu_0,\sigma_0^2)$.

3.2 Bayesian estimate of univariate Gaussian: posteriori $p(\mu|\mathcal{D})$

Posteriori $p(\mu|\mathcal{D})$

$$p(\mu|\mathcal{D}) = \alpha \prod_{k=1}^{n} p(x_k|\mu)p(\mu) \tag{26}$$

$$= \alpha \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big[-\frac{1}{2}\Big(\frac{x_k-\mu}{\sigma}\Big)^2\Big] \frac{1}{\sqrt{2\pi}\sigma_0} \exp\Big[-\frac{1}{2}\Big(\frac{\mu-\mu_0}{\sigma_0}\Big)^2\Big] \tag{27}$$

$$= \alpha' \exp\Big[-\frac{1}{2}\Big(\sum_{k=1}^{n}(\frac{\mu-x_k}{\sigma})^2 + (\frac{\mu-\mu_0}{\sigma_0})^2\Big)\Big] \tag{28}$$

$$= \alpha'' \exp\Big[-\frac{1}{2}\Big[(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})\mu^2 - 2(\frac{1}{\sigma^2}\sum_{k=1}^{n} x_k + \frac{\mu_0}{\sigma_0^2})\mu\Big]\Big] \tag{29}$$

$$\therefore \quad p(\mu|\mathcal{D}) \sim \mathcal{N}(\mu_n, \sigma_n^2), \; i.e. \; p(\mu|\mathcal{D}) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\Big[-\frac{1}{2}\Big(\frac{\mu-\mu_n}{\sigma_n}\Big)^2\Big] \tag{30}$$

$$\Longrightarrow \frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}, \; \frac{\mu_n}{\sigma_n^2} = \frac{1}{\sigma^2}\sum_{k=1}^{n} x_k + \frac{\mu_0}{\sigma_0^2} = \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \tag{31}$$

$$\Longrightarrow \mu_n = \Big(\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\Big)\hat{\mu}_n + \frac{\sigma^2}{\sigma_0^2+\sigma^2}\mu_0, \; \sigma_0^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2} \tag{32}$$

3.3 Bayesian estimate of univariate Gaussian: conditional probability density $p(x|w_i, \mathcal{D}_i)$

Conditional probability density $p(x|w_i, \mathcal{D}_i)$

test