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: $\hat{\Sigma}$ 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

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

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

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

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

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

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

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

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 $\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \hat{\mu})$.

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

Univariate Gassian Case

$$E[\frac{1}{n}\sum_{k=1}^{n}(x_k-\hat{\mu})^2] = E\Big[\frac{1}{n}\sum_{k=1}^{n}(x_k-\frac{1}{n}\sum_{j=1}^{n}x_j)^2\Big] \tag{12}$$

$$= E\Big[\frac{1}{n}\sum_{k=1}^{n}(x_k^2-\frac{2}{n}x_k\sum_{j=1}^{n}x_j+\frac{1}{n^2}(\sum_{j=1}^{n}x_j)^2)\Big] \tag{13}$$

$$= E\Big[\frac{1}{n}\Big(\sum_{k=1}^{n}x_k^2-\frac{2}{n}(\sum_{k=1}^{n}x_k)^2+\frac{n}{n^2}(\sum_{k=1}^{n}x_k)^2\Big)\Big] \tag{14}$$

$$= E\Big[\frac{1}{n}\Big(\sum_{k=1}^{n}x_k^2-\frac{1}{n}(\sum_{k=1}^{n}x_k)^2\Big)\Big] \tag{15}$$

$$= \frac{1}{n}E[\sum_{k=1}^{n}x_k^2]-\frac{1}{n^2}E[(\sum_{k=1}^{n}x_k)^2] \tag{16}$$

$$= E[x^2]-\frac{1}{n^2}E\Big[\sum_{k=1}^{n}x_k^2+\sum_{i\neq j}x_ix_j\Big] \tag{17}$$

$$= E[x^2]-\frac{1}{n^2}E[\sum_{k=1}^{n}x_k^2]-\frac{1}{n^2}E[\sum_{i\neq j}x_ix_j] \tag{18}$$

$$= E[x^2]-\frac{1}{n}E[x^2]-\frac{n^2-n}{n^2}E[x_ix_j] = \frac{n-1}{n}\Big(E[x^2]-(E[x])^2\Big) \tag{19}$$

$$= \frac{n-1}{n}\sigma^2 \tag{20}$$

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: $\hat{\mathbf{\Sigma}}$ 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: $\hat{\mathbf{\Sigma}}$ 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 $p(x|w_i, \mathcal{D}_i)$