Robotics

Estimation and Learning
with Dan Lee

Supplementary Notes on MLE for Multivariate Gaussian

Penn Engineering

Maximum Likelihood Estimate of Multivariate Gaussian Parameters

- Objective

  Estimate the mean and the variance given observed data

$$\hat{\boldsymbol{\mu}}, \hat{\Sigma} = \arg\max_{\boldsymbol{\mu},\Sigma} p(\{\mathbf{x}_i\}|\boldsymbol{\mu}, \Sigma)$$

- Assuming independence of observations,

$$\hat{\boldsymbol{\mu}}, \hat{\Sigma} = \arg\max_{\boldsymbol{\mu},\Sigma} \prod_{i=1}^{N} p(\mathbf{x}_i|\boldsymbol{\mu}, \Sigma)$$

Maximum Likelihood Estimate of Multivariate Gaussian Parameters

- To obtain $\hat{\boldsymbol{\mu}}, \hat{\Sigma} = \arg\max_{\boldsymbol{\mu},\Sigma} \prod_{i=1}^{N} p(\mathbf{x}_i | \boldsymbol{\mu}, \Sigma)$

(1) $\arg\max_{\boldsymbol{\mu},\Sigma} \prod_{i=1}^{N} p(\mathbf{x}_i | \boldsymbol{\mu}, \Sigma) = \arg\max_{\boldsymbol{\mu},\Sigma} \sum_{i=1}^{N} \ln p(\mathbf{x}_i | \boldsymbol{\mu}, \Sigma)$

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

Maximum Likelihood Estimate of Multivariate Gaussian Parameters

$$\hat{\boldsymbol{\mu}}, \hat{\Sigma} = \arg\max_{\boldsymbol{\mu},\Sigma} \prod_{i=1}^{N} p(\mathbf{x}_i | \boldsymbol{\mu}, \Sigma)$$

$$= \arg\max_{\mu,\sigma} \sum_{i=1}^{N} \ln p(\mathbf{x}_i | \boldsymbol{\mu}, \Sigma)$$ from (1)

$$= \arg\max_{\mu,\sigma} \sum_{i=1}^{N} \left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^{\boldsymbol{T}} \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) - \frac{1}{2} \ln|\Sigma| + c \right\}$$ from (2)

$$= \arg\min_{\mu,\sigma} \sum_{i=1}^{N} \left\{ \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^{\boldsymbol{T}} \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) + \frac{1}{2} \ln|\Sigma| \right\}$$

Maximum Likelihood Estimate of Multivariate Gaussian Parameters

Let $J(\boldsymbol{\mu}, \Sigma) = \sum_{i=1}^{N} \left\{ \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) + \frac{1}{2} \ln|\Sigma| \right\}$

Then $\hat{\boldsymbol{\mu}}, \hat{\Sigma} = \arg\min_{\boldsymbol{\mu}, \Sigma} J(\boldsymbol{\mu}, \Sigma)$

① $\frac{\partial J}{\partial \boldsymbol{\mu}} = \mathbf{0} \longrightarrow \hat{\boldsymbol{\mu}}$ ② $\frac{\partial J(\hat{\boldsymbol{\mu}}, \Sigma)}{\partial \Sigma} = 0 \longrightarrow \hat{\Sigma}$

Maximum Likelihood Estimate of Multivariate Gaussian Parameters

① $$\boxed{\frac{\partial J}{\partial \boldsymbol{\mu}} = \mathbf{0} \longrightarrow \hat{\boldsymbol{\mu}}}$$

$$\frac{\partial J}{\partial \boldsymbol{\mu}} = \frac{\partial}{\partial \boldsymbol{\mu}} \sum_{i=1}^{N} \left\{ \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^{T} \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) + \frac{1}{2} |\Sigma| \right\}$$

$$= \frac{\partial}{\partial \boldsymbol{\mu}} \sum_{i=1}^{N} \left\{ \frac{1}{2} \boldsymbol{\mu}^{T} \Sigma^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^{T} \Sigma^{-1} \mathbf{x}_i \right\}$$

$$= \Sigma^{-1} \sum_{i=1}^{N} \{ \boldsymbol{\mu} - \mathbf{x}_i \} = \mathbf{0} \longrightarrow \hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$$

Maximum Likelihood Estimate of Multivariate Gaussian Parameters

② $$\boxed{\frac{\partial J(\hat{\boldsymbol{\mu}}, \Sigma)}{\partial \Sigma} = 0 \longrightarrow \hat{\Sigma}}$$

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^\intercal$$

$$\frac{\partial J}{\partial \Sigma} = \frac{\partial}{\partial \Sigma}\sum_{i=1}^{N}\left\{\frac{1}{2}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^{T}\Sigma^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}) + \frac{1}{2}\ln|\Sigma|\right\}$$

$$= \frac{1}{2}\sum_{i=1}^{N}\{-\Sigma^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^{T}\Sigma^{-1} + \Sigma^{-1}\}$$

$$= \frac{1}{2}\Sigma^{-1}\left[-\left\{\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^{T}\right\}\Sigma^{-1} + N \cdot \mathrm{I}\right] = 0$$

Maximum Likelihood Estimate of Multivariate Gaussian Parameters

- In summary, we have

$$\widehat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$$

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}})(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})^\intercal$$