MLE for Multivariate Gaussian

Piyush Rai

Introduction to Machine Learning (CS771A)

(Supplementary Slides)

Gaussian Distribution

- The (multivariate) Gaussian with mean $\boldsymbol{\mu}$ and cov. matrix $\boldsymbol{\Sigma}$

$$\begin{aligned}\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}\\ &= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}}\exp\left\{-\frac{1}{2}\text{trace}\left[\boldsymbol{\Sigma}^{-1}\mathbf{S}\right]\right\} \qquad \text{where } \mathbf{S}=(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top\end{aligned}$$

- An alternate representation: The "information form"

$$\mathcal{N}_c(\boldsymbol{x}|\boldsymbol{\xi},\boldsymbol{\Lambda}) = (2\pi)^{-D/2}|\boldsymbol{\Lambda}|^{1/2}\exp\left\{-\frac{1}{2}\left(\boldsymbol{x}^\top\boldsymbol{\Lambda}\boldsymbol{x}+\boldsymbol{\xi}^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{\xi}-2\boldsymbol{x}^\top\boldsymbol{\xi}\right)\right\}$$

  where $\boldsymbol{\Lambda}=\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\xi}=\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ are the "natural parameters" (recall exp. family).

- Note that there is a term quadratic in $\boldsymbol{x}$ (involves $\boldsymbol{\Lambda}=\boldsymbol{\Sigma}^{-1}$) and linear in $\boldsymbol{x}$ (involves $\boldsymbol{\xi}=\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$)
- Information form can help recognize $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of a Gaussian when doing algebraic manipulations

Estimating Parameters of Gaussian: MLE

- Given: N i.i.d. observations $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ from a multivariate Gaussian $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Goal: Estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Simple to do MLE for this task

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \log p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \log p(\boldsymbol{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \log \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- Plugging in $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} \exp\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\}$ and ignoring the constants

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{N}{2} \log |\boldsymbol{\Sigma}|^{-1} - \frac{1}{2} \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}) \\
&= \frac{N}{2} \log |\boldsymbol{\Sigma}|^{-1} - \frac{1}{2} \sum_{n=1}^{N} \text{trace}[\boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^\top] \\
&= \frac{N}{2} \log |\boldsymbol{\Sigma}|^{-1} - \frac{1}{2} \text{trace}[\boldsymbol{\Sigma}^{-1} \mathbf{S}_{\mu}] \qquad \left[\text{where} \quad \mathbf{S}_{\mu} = \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^\top\right]
\end{aligned}$$

Estimating Parameters of Gaussian: MLE

- Taking (partial) derivatives w.r.t. $\boldsymbol{\mu}$ and setting to zero

$$\frac{\partial}{\partial \boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\partial}{\partial \boldsymbol{\mu}} \left[\frac{N}{2} \log |\boldsymbol{\Sigma}|^{-1} - \frac{1}{2} \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}) \right] = -\frac{1}{2} \sum_{n=1}^{N} (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-\top})(\boldsymbol{x}_n - \boldsymbol{\mu}) = 0$$

which gives the following MLE solution for the multivariate Gaussian's mean

$$\boxed{\boldsymbol{\mu}_{ML} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n}$$

- Taking derivatives w.r.t. $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ (instead of $\boldsymbol{\Sigma}$; leads to simpler derivatives) and setting to zero

$$\frac{\partial}{\partial \boldsymbol{\Lambda}} \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \frac{\partial}{\partial \boldsymbol{\Lambda}} \left[\frac{N}{2} \log |\boldsymbol{\Lambda}| - \frac{1}{2} \text{trace}[\boldsymbol{\Lambda} \mathbf{S}_\mu] \right] = \frac{N}{2} \boldsymbol{\Lambda}^{-\top} - \frac{1}{2} \mathbf{S}_\mu^\top = \frac{N}{2} \boldsymbol{\Lambda}^{-1} - \frac{1}{2} \mathbf{S}_\mu = \frac{N}{2} \boldsymbol{\Sigma} - \frac{1}{2} \mathbf{S}_\mu = 0$$

which gives the following MLE solution for the multivariate Gaussian's covariance matrix

$$\boxed{\boldsymbol{\Sigma}_{ML} = \frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu}_{ML})(\boldsymbol{x}_n - \boldsymbol{\mu}_{ML})^\top}$$