假设数据集为 $\mathcal{D}=\{x_i\}_{i=1}^n$，其中单样本 $x_i\in\mathbb{R}^d, d>1$ 且服从高维高斯分布:

$$p(x_i;\mu,\Sigma)=(2\pi)^{-\frac{d}{2}}det(\Sigma)^{-\frac{1}{2}}exp\left\{-\frac{1}{2}(x_i-\mu)^\top\Sigma^{-1}(x_i-\mu)\right\}$$

试利用MLE求解 μ,Σ .

注：可能会用到的公式

$$\frac{\partial\, log\, det(\Sigma)}{\partial\Sigma}=(\Sigma^{-1})^\top$$

$$\frac{\partial\,(x^\top A^{-1}x)}{\partial A}=-(A^{-1})^\top xx^\top(A^{-1})^\top$$

更多公式可参考这里

Ans:

在整个数据集上，有

$$\begin{aligned}\log P(D;\mu,\Sigma)&=\log\left(\prod_{i=1}^n p(x_i;\mu,\Sigma)\right)\\&=\sum_{i=1}^n\log p(x_i;\mu,\Sigma)\\&=\sum_{i=1}^n-\frac{1}{2}(x_i-\mu)^\top\Sigma^{-1}(x_i-\mu)-\frac{1}{2}\log det(\Sigma)+const\end{aligned}$$

（1）关于μ求导，并令其为0：

$$\frac{\partial\log P(D;\mu,\Sigma)}{\partial\mu}=\sum_{i=1}^n\left[\Sigma^{-1}(x_i-\mu)\right]\doteq 0$$

$$\Longrightarrow\mu=\frac{1}{n}\sum_{i=1}^n x_i$$

（2）关于Σ求导，并令其为0：

$$\frac{\partial\log P(D;\mu,\Sigma)}{\partial\Sigma}=\sum_{i=1}^n\left[\frac{1}{2}\Sigma^{-1}(x_i-\mu)(x_i-\mu)^\top\Sigma^{-1}-\frac{1}{2}\Sigma^{-1}\right]\doteq 0$$

$$\Longrightarrow\Sigma=\frac{1}{n}\sum_{i=1}^n(x_i-\mu)(x_i-\mu)^\top$$