TDA231 - Algorithms for Machine Learning & Inference

Chalmers University of Technology

LP3– 2017-01-31

By:
Karabo Ikaneng, 891024-8239, kara.ikaneng@gmail.com
Elias Svensson, 920406-3052, elias.svensson.1992@gmail.com

Goal

Maximum likelihood estimation (MLE), Maximum a posteriori (MAP)

1 Theoretical problems

1.1 Maximum likelihood estimator (MLE), 4 points

Information:

$$Dataset: x_1, ..., x_n$$

$$X \sim \mathcal{N}(\mu, \sigma^2 I)$$

$$\mu \in \Re^p$$

I identity matrix p x p

σ^2scalar

Quesiton:

$$\hat{\sigma}_{MLE} =?$$

Formulas:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2} * |\Sigma|^{1/2}} * e^{\frac{-((x-\mu)^T * \Sigma^{-1} * (x-\mu))}{2}} \tag{1}$$

$$L(\sigma) = p(x|\mu, \sigma^2 I) = \prod_{i=1}^{n} \sim N(x_i|\mu, \sigma^2 I) \tag{2}$$

$$ln(L(\sigma)) = -\frac{1}{2\sigma^2}\Sigma_{i=1}^{n}(x_i - \mu)^2 - \frac{n}{2}ln(\sigma^2) - \frac{n}{2}ln(2\pi) \tag{3}$$

$$\frac{\partial l(\sigma)}{\partial \sigma} = 0 \tag{4}$$

Solution:
Given that $\Sigma = \sigma^2$ I, it means that σ is the same in every dimension. Therefore, we can use X $\sim N(x|\mu, \sigma^2)$ univariate to find σ for all features, as seen in equation 1. The likelihood function of this distribution is seen in equation 2. From there, the log-likelihood can be derived, as seen in equation 3. The MLE of σ is given by $\frac{\partial l(\sigma)}{\partial \sigma} = 0$, as seen in equation 4.

$$\frac{\partial l(\sigma)}{\partial \sigma} = 0$$

$$\Rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{N} \sum_{n=1}^{N} ||x_i - \mu||$$

$$\hat{\sigma}_{MLE} = \sqrt{\hat{\sigma}^2_{MLE}}$$

1.2 Posterior distributions, 6 points

Information:

a)

$$Dataset : x_1, ..., x_n$$

$$X \sim \mathcal{N}(\mu, \sigma^2 I)$$

$$\mu = [\mu_1, \mu_2]^T \in \Re^p$$

I identity matrix 2 x 2

σ^2scalar

$$\mathcal{P}(X = x|\mu^2) = \frac{1}{2\pi\sigma^2} * e^{\dfrac{-((x-\mu)^T * (x-\mu))}{2\sigma^2}}$$

$$\mathcal{P}(\sigma^2 = s|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} * s^{-\alpha-1} * e^{-\dfrac{\beta}{s}}$$

b)

$\alpha_A and \beta_A$(Model M_A)

$\alpha_B and \beta_B$(Model M_B)

c)

$$P(M_A) = P(M_B) = \frac{1}{2}$$

Use the MAP estimate for σ^2

Question:

a) $P(\sigma^2 = s|x_i, ..., x_n; \alpha, \beta) =?$

b) $BF_{a,b} =?$

c) $BF_{a,b} =?$

Formulas:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

$$\frac{P(D|M_1)}{P(D|M_2)} = \frac{\int_P (\theta_1|M_1)P(D|\theta_1, M_1)\partial\theta_1}{\int_P (\theta_2|M_2)P(D|\theta_2, M_2)\partial\theta_2} \tag{2}$$

Solution:

a) Multiplying the probability distribution with the inverse-gamma prior distribution produces the posterior distribution. $P(\sigma^2 = s|x_1, ..., x_n; \alpha, \beta) \propto InverseGamma(\alpha + 1, \beta + \frac{1}{2}\Sigma(x - \mu)^T(x - \mu))$

b) $BF_{a,b} = \frac{P(D|M_a)}{P(D|M_b)} = \frac{\int_{\sigma^2} P(\sigma^2|M_a)P(D|\sigma^2, M_a)\partial\sigma^2}{\int_{\sigma^2} P(\sigma^2|M_2)P(D|\sigma^2, M_2)\partial\sigma^2}$

c) $BF_{a,b} = \frac{P(M_A|\sigma^2 = \sigma^2_{MLE})}{P(M_B|\sigma^2 = \sigma^2_{MLE})}$

2 Practical problems

2.1 Spherical Gaussian estimation, 5 points

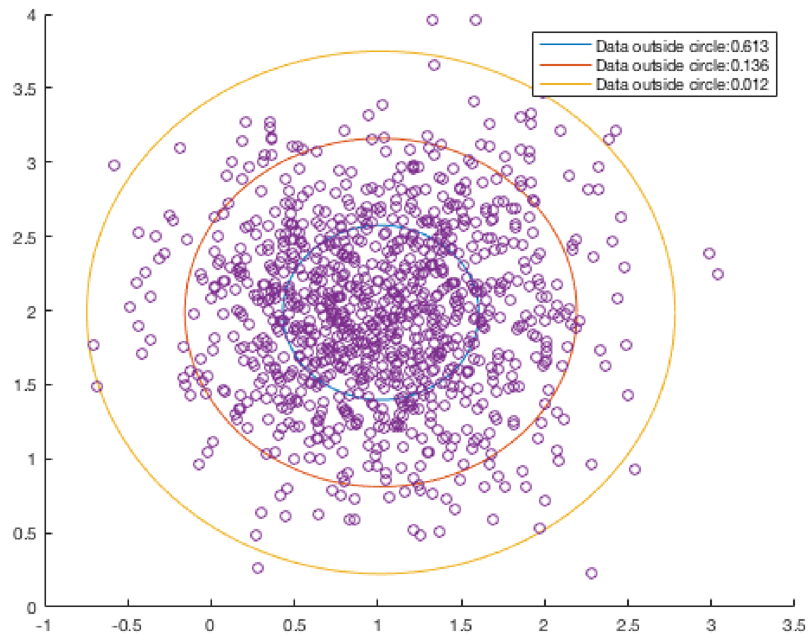

Figure 1: Scatter plot

The plot in Figure. 1 shows a scatter plot of data, with the legend displaying the fraction of data lying outside each circle. The circles correspond with standard deviation 1, 2, 3.

2.2 MAP estimation, 5 points

a)

The plot in Figure.2 and 3 respectively show the prior and posterior distribution for two set values. The posterior distribution is clearly more precise in determining σ^2.

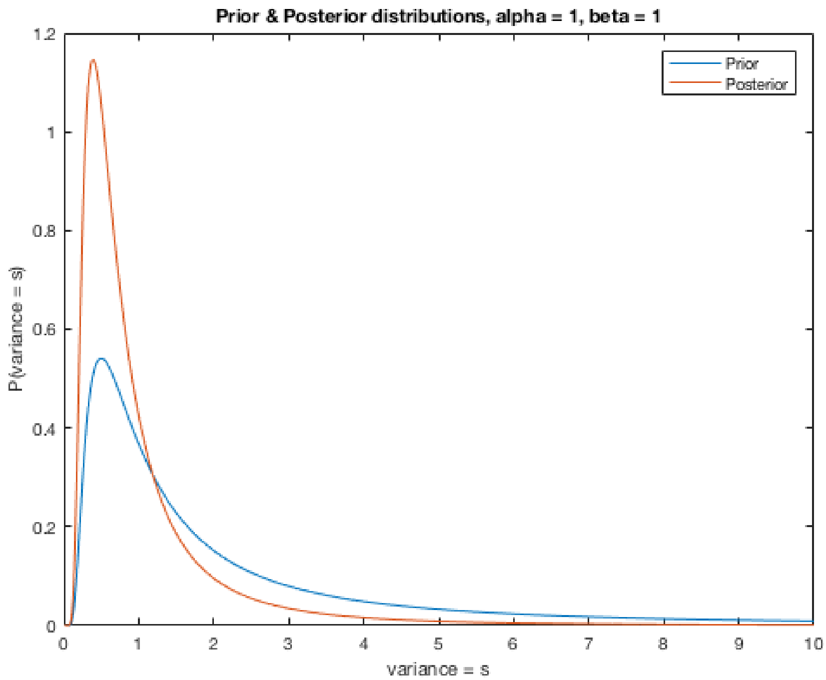

Figure 2: Prior and posterior distribution, alpha = 1, beta = 1

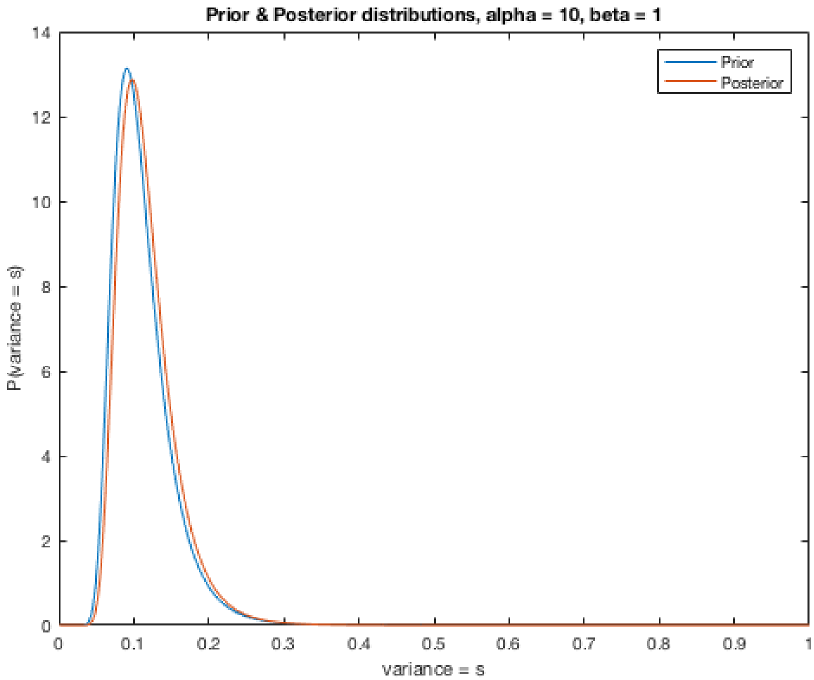

Figure 3: Prior and posterior distribution, alpha = 10, beta = 1

b)

$\sigma^2_{MAP,a} = 0.3547$

$\sigma^2_{MAP,b} = 0.3516$

c)

$BF_{a,b} = 0.8245$

As the result is $BF < 1$, Model M_b is the better model.