Homework 1: MLE & MAP

TDA231 - Algorithms for Machine Learning & Inference

Theodor Åstrand, theast@student.chalmers.se, 931109-9114
Linnéa Otterlind, linott@student.chalmers.se, 921126-3620

Chalmers University of Technology

January, 2016

Problem 1.1

A vector-valued random variable X is said to have a multivariate Gaussian distribution if its probability density function is given by:

$P(\mathbf{X}=\mathbf{x}|\mu,\Sigma)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\big(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\big)$

The likelihood function is given by:

$L(\mu,\Sigma|\mathbf{x}_1,...,\mathbf{x}_n)=\prod_{i=1}^{n}P(\mathbf{X}=\mathbf{x}|\mu,\Sigma)=$

$=\prod_{i=1}^{n}\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\big(-\frac{1}{2}(\mathbf{x}_i-\mu)^T\Sigma^{-1}(\mathbf{x}_i-\mu)\big)=$

$=\frac{1}{(2\pi)^{pn/2}|\Sigma|^{n/2}}\exp\big(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)^T\Sigma^{-1}(\mathbf{x}_i-\mu)\big)$

The log likelihood function is given by:

$l(\mu,\Sigma|\mathbf{x}_1,...,\mathbf{x}_n)=\ln(L(\mu,\Sigma|\mathbf{x}_1,...,\mathbf{x}_n))=$

$=\ln\Big(\frac{1}{(2\pi)^{pn/2}|\Sigma|^{n/2}}\exp\big(-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)^T\Sigma^{-1}(\mathbf{x}_i-\mu)\big)\Big)=$

$=-\frac{pn}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^{2p})-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)^T\frac{1}{\sigma^2}I(\mathbf{x}_i-\mu)=$

$=-\frac{pn}{2}\ln(2\pi)-np\ln(\sigma)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)$

To find the MLE for σ^2 we differentiate the log likelihood function with respect to σ and find the point where it is equal to zero.

$\frac{d}{d\sigma}l(\mu,\Sigma|\mathbf{x}_1,...,\mathbf{x}_n)=\frac{d}{d\sigma}\big(-\frac{pn}{2}\ln(2\pi)-np\ln(\sigma)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)\big)=$

$=-\frac{np}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)$

$-\frac{np}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)=0\Leftrightarrow$

$\Leftrightarrow\hat{\sigma}^2=\frac{1}{np}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)$

Problem 1.2

(a)

We have the probability density function:

$P(\mathbf{X}=\mathbf{x}|\sigma^2)=\frac{1}{2\pi\sigma^2}\exp\big(-\frac{(\mathbf{x}-\mu)^T(\mathbf{x}-\mu)}{2\sigma^2}\big)$

Which yields the likelihood function:

$L(\mu,\Sigma|\mathbf{x}_1,...,\mathbf{x}_n)=\prod_{i=1}^n\frac{1}{2\pi\sigma^2}\exp\big(-\frac{(\mathbf{x}-\mu)^T(\mathbf{x}-\mu)}{2\sigma^2}\big)=$

$=\frac{1}{(2\pi)^n\sigma^{2n}}\exp\big(-\frac{1}{2\sigma^2}\sum_{i=1}^n(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)\big)$

The posterior distribution is proportional to the likelihood times the prior distribution:

$p(\sigma^2=s|\mathbf{x}_1,...,\mathbf{x}_n;\alpha,\beta)\propto L(\mathbf{x}_1,...,\mathbf{x}_n,\mu,\sigma^2)P(\sigma^2=s|\alpha,\beta)=$

$=\frac{1}{(2\pi)^n\sigma^{2n}}\exp\big(-\frac{1}{2\sigma^2}\sum_{i=1}^n(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)\big)\cdot\frac{\beta^\alpha}{\Gamma(\alpha)}s^{-\alpha-1}\exp(\frac{\beta}{s})=\{\sigma^2=s\}=$

$=\frac{\beta s^{-\alpha-1}}{(2\pi)^n s^n\Gamma(\alpha)}\exp\big(-\frac{1}{s}\big(\beta+\frac{1}{2}\sum_{i=1}^n(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)\big)\big)$

(b)

Bayes' factor K can be found by dividing the integral of the posterior distribution of model A by the integral of the posterior distribution of model B. Since both models follow the form of (a), we can write the factor like so:

$$K=\frac{\int\frac{\beta_A s_1^{-\alpha_A-1}}{(2\pi)^n s_1^n\Gamma(\alpha_A)}\exp\Big(-\frac{1}{s_1}\Big(\beta_A+\frac{1}{2}\sum_{i=1}^n(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)\Big)\Big)ds_1}{\int\frac{\beta_B s_2^{-\alpha_B-1}}{(2\pi)^n s_2^n\Gamma(\alpha_B)}\exp\Big(-\frac{1}{s_2}\Big(\beta_B+\frac{1}{2}\sum_{i=1}^n(\mathbf{x}_i-\mu)^T(\mathbf{x}_i-\mu)\Big)\Big)ds_2}$$

Problem 2.1

(a)

$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$

$\hat{\sigma}^2 = \frac{1}{np}\sum_{i=1}^{n}(\mathbf{x}_i - \mu)^T(\mathbf{x}_i - \mu)$

(d)

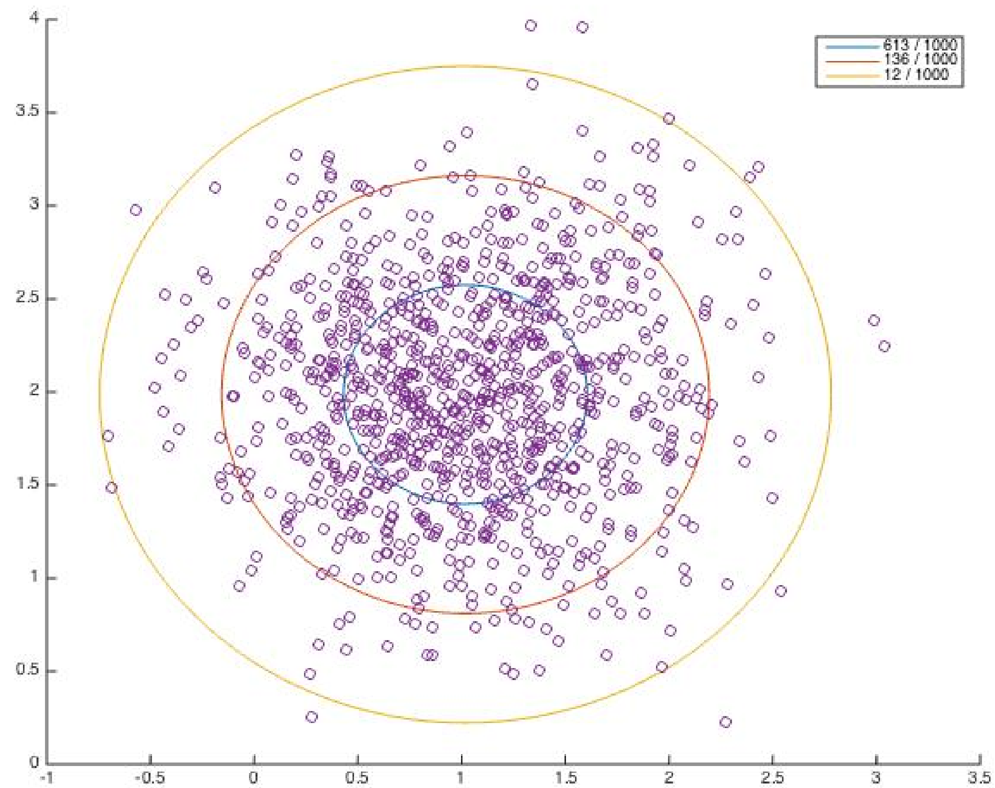

Figure 1: Plot from function created in 2.1(c) with dataset1.mat as input.

Problem 2.2

(a)

We observe that in figure 2 and 3 the posterior distribution remains the same while prior distribution changes considerably. The prior distribution in figure 2 is closer to the posterior distribution than the prior distribution in figure 3. So we can see that the prior distribution does not affect the posterior distribution greatly under these circumstances.

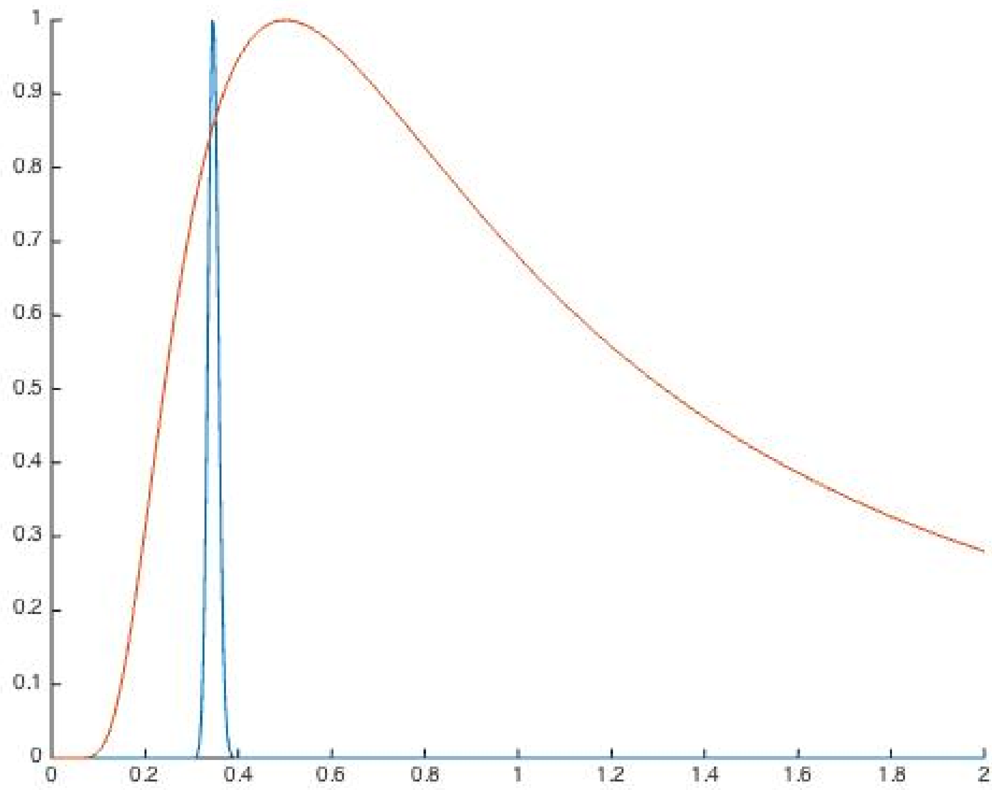

Figure 2: Graph of prior (orange) and posterior (blue) distributions with alpha=beta=1.

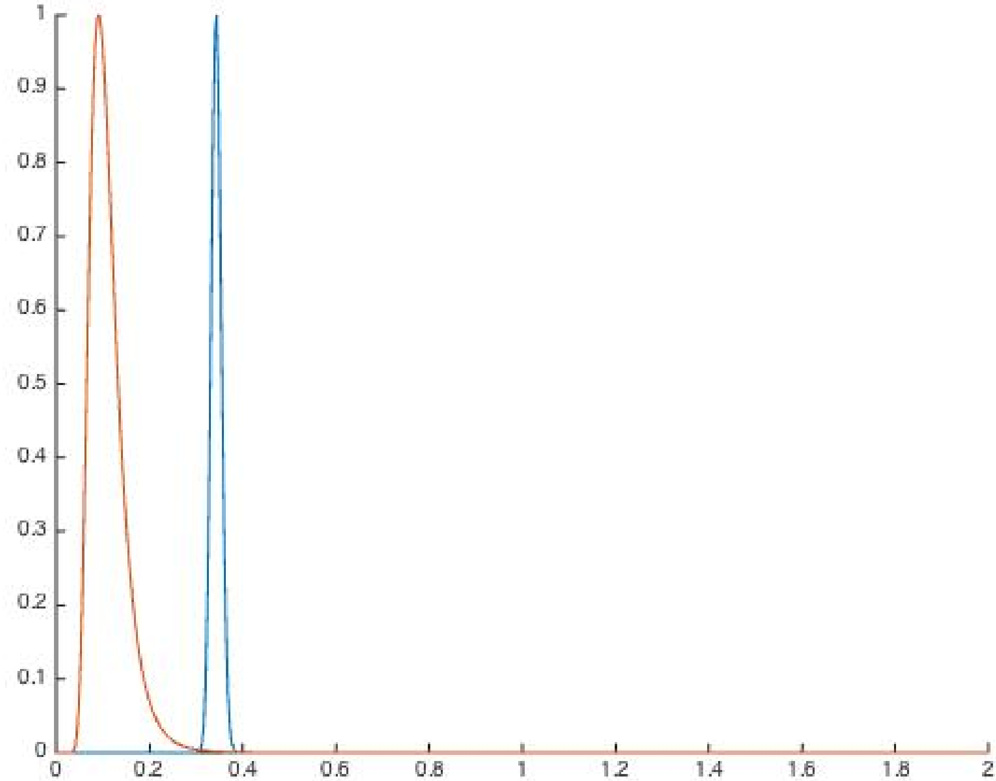

Figure 3: Graph of prior (orange) and posterior (blue) distributions with alpha=10 and beta=1.