Homework 1 Machine Learning

Oskar Hulthen 950801-1195 huoskar@student.chalmers.se
Alexander Branzell 931003-1977 alebra@student.chalmers.se

April 2018

1 Maximum likelihood estimator

Observing a single variable in the multivariate Gaussian pdf gives us:

$$N(x_i \mid \mu, \Sigma) \;=\; \frac{1}{(2\pi)^{\frac{p}{2}} \;\cdot\; |\Sigma|^{\frac{1}{2}}} \;\cdot\; exp(-\frac{1}{2} \;\cdot\; (x_i \;-\; \mu)^T \Sigma^{-1} (x_i \;-\; \mu))$$

Where p is the number of dimensions

With n observations, to get the likelihood function we take the pdf over all elements in X. In other words the following product:

$$L(X \mid \mu, \Sigma) = \prod_{i=1}^{n} N(x_i \mid \mu, \Sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{p}{2}} \;\cdot\; |\Sigma|^{\frac{1}{2}}} \;\cdot\; exp(-\frac{1}{2} \;\cdot\; (x_i \;-\; \mu)^T \Sigma^{-1} (x_i \;-\; \mu))$$

$$= \;(\frac{1}{(2\pi)^{\frac{p}{2}} \;\cdot\; |\Sigma|^{\frac{1}{2}}})^n \;\cdot\; exp(-\frac{1}{2} \;\cdot\; \sum_{i=1}^{n} (x_i \;-\; \mu)^T \Sigma^{-1} (x_i \;-\; \mu))$$

$$= \;\frac{1}{(2\pi)^{\frac{np}{2}} \;\cdot\; |\Sigma|^{\frac{n}{2}}} \;\cdot\; exp(-\frac{1}{2} \;\cdot\; \sum_{i=1}^{n} (x_i \;-\; \mu)^T \Sigma^{-1} (x_i \;-\; \mu))$$

Since Σ is of the form $\sigma^2 I$ we know that $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \cdot & 0 & 0 \\ 0 & 0 & \cdot & 0 \\ 0 & 0 & 0 & \sigma_p^2 \end{bmatrix}$

This means that $|\Sigma| \;=\; \sqrt{p}\sigma^2$ and that $\Sigma^{-1} \;=\; \frac{1}{\sigma^2} I$. Combining this with what we calculated earlier we get the following equation:

$$\frac{1}{(2\pi)^{\frac{np}{2}} \;\cdot\; (\sqrt{p}\sigma^2)^{\frac{n}{2}}} \;\cdot\; exp(-\frac{1}{2\sigma^2} \;\cdot\; \sum_{i=1}^{n} (x_i \;-\; \mu)^T (x_i \;-\; \mu))$$

Since we cannot simplify $\sum_{i=1}^{n} (x_i \;-\; \mu)^T (x_i \;-\; \mu))$ further we will substitute it with the constant c. We now take the logarithm of the likelihood function to simplify the calculations.

$$ln(\frac{1}{(2\pi)^{\frac{np}{2}} \;\cdot\; (\sqrt{p}\sigma^2)^{\frac{n}{2}}} \;\cdot\; exp(-\frac{c}{2\sigma^2})) \;=\; ln(\frac{1}{(2\pi)^{\frac{np}{2}} \;\cdot\; (\sqrt{p}\sigma^2)^{\frac{n}{2}}}) \;-\; \frac{c}{2\sigma^2}$$

$$= \;-ln((2\pi)^{\frac{np}{2}} \;\cdot\; (\sqrt{p}\sigma^2)^{\frac{n}{2}}) \;-\; \frac{c}{2\sigma^2} \;=\; -ln((2\pi)^{\frac{np}{2}}) \;-\; ln((\sqrt{p}\sigma^2)^{\frac{n}{2}}) \;-\; \frac{c}{2\sigma^2}$$

$$= -\frac{np}{2}ln(2\pi) - \frac{n}{2}ln(\sqrt{p}\sigma^2) - \frac{c}{2\sigma^2} = -\frac{np}{2}ln(2\pi) - \frac{n}{2}(ln(\sqrt{p}) + ln(\sigma^2)) - \frac{c}{2\sigma^2}$$
$$= -\frac{np}{2}ln(2\pi) - \frac{n}{4}ln(p) - n \cdot ln(\sigma) - \frac{c}{2\sigma^2}$$

To find the maximum, we derive this equation with respect to σ and set the derivative to 0

$$\frac{\partial\ ln(L(X \mid \mu, \Sigma))}{\partial\sigma} = -\frac{n}{\sigma} + \frac{c}{\sigma^3} = -\frac{n}{\sigma} + \frac{c}{\sigma^3} = 0$$
$$\rightarrow \frac{c}{\sigma^3} = \frac{n}{\sigma} \rightarrow \frac{c}{\sigma^2} = n \rightarrow \frac{c}{n} = \sigma^2 \rightarrow \sigma = \sqrt{\frac{c}{n}}$$

$$Where\ c = \sum_{i=1}^{n}(x_i - \mu)^T(x_i - \mu))$$

2 Posterior distributions

a.)

From the lectures we have that $P(\mu \mid X) \; \propto \; P(X \mid \mu)P(\mu)$. If we enter the functions given in the task we get:

$$P(\sigma^2 \; = \; S|x_1, \; ..., \; x_n \; ; \; \alpha, \; \beta) \; \propto \; P(x_1, \; ..., \; x_n \mid \sigma^2)P(\sigma^2 \; = \; S \mid \alpha, \; \beta)$$

$P(X \; = \; x \mid \sigma^2)$ is given by the task, so for $P(x_1, \; ..., \; x_n \mid \sigma^2)$ we need to multiply according to: $\prod_{i=1}^{n} P(x_i \mid \sigma^2)$

$$\prod_{i=1}^{n} P(x_i \mid \sigma^2) \; = \; (\frac{1}{2\pi\sigma^2})^n \; \cdot \; exp(\frac{\sum_{i=0}^{n}(x_i \; - \; \mu)^T(x_i \; - \; \mu)}{2\sigma^2})$$

$$= \; \frac{1}{(2\pi\sigma^2)^n} \; \cdot \; exp(\frac{1}{2\sigma^2}\sum_{i=0}^{n}(x_i \; - \; \mu)^T(x_i \; - \; \mu))$$

Now we can multiply it with with $P(\sigma \; = \; S \mid \alpha, \; \beta)$, substituting σ^2 with S in the earlier equation and $\sum_{i=0}^{n}(x_i \; - \; \mu)^T(x_i \; - \; \mu)$ with c.

$$\frac{1}{(2\pi S)^n} \cdot exp(\frac{c}{2S}) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}S^{-\alpha \; - \; 1} \cdot exp\frac{-\beta}{S} \; = \; \frac{1}{(2\pi S)^n} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}S^{-\alpha \; - \; 1} \cdot exp(\frac{c}{2S} + \frac{-\beta}{S})$$

$$= \; \frac{1}{(2\pi)^n} \; \cdot \; \frac{1}{S^n} \cdot \; \frac{\beta^\alpha}{\Gamma(\alpha)}S^{-\alpha \; - \; 1} \; \cdot \; exp(\frac{c}{2S} \; + \; \frac{-\beta}{S})$$

$$= \; \frac{1}{(2\pi)^n} \; \cdot \; \frac{\beta^\alpha}{\Gamma(\alpha)}S^{-\alpha \; - \; n- \; 1} \; \cdot \; exp(\frac{c}{2S} \; + \; \frac{-\beta}{S})$$

From the hint we know that $P(\sigma^2 \; = \; S|x_1, \; ..., \; x_n \; ; \; \alpha, \; \beta)$ will be of the form $\frac{\beta^\alpha}{\Gamma(\alpha)}S^{-\alpha \; - \; 1} \cdot \exp(\frac{-\beta}{S})$. Then we know from our earlier statements that:

$$\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}S^{-\alpha_1 \; - \; 1} \; \cdot \; exp(\frac{-\beta_1}{S}) \propto \frac{1}{(2\pi)^n} \; \cdot \; \frac{\beta^\alpha}{\Gamma(\alpha)}S^{-\alpha \; - \; n- \; 1} \; \cdot \; exp(\frac{c}{2S} \; + \; \frac{-\beta}{S})$$

We can remove all the values that do not depend on S, as they are considered constant and now the left hand side will be equal to the right hand side.

$$S^{-\alpha_1 \; - \; 1} \; \cdot \; exp(\frac{-\beta_1}{S}) \; = \; S^{-\alpha \; - \; n- \; 1} \; \cdot \; exp(\frac{c}{2S} \; + \; \frac{-\beta}{S})$$

$$= \; S^{-(a \; + \; n) \; - \; 1} \; \cdot \; exp(-\frac{1}{S}(\frac{c}{2} + \beta))$$

From this we can see that $\alpha_1 \; = \; \alpha + n$ and $\beta_1 \; = \; \beta \; + \; \frac{c}{2}$

This gives us the posterior distribution of:

$$P(\sigma^2 \; = \; S|x_1, \; ..., \; x_n \; ; \; \alpha, \; \beta) = \frac{(\beta \; + \; \frac{c}{2})^{\alpha+n}}{\Gamma(\alpha+n)}S^{-(\alpha+n) \; - \; 1} \; \cdot \; exp(\frac{-\beta \; + \; \frac{c}{2}}{S})$$

Where $c = \sum_{i=0}^{n}(x_i \; - \; \mu)^T(x_i \; - \; \mu)$

b.)

To maximize the posterior distribution, we started by taking the logarithm of the posterior distribution to simplify the calculations:

$$ln(P(\sigma^2 \ = \ S|x_1, \ ..., \ x_n \ ; \ \alpha, \ \beta)) = ln(\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}) + ln(S^{-\alpha_1 \ - \ 1}) \ + \frac{-\beta_1}{S}$$

$$= \alpha_1 \cdot ln(\beta_1) - ln(\Gamma(\alpha_1)) + (-\alpha_1 \ - \ 1) \cdot ln(S) \ + \frac{-\beta_1}{S}$$

From the logarithmic posterior distribution, we can derive with respect to S and later set to zero to find the S that maximizes the posterior distribution:

$$\frac{\delta \ ln(P(\sigma^2 \ = \ S|x_1, \ ..., \ x_n \ ; \ \alpha, \ \beta))}{\delta S} = \frac{(-\alpha_1 \ - \ 1)}{S} + \frac{\beta_1}{S^2}$$

$$0 = \frac{(-\alpha_1 \ - \ 1)}{S} + \frac{\beta_1}{S^2} \rightarrow \frac{S^2}{S} = S = \frac{\beta_1}{\alpha_1 + 1}$$

So to maximize the posterior distribution S should be equal to $\frac{\beta_1}{\alpha_1+1}$.
For model A, where $\alpha = 1$ and $\beta = 1$ then $\alpha_1 = 1 + n$ and $\beta_1 = \frac{C}{2} + 1$. Yielding the following S:

$$S_{MA} = \frac{\frac{C}{2} + 1}{1 + n + 1} = \frac{C + 2}{2(n + 2)} = \frac{2 + \sum_{i=0}^{n}(x_i \ - \ \mu)^T(x_i \ - \ \mu)}{2n + 4}$$

For model B, where $\alpha = 10$ and $\beta = 1$ then $\alpha_1 = 10 + n$ and $\beta_1 = \frac{C}{2} + 1$. Yielding the following S:

$$S_{MB} = \frac{\frac{C}{2} + 1}{10 + n + 1} = \frac{C + 2}{2(n + 11)} = \frac{2 + \sum_{i=0}^{n}(x_i \ - \ \mu)^T(x_i \ - \ \mu)}{2n + 22}$$