Homework set 2 TDA231 Algorithms for Machine Learning & Inference

Kevin Jaquier 921119T334
Sandra Viknander 9003012482

March 2018

1 Theoretical problems

1.1 Maximum likelihood estimator (MLE)

We consider a set $x_1,...,x_n$ generated from a spherical multivariate Gaussian distribution $N(\mu,\sigma^2\mathbf{I})$ where $\mu\epsilon\mathbf{R}^p$. Thus our data is a set in space of dimension p. The likelihood of our dataset $\mathbf{x}$ given the parameter $\sigma^2\mathbf{I}$ is given in equation 1.taking the logarithm of the likelyhood will help us when we are going to derive the likelihood to find the MLE, this new log likelihood can be seen in equation2. Now taking the derivative, with respect to σ^2 of the logarithmic likelihood and setting it equal to zero we arrive at the MLE for σ^2 shown in equation **??**

$$\begin{aligned} P(\mathbf{x}|\sigma^2\mathbf{I}) = \prod_{i=1}^{n} P(x_i|\sigma^2\mathbf{I}) = \\ \frac{1}{(2\sigma^2)^{\frac{p}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \end{aligned} \tag{1}$$

$$-\frac{pn}{2} Log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} \tag{2}$$

$$\begin{aligned} 0 = -\frac{pn\pi\sigma 4}{2 * 2\pi\sigma^2} + \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^3} \rightarrow \\ \frac{pn}{\sigma} = \sum_{i=1} n \frac{(x_i-\mu)^2}{\sigma^3} \rightarrow \\ \sigma^2 = \frac{\sum_{i=1} n}{pn} \end{aligned} \tag{3}$$

1.2 Posterior

1.2.1 Deriving the posterior

If the inverse gamma distribution really are a conjugate to the spherical Gaussian distribution then the posterior should also be a inverse gamma distribution. Calculating the posterior with the basis the inverse gamma function is conjugate to the spherical Gaussian distribution we start with equation 4. Ignoring the constants that are not dependent on σ^2, sins we only care about the proportionality with the posterior, $(prior : \frac{\beta^\alpha}{\Gamma(\alpha)}; Likelihood : (2\pi)^{-n})$ and combining the two expressions we then get the expression in equation 5. We can then see that this is just another inverse gamma distribution with new α and β values $f(\sigma^2{,}\alpha_{new}{,}\beta_{new} = f(\sigma^2{,}\alpha + n, \beta + \sum_{i=n}^{n} \frac{||x_i-\mu||^2}{2}))$

$$\begin{aligned} P(\sigma^2 = s|\mathbf{x}) \propto P(X = \mathbf{x}|\sigma^2)P(\sigma^2 = s|\alpha{,}\beta) \\ \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} e^{-\frac{||x_i-\mu||^2}{2\sigma^2}} \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} e^{\frac{-\beta}{\sigma^2}} \end{aligned} \tag{4}$$

$$(\sigma^2)^{-\alpha-1-n} e^{\frac{1}{\sigma^2}(\beta + \sum_{i=n}^{n} \frac{||x_i-\mu||^2}{2})} \tag{5}$$

1.2.2 Computing the MAP

To compute the MAP we begin to take the log of the posterior sins this will make it easier to compute the derivative. We can see this in equation 6. Now deriving with respect to σ^2 and setting it equal to zero we get 7. Now calculating the σ^2_{MAP} with the parameters $(\alpha{,}\beta) = (1{,}1) and (\alpha{,}\beta) = (10{,}1)$ is shown in table 1

$$\begin{aligned} log((\sigma^2)^{-(\alpha+1+n)} e^{\frac{1}{\sigma^2}(\beta + \sum_{i=n}^{n} \frac{||x_i-\mu||^2}{2}}) \rightarrow \\ log((\sigma^2)^{-(\alpha+1+n)}) - \frac{1}{\sigma^2}(\beta + \sum_{i=n}^{n} \frac{||x_i-\mu||^2}{2}) \end{aligned} \tag{6}$$

$$\begin{aligned} 0 = -(\alpha+1+n)\frac{1}{(\sigma^2)} + \frac{1}{(\sigma^2)^2} * (\beta + \sum_{i=1}^{n} \frac{||x_i-\mu||^2}{2}) \rightarrow \\ \frac{\alpha+1+n}{\sigma^2} = \frac{1}{(\sigma^2)^2}(\beta + \sum_{i=1}^{n} \frac{||x_i-\mu||^2}{2}) \rightarrow \\ \sigma^2 = \frac{(\beta + \sum_{i=1}^{n} \frac{||x_i-\mu||^2}{2})}{\alpha+1+n} \end{aligned} \tag{7}$$

2 Practical problems

See the completed Jupyter notebook.

$n << \infty$		
	Parameters (α,β)	σ^2
	(1,1)	$\frac{1+\sum_{i=1}^{n}\frac{\|\|x_i-\mu\|\|^2}{2}}{2+n}$
	(1,10)	$\frac{1+\sum_{i=1}^{n}\frac{\|\|x_i-\mu\|\|^2}{2}}{11+n}$
$n \rightarrow \infty$		
	(1,1)	$MLE: \sum_{i=1}^{n}\frac{\|\|x_i-\mu\|\|^2}{2n}$
	(1,10)	$MLE: \sum_{i=1}^{n}\frac{\|\|x_i-\mu\|\|^2}{2n}$

Table 1: σ^2 for different values of parameter α and β. With both a small data set $\mathbf{x}$ $n << \infty$ and when the data set approaches infinity