MAP Problem with Solution

Course: Introduction to Data Mining
Professor: Dr. Tahaei
Author: Parinaz Kanan

December 23, 2024

Problem 1: Gaussian MAP example: Course hours. We started taking an advanced systems class at CMU. We're part way through the semester and are starting to wonder what the mean number of hours per week students are spending on this class. Asking a few friends in the course, we collect a dataset of their hours per week:

$$D = \{x^{(i)}\}_{i=1}^{4} = \{18, 20, 14, 10\}$$

We do a quick MLE assuming a Gaussian distribution on our friends' hours, $x \sim \mathcal{N}(\mu, \sigma)$, and find that the mean is $\mu_{\text{MLE}} = 15.5$.

Not wanting to overfit our small dataset, we do a little digging in past course evaluation data and find statistics saying that over the past several years the reported hours per week have a mean of $\nu = 23.9$ and a standard deviation of $\tau = 1.56$. We use these statistics to create a Gaussian prior on our mean parameter, $\mu \sim \mathcal{N}(\nu = 23.9, \tau = 1.56)$. (To avoid confusion, we chose different symbols for this second set of mean and standard deviation.)

Note that we now have two different Gaussian distributions, one for our hours data, $x \sim \mathcal{N}(\mu, \sigma)$, (with unknown parameters, μ and σ) and one for our mean parameter, $\mu \sim \mathcal{N}(\nu = 23.9, \tau = 1.56)$.

In this example problem, we are going to focus on using MAP to estimate the mean for our data distribution, μ. To do this, we'll hold the standard deviation, σ, constant. (The mean and standard deviation parameters of the prior distribution, ν and τ, are already constants.)

Solution. **Formulate the likelihood times the prior**

$$\begin{aligned} p(\mu|D) &\propto p(\mu)\prod_{i=1}^{4} p(x^{(i)}|\mu) \\ &= \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\mu-\nu)^2} \prod_{i=1}^{4} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2} \end{aligned}$$

Set the objective function to the negative log likelihood times the prior

$$\begin{aligned} J(\mu) &= -\log\left(p(\mu)\prod_{i=1}^{4} p(x^{(i)}|\mu)\right) \\ &= -\log\left(\frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\mu-\nu)^2} \prod_{i=1}^{4} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2}\right) \\ &= -\log\left(\frac{1}{\sqrt{2\pi\tau^2}}\right) + \frac{1}{2\tau^2}(\mu-\nu)^2 - N\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \sum_{i=1}^{4}\left(\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2\right) \end{aligned}$$

Compute derivative of objective, $\frac{\partial J}{\partial \theta}$

$$\begin{aligned} \frac{\partial J}{\partial \mu} &= \frac{\partial}{\partial \mu}\left(-\log\left(\frac{1}{\sqrt{2\pi\tau^2}}\right) + \frac{1}{2\tau^2}(\mu-\nu)^2 - \log\left(\sqrt{\frac{N}{2\pi\sigma^2}}\right) + \sum_{i=1}^{N}\frac{1}{2\sigma^2}(x^{(i)}-\mu)^2\right) \\ &= \frac{1}{\tau^2}(\mu-\nu) - \sum_{i=1}^{N}\frac{1}{\sigma^2}(x^{(i)}-\mu) \end{aligned}$$

Find $\hat{\theta}$ by setting derivative equal to zero and solve for θ

$$\begin{aligned} \frac{\partial J}{\partial \mu} &= 0 \\ 0 &= \frac{1}{\tau^2}(\mu-\nu) - \sum_{i=1}^{4}\frac{1}{\sigma^2}(x^{(i)}-\mu) \\ 0 &= \frac{1}{\tau^2}\mu - \frac{\nu}{\tau^2} + \frac{N}{\sigma^2}\mu - \frac{1}{\sigma^2}\sum_{i=1}^{4} x^{(i)} \\ \frac{1}{\tau^2}\mu + \frac{N}{\sigma^2}\mu &= \frac{\nu}{\tau^2} + \frac{1}{\sigma^2}\sum_{i=1}^{4} x^{(i)} \\ \sigma^2\mu + N\tau^2\mu &= \sigma^2\nu + \tau^2\sum_{i=1}^{4} x^{(i)} \\ (\sigma^2 + N\tau^2)\mu &= \sigma^2\nu + \tau^2\sum_{i=1}^{4} x^{(i)} \\ \hat{\mu}_{MAP} &= \frac{1}{\sigma^2 + N\tau^2}\left(\sigma^2\nu + \tau^2\sum_{i=1}^{4} x^{(i)}\right) \end{aligned}$$

if we (rather naively) assume that the standard deviation for this semester's data is the same as the prior standard deviation, $\sigma = \tau$:

$$\hat{\mu} = \frac{1}{1+N}\left(\nu + \sum_{i=1}^{4} x^{(i)}\right)$$

This estimate is the same as if we treated the prior mean as just one other data point and included it into our MLE mean!

$$\hat{\mu} = \frac{1}{5}(23.9 + (18 + 20 + 14 + 10)) = 17.8 \text{ hours per week}$$

$\square$