4.14

Intro

In this exercise, we will perform MAP estimation for the mean of a 1D Gaussian. One important aspect of this exercise is that it highlights the contributions of the prior and likelihood to the posterior. We will see the details in the solution.

Solution

a)

The mean posterior is $\mu|D \sim \mathcal{N}(\mu|m_n, s_n^2)$. The MAP estimation is equal to the mode, which in turn is equal to the mean for Gaussian distributions. Therefore $\mu_{MAP} = m_n$. Using the posterior distribution of the mean given on section 4.6.1:

$$\mu_{MAP} = m_n = s_n^2 \left(\frac{n\bar{x}}{\sigma^2} + \frac{m}{s^2}\right) = \left(\frac{1}{s^2} + \frac{n}{\sigma^2}\right)^{-1} \left(\frac{n\bar{x}}{\sigma^2} + \frac{m}{s^2}\right) = \frac{m\sigma^2 + ns^2\bar{x}}{\sigma^2 + ns^2} \tag{1}$$

where s_n is the posterior variance, and its expression is also deduced in section 4.6.1. Therefore, the MAP is a weighted average between the prior mean m and the sample average $\bar{x}$.

b)

We want to prove that when the number of samples increases, the MAP convergers to the MLE:

$$\lim_{n\to\infty} \frac{m\sigma^2 + ns^2\bar{x}}{\sigma^2 + ns^2} = \lim_{n\to\infty} \frac{ns^2\bar{x}}{ns^2} = \bar{x} = \mu_{MLE} \tag{2}$$

c)

Now, we will consider n a small number and increase the prior variance s^2

$$\lim_{s^2\to\infty} \frac{m\sigma^2 + ns^2\bar{x}}{\sigma^2 + ns^2} = \lim_{s^2\to\infty} \frac{ns^2\bar{x}}{ns^2} = \bar{x} = \mu_{MLE} \tag{3}$$

The result is the same as in item b. We will comment on that on the conclusion.

d)

At last, we will see what happens when we consider n a small number and decrease the prior variance s^2

$$\lim_{s^2\to o} \frac{m\sigma^2 + ns^2\bar{x}}{\sigma^2 + ns^2} = \frac{m\sigma^2}{\sigma^2} = m \tag{4}$$

The result is the prior mean.

Conclusion

In this exercise, we performed MAP estimation for the mean of a 1D Gaussian. We discovered that the MAP is a weighted average between the prior mean m and the sample averege $\bar{x}$. What is interesting to note about this expression is the influence of the means and variances on the over the final result.

First note that when the variance of the prior/original distribution increases this penalizes the prior/likelihood by increasing the weight of its competitor (likelihood/prior) in the MAP expression. To put it more simply, big prior variances makes the prior unrelaible compared to the likelihood. So the latter term dominates. On the other hand, big original distribution variances makes our data unrelaible compared with the prior. So the latter takes over.

The second factor that has influence over the expression is the number of samples. We can see that the likelihood gains strenght in the weighted average not only by the variance of the prior, but also by the number of samples. That's why when we increase the number of samples, the MLE takes over.