Essay title

Machine Learning, Advanced Course/DD2434/mladv24

Aurhor
KTH Royal Institute of Technology
School of Engineering Sciences in Chemistry, Biotechnology and Health

November 19, 2024

Andrea Stanziale, Leonardo Lüder
stanz@kth.se, luder@kth.se

Contents

0.1 1.2.8 1

Andrea Stanziale, Leonardo Lüder
stanz@kth.se, luder@kth.se

1.1

1.2

1.2.7

We implement a function that generates a dataset of N points in the plane, where each point is drawn from a normal distribution with mean μ and precision τ. For reproducibility we set the seed of the random number generator to 0. We generate datasets for $N = 10, 100, 1000$ and plot the generated values as histograms, which results in figure Figure 4.

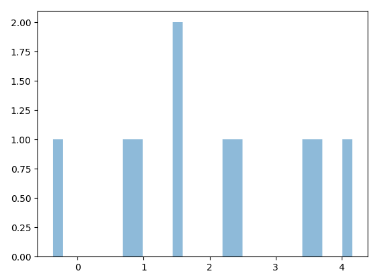

Figure 1: N=10

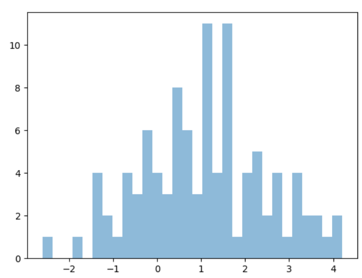

Figure 2: N=100

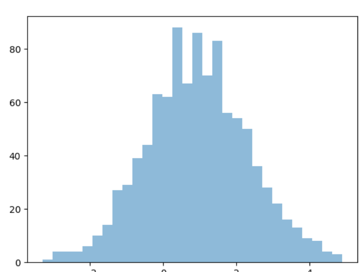

Figure 3: N=1000

Figure 4: Histograms of generated datasets

We see that the more datapoints we have the more the histogram resembles a normal distribution which is as expected.

0.1 1.2.8

We derive the ML estimates for μ and τ. The posterior distribution is proportional to the product of the likelihood and the prior:

$$p(\mu, \tau|X) \propto p(X|\mu, \tau)p(\mu|\tau)p(\tau) \tag{1}$$

The likelihood is given by:

$$p(X|\mu, \tau) = \prod_{i=1}^{N} p(X_i|\mu, \tau) = \prod_{i=1}^{N} \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(X_i-\mu)^2} \tag{2}$$

where we assume $X_i|\mu, \tau \sim N(\mu, \frac{1}{\tau})$.

The prior distributions are:

$$p(\mu|\tau) = N(\mu_0, \frac{1}{\lambda_0 \tau}); \quad p(\tau) = \text{Gamma}(\alpha_0, \beta_0) \tag{3}$$

This gives the joint posterior:

$$p(\mu, \tau|X) \propto \prod_{i=1}^{N} \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(X_i-\mu)^2} \times \sqrt{\frac{\lambda_0 \tau}{2\pi}} e^{-\frac{\lambda_0 \tau}{2}(\mu-\mu_0)^2} \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \tau^{\alpha_0-} \tag{4}$$

Simplifying, we get:

$$p(\mu, \tau|X) \propto \left(\frac{\tau}{2\pi}\right)^{\frac{N+1}{2}} e^{-\frac{\tau}{2}\left(\sum_{i=1}^{N}(X_i-\mu)^2+\lambda_0(\mu-\mu_0)^2+2\beta_0\right)} \tau^{\alpha_0+\frac{N}{2}-1} \tag{5}$$

After completing the square for terms involving μ and simplifying, the posterior distribution of μ given X and τ is:

$$p(\mu|X, \tau) \sim N\left(\frac{\sum X_i + \lambda_0 \mu_0}{N + \lambda_0}, \frac{1}{(N + \lambda_0)\tau}\right) \tag{6}$$