Bayesian linear regression with fixed noise precision

Jan 2, 2023

1 Preliminaries

Definition 1. *(Linear Gaussian system; originally 4.124 on p119)*
Let $\boldsymbol{x} \in \mathbb{R}^{D_x}$ and $\boldsymbol{y} \in \mathbb{R}^{D_y}$ be two random variables. The following generative model

$$\begin{aligned} p(\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \\ p(\boldsymbol{y} | \boldsymbol{x}) &= \mathcal{N}(\boldsymbol{y} | \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \boldsymbol{\Sigma}_y) \end{aligned}$$

is the linear Gaussian system, where $\boldsymbol{A} \in \mathbb{R}^{D_y \times D_x}$ and $\boldsymbol{b} \in \mathbb{R}^{D_y}$.

Theorem 2. *(Bayes rule for linear Gaussian sytems, originally 4.125 on pp119)*
Given a linear Gaussian system, the posterior is given by

$$\begin{aligned} p(\boldsymbol{x} | \boldsymbol{y}) &= \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}}) \\ \boldsymbol{\Sigma}_{\text{post}}^{-1} &= \boldsymbol{\Sigma}_x^{-1} + \boldsymbol{A}^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{A} \\ \boldsymbol{\mu}_{\text{post}} &= \boldsymbol{\Sigma}_{\text{post}} [\boldsymbol{A}^T \boldsymbol{\Sigma}_y^{-1} (\boldsymbol{y} - \boldsymbol{b}) + \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\mu}_x] \end{aligned}$$

Note that the posterior uncertainty (i.e., the covariance is independent on $\boldsymbol{y}$) and the posterior mean is a linear function of $\boldsymbol{y}$.

2 Core

Likelihood.

$$p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{w}, \beta) = \prod_{i=1}^{N} \mathcal{N}(y_i | \mu_i, \beta^{-1}) = \mathcal{N}(\boldsymbol{y} | \boldsymbol{X}\boldsymbol{w}, \beta^{-1}\boldsymbol{I})$$

with $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{w}$. Let's call β the noise precision.

Prior.

$$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w} | \boldsymbol{0}, \alpha^{-1}\boldsymbol{I})$$

Let's call α the prior precision. Of course, we could also have used a Gaussian with an arbitrary mean and covariance. But let's keep it simple for now.

Posterior.

The key insight is that $\boldsymbol{X}$ transforms $\boldsymbol{w}$ into the mean of $p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{w}, \beta)$. So we can apply Theorem 2 to obtain the posterior:

$$\begin{aligned} p(\boldsymbol{w} | \boldsymbol{y}, \boldsymbol{X}, \beta^{-1}) &= \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}}) \\ \\ \boldsymbol{\Sigma}_{\text{post}}^{-1} &= \boldsymbol{\Sigma}_w^{-1} + \boldsymbol{X}^T \boldsymbol{\Sigma}_y^{-1} \boldsymbol{X} \\ &= (\alpha^{-1}\boldsymbol{I})^{-1} + \boldsymbol{X}^T (\beta^{-1}\boldsymbol{I})^{-1} \boldsymbol{X} \\ &= \alpha\boldsymbol{I} + \beta \boldsymbol{X}^T \boldsymbol{X} \\ \\ \boldsymbol{\mu}_{\text{post}} &= \boldsymbol{\Sigma}_{\text{post}} [\boldsymbol{X}^T \boldsymbol{\Sigma}_y^{-1} (\boldsymbol{y} - \boldsymbol{0}) + \boldsymbol{\Sigma}_w^{-1} \boldsymbol{\mu}_w] \\ &= \boldsymbol{\Sigma}_{\text{post}} [\boldsymbol{X}^T (\beta^{-1}\boldsymbol{I})^{-1} \boldsymbol{y} + (\alpha^{-1}\boldsymbol{I})^{-1} \boldsymbol{0}] \\ &= \beta \boldsymbol{\Sigma}_{\text{post}} \boldsymbol{X}^T \boldsymbol{y} \end{aligned}$$

3 Experiment