Bayesian linear regression with fixed noise precision

Jan 2, 2023

1 Preliminaries

Definition 1. *(Linear Gaussian system; originally 4.124 on p119 of Murphy)*
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 of Murphy)*
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.

$$p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{w}, \beta) = \prod_{i=1}^{N} \mathcal{N}(y_i | \mu_i, \sigma^2) = \mathcal{N}(\boldsymbol{y} | \boldsymbol{X}\boldsymbol{w}, \sigma^2 \boldsymbol{I})$$

with $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{w}$.

Prior.

$$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w} | \boldsymbol{w}_0, \boldsymbol{V}_0)$$

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:

$$\begin{aligned} p(\boldsymbol{w} | \boldsymbol{y}, \boldsymbol{X}, \sigma^2) &= \mathcal{N}(\boldsymbol{w} | \boldsymbol{w}_N, \boldsymbol{V}_N) \\ \boldsymbol{w}_N &= \boldsymbol{V}_N (\boldsymbol{X}^T (\sigma^2 \boldsymbol{I})^{-1} \boldsymbol{y} + \boldsymbol{V}_0^{-1} \boldsymbol{w}_0) \\ &= \boldsymbol{V}_N (\boldsymbol{V}_0^{-1} \boldsymbol{w}_0 + (1/\sigma^2) \boldsymbol{X}^T \boldsymbol{y}) \\ \boldsymbol{V}_N &= (\boldsymbol{V}_0^{-1} + \boldsymbol{X}^T (\sigma^2 \boldsymbol{I})^{-1} \boldsymbol{X})^{-1} \\ &= (\boldsymbol{V}_0^{-1} + (1/\sigma^2) \boldsymbol{X}^T \boldsymbol{X})^{-1} \end{aligned}$$

Posterior predictive.

$$\begin{aligned} p(\tilde{\boldsymbol{y}} | \tilde{\boldsymbol{X}}, \boldsymbol{X}, \boldsymbol{y}, \sigma^2) &= \int \mathcal{N}(\tilde{\boldsymbol{y}} | \tilde{\boldsymbol{X}}^T \boldsymbol{w}, \sigma^2 \boldsymbol{I}_m) \mathcal{N}(\boldsymbol{w} | \boldsymbol{w}_N, \boldsymbol{V}_N) \, d\boldsymbol{w} \\ &= \int \mathcal{N}(\tilde{\boldsymbol{y}} | \tilde{\boldsymbol{X}}^T \boldsymbol{w}_N, \sigma^2 \boldsymbol{I}_m + \tilde{\boldsymbol{X}}^T \boldsymbol{V}_N \tilde{\boldsymbol{X}}) \quad \text{(By Eq 4.126)} \end{aligned}$$