PROOF OF THE POSTERIOR OF BAYSIAN LM

Proof:

We want to show that

- for a Gaussian prior on $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I}_p)$
- for a Gaussian Likelihood $y \mid \mathbf{X}, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{X}^{\top}\boldsymbol{\theta}, \sigma^{2}\boldsymbol{I}_{n})$

the resulting posterior is Gaussian $\mathcal{N}(\sigma^{-2}\mathbf{A}^{-1}\mathbf{X}^{\top}\mathbf{y}, \mathbf{A}^{-1})$ with $\mathbf{A} := \sigma^{-2}\mathbf{X}^{\top}\mathbf{X} + \frac{1}{\tau^{2}}\mathbf{I}_{p}$. Plugging in Bayes' rule and multiplying out yields

$$p(\theta | \mathbf{X}, \mathbf{y}) \propto p(\mathbf{y} | \mathbf{X}, \theta) q(\theta) \propto \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\theta)^\top (\mathbf{y} - \mathbf{X}\theta) - \frac{1}{2\tau^2} \theta^\top \theta\right]$$

$$= \exp\left[-\frac{1}{2} \left(\underbrace{\sigma^{-2} \mathbf{y}^\top \mathbf{y}}_{\text{doesn't depend on } \theta} - 2\sigma^{-2} \mathbf{y}^\top \mathbf{X}\theta + \sigma^{-2} \theta^\top \mathbf{X}^\top \mathbf{X}\theta + \tau^{-2} \theta^\top \theta\right)\right]$$

$$\propto \exp\left[-\frac{1}{2} \left(\sigma^{-2} \theta^\top \mathbf{X}^\top \mathbf{X}\theta + \tau^{-2} \theta^\top \theta - 2\sigma^{-2} \mathbf{y}^\top \mathbf{X}\theta\right)\right]$$

$$= \exp\left[-\frac{1}{2} \theta^\top \underbrace{\sigma^{-2} \mathbf{X}^\top \mathbf{X} + \tau^{-2} \mathbf{I}_p}_{\mathbf{y}} \theta + \sigma^{-2} \mathbf{y}^\top \mathbf{X}\theta\right]$$

This expression resembles a normal density - except for the term in red!

XX

PROOF OF THE POSTERIOR OF BAYSIAN LM / 2

Note: We need not worry about the normalizing constant since its mere role is to convert probability functions to density functions with a total probability of one. We subtract a (not yet defined) constant *c* while compensating for this change by adding the respective terms ("adding 0"), emphasized in green:

$$p(\theta|\mathbf{X}, \mathbf{y}) \propto \exp\left[-\frac{1}{2}(\theta - c)^{\top} \mathbf{A}(\theta - c) - c^{\top} \mathbf{A}\theta + \underbrace{\frac{1}{2} c^{\top} \mathbf{A} c}_{\text{doesn't depend on } \theta} + \sigma^{-2} \mathbf{y}^{\top} \mathbf{X} \theta\right]$$
$$\propto \exp\left[-\frac{1}{2}(\theta - c)^{\top} \mathbf{A}(\theta - c) - c^{\top} \mathbf{A}\theta + \sigma^{-2} \mathbf{y}^{\top} \mathbf{X} \theta\right]$$

If we choose *c* such that $-c^{\top} \mathbf{A} \theta + \sigma^{-2} \mathbf{y}^{\top} \mathbf{X} \theta = 0$, the posterior is normal with mean *c* and covariance matrix \mathbf{A}^{-1} . Taking into account that **A** is symmetric, this is if we choose

$$\sigma^{-2} \mathbf{y}^{\mathsf{T}} \mathbf{X} = \boldsymbol{c}^{\mathsf{T}} \mathbf{A}$$
$$\Leftrightarrow \quad \sigma^{-2} \mathbf{y}^{\mathsf{T}} \mathbf{X} \mathbf{A}^{-1} = \boldsymbol{c}^{\mathsf{T}}$$
$$\Leftrightarrow \quad \boldsymbol{c} = \sigma^{-2} \mathbf{A}^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{y}$$

as claimed.

PREDICTIVE DISTRIBUTION

Based on the posterior distribution

$$\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{y} \sim \mathcal{N}(\sigma^{-2} \boldsymbol{A}^{-1} \mathbf{X}^{\top} \mathbf{y}, \boldsymbol{A}^{-1})$$

we can derive the predictive distribution for a new observations \mathbf{x}_* . The predictive distribution for the Bayesian linear model, i.e. the distribution of $\boldsymbol{\theta}^{\top}\mathbf{x}_*$, is

$$y_* \mid \mathbf{X}, \mathbf{y}, \mathbf{x}_* \sim \mathcal{N}(\sigma^{-2} \mathbf{y}^\top \mathbf{X} \mathbf{A}^{-1} \mathbf{x}_*, \mathbf{x}_*^\top \mathbf{A}^{-1} \mathbf{x}_*)$$

Note that $y_* = \theta^T \mathbf{x}_* + \epsilon$, where both the posterior of θ and ϵ are Gaussians. By applying the rules for linear transformations of Gaussians, we can confirm that $y_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*$ is a Gaussian, too.

× 0 0 × × ×