

# PROOF OF THE POSTERIOR OF BAYSIAN LM

## Proof:

We want to show that

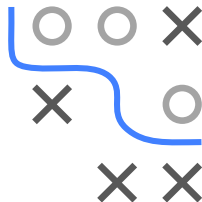
- for a Gaussian prior on  $\theta \sim \mathcal{N}(\mathbf{0}, \tau^2 \mathbf{I}_p)$
- for a Gaussian Likelihood  $y | \mathbf{X}, \theta \sim \mathcal{N}(\mathbf{X}^\top \theta, \sigma^2 \mathbf{I}_n)$

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

$$\begin{aligned} p(\theta | \mathbf{X}, \mathbf{y}) &\propto p(\mathbf{y} | \mathbf{X}, \theta) q(\theta) \propto \exp \left[ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\theta)^\top (\mathbf{y} - \mathbf{X}\theta) - \frac{1}{2\tau^2} \theta^\top \theta \right] \\ &= \exp \left[ -\frac{1}{2} \left( \underbrace{\sigma^{-2} \mathbf{y}^\top \mathbf{y}}_{\text{doesn't depend on } \theta} - 2\sigma^{-2} \mathbf{y}^\top \mathbf{X}\theta + \sigma^{-2} \theta^\top \mathbf{X}^\top \mathbf{X}\theta + \tau^{-2} \theta^\top \theta \right) \right] \\ &\propto \exp \left[ -\frac{1}{2} \left( \sigma^{-2} \theta^\top \mathbf{X}^\top \mathbf{X}\theta + \tau^{-2} \theta^\top \theta - 2\sigma^{-2} \mathbf{y}^\top \mathbf{X}\theta \right) \right] \\ &= \exp \left[ -\frac{1}{2} \theta^\top \underbrace{\left( \sigma^{-2} \mathbf{X}^\top \mathbf{X} + \tau^{-2} \mathbf{I}_p \right)}_{:=\mathbf{A}} \theta + \sigma^{-2} \mathbf{y}^\top \mathbf{X}\theta \right] \end{aligned}$$

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



# 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.

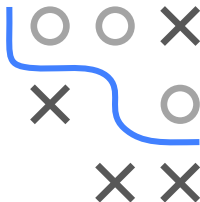
We subtract a (not yet defined) constant  $c$  while compensating for this change by adding the respective terms (“adding 0”), emphasized in green:

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

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

$$\begin{aligned} \sigma^{-2} \mathbf{y}^\top \mathbf{X} &= \mathbf{c}^\top \mathbf{A} \\ \Leftrightarrow \sigma^{-2} \mathbf{y}^\top \mathbf{X} \mathbf{A}^{-1} &= \mathbf{c}^\top \\ \Leftrightarrow \mathbf{c} &= \sigma^{-2} \mathbf{A}^{-1} \mathbf{X}^\top \mathbf{y} \end{aligned}$$

as claimed.



# PREDICTIVE DISTRIBUTION

Based on the posterior distribution

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

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

