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

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

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

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 $y_* = \boldsymbol{\theta}^T\mathbf{x}_* + \epsilon$, where both the posterior of $\boldsymbol{\theta}$ and ϵ 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.