

**Solution 1:**

- (a) • Hypothesis space  $\mathcal{H}$  is defined as:

$$\mathcal{H} = \{f(\mathbf{x}) = \mathbf{X}\boldsymbol{\beta} \mid \boldsymbol{\beta} \in \mathbb{R}^p\}$$

- We fit a linear model, ergo using the  $L2$  loss makes sense (e.g., because of the link to Gaussian MLE):

$$L(y^{(i)}, f(\mathbf{x}^{(i)}|\boldsymbol{\beta})) = L(y^{(i)}, \mathbf{x}^{(i)\top}\boldsymbol{\beta}) = 0.5 \left(y^{(i)} - \mathbf{x}^{(i)\top}\boldsymbol{\beta}\right)^2$$

and the theoretical risk is

$$\mathcal{R}(f) = \mathcal{R}(\boldsymbol{\beta}) = \int (y - f(\mathbf{x}))^2 d\mathbb{P}_{xy} = \int (y - \mathbf{x}^\top \boldsymbol{\beta})^2 d\mathbb{P}_{xy}.$$

- (b) The Bayes regret is  $\mathcal{R}_L(\hat{f}) - \mathcal{R}_L^*$  and can be decomposed into an estimation error  $[\mathcal{R}_L(\hat{f}) - \inf_{f \in \mathcal{H}} \mathcal{R}_L(f)]$  and an approximation error  $[\inf_{f \in \mathcal{H}} \mathcal{R}_L(f) - \mathcal{R}_L^*]$ .

- (i) If  $f^* \in \mathcal{H}$ ,  $\mathcal{R}_L^* = \inf_{f \in \mathcal{H}} \mathcal{R}_L(f)$ , i.e., the approximation error is 0 and for  $n \rightarrow \infty$  our Bayes regret  $\rightarrow 0$ .  
(ii) If  $f^* \notin \mathcal{H}$ , the Bayes regret typically consists of both parts, but as  $n \rightarrow \infty$ , we are left with the approximation error.

- (c) • Our empirical risk is

$$\mathcal{R}_{emp}(\boldsymbol{\beta}) = 0.5 \sum_{i=1}^n \left(y^{(i)} - \mathbf{x}^{(i)\top}\boldsymbol{\beta}\right)^2 = 0.5 \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2.$$

- Optimization = minimization of the empirical risk can either be done analytically (the preferred solution in this case!) or using, e.g., gradient descent.

$$\nabla_{\boldsymbol{\beta}} \mathcal{R}(\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = -\mathbf{X}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

- (d) For convex objectives, every local minimum corresponds to a global minimum. To show convexity, calculate the second derivatives:

$$\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}^\top} \mathcal{R}(\boldsymbol{\beta}) = \mathbf{X}^\top \mathbf{X}.$$

Since  $\mathbf{z}^\top \mathbf{X}^\top \mathbf{X} \mathbf{z}$  is the inner product of a vector  $\tilde{\mathbf{z}} = \mathbf{X} \mathbf{z}$  with itself, i.e.

$$\mathbf{z}^\top \mathbf{X}^\top \mathbf{X} \mathbf{z} = \tilde{\mathbf{z}}^\top \tilde{\mathbf{z}} = \sum_{j=1}^p \tilde{z}_j^2$$

it is  $\geq 0$  and hence  $\mathbf{X}^\top \mathbf{X}$  psd and therefore  $\mathcal{R}(\boldsymbol{\beta})$  convex.