MALTA Exercises

Giovanni Neglia and Samir M. Perlaza

October 4, 2024

Exercises Week 4

Ex. 1 — Consider the regression problem over the class of linear regressors $\mathcal{H} = \{h_{\mathbf{w}} : \mathbb{R}^d \to \mathbb{R}, \mathbf{w} \in \mathbb{R}^d\}$ where $h_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^\intercal \mathbf{x}$. With the classic squared loss over a dataset $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, the empirical loss is $L_S(\mathbf{w}) = \frac{1}{m}\sum_{i=1}^m (\mathbf{w}^\intercal \mathbf{x}_i - y_i)^2$. We want to consider a different empirical loss by adding an additional *regularization term*. In particular, we consider $\tilde{L}_S(\mathbf{w}) = \frac{1}{m}\sum_{i=1}^m (\mathbf{w}^\intercal \mathbf{x}_i - y_i)^2 + \lambda\|\mathbf{w}\|^2$, where $\lambda > 0$.

1. Is the minimization of $\tilde{L}_S(\mathbf{w})$ a convex optimization problem?
2. Show that minimizing $\tilde{L}_S(\mathbf{w})$ leads to solving a linear system.
3. How many solutions has this linear system?

Ex. 2 — Consider the regression problem over the class of linear regressors $\mathcal{H} = \{h_w : \mathbb{R} \to \mathbb{R}, w \in \mathbb{R}\}$ where $h_w(x) = wx$ with the loss $\ell(w, (x, y)) = |wx - y|$. We want to study the ERM predictor.

1. Given a dataset $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, the empirical risk is $L_S(\mathbf{w}) = \frac{1}{m}\sum_{i=1}^m |\mathbf{w}^\intercal \mathbf{x}_i - y_i|$. Is the empirical risk minization a convex optimization problem?
2. Given a sample (x_i, y_i), for which values of w is the loss $\ell(w, (x_i, y_i))$ differentiable with respect to w? What about the empirical loss over the whole dataset $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$, i.e., $L_S(\mathbf{w}) = \frac{1}{m}\sum_{i=1}^m |\mathbf{w}^\intercal \mathbf{x}_i - y_i|$?
3. Can you propose an algorithm to find an ERM regressor $w^* \in \arg\min_{w\in\mathbb{R}} L_S(w)$? Hint: it should be $\frac{\mathrm{d}L_S(w)}{\mathrm{d}w} \leq 0$ for $w \leq w^*$; and $\frac{\mathrm{d}L_S(w)}{\mathrm{d}w} \geq 0$ for $w \geq w^*$.

4. What is the time-complexity of the proposed algorithm?