Queen Mary University of London

MTH786, Semester A, 2023/24

Coursework 4 **N. Perra**

Problem 1. For this exercise we consider ridge regression problems of the form

$$\mathbf{w}_\alpha = \arg\min_{\mathbf{w}\in\mathbb{R}^{d+1}} \left\{ \frac{1}{2s}\|\mathbf{X}\mathbf{w}-\mathbf{y}\|^2 + \frac{\alpha}{2}\|\mathbf{w}\|^2 \right\}, \tag{1}$$

for data $\mathbf{y}\in\mathbb{R}^s$, a data matrix $\mathbf{X}\in\mathbb{R}^{s\times(d+1)}$ and a regularisation parameter $\alpha > 0$.

1. Compute the ridge regression solution for the original data samples given in the previous question on unstable regression problems. I.e., consider data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.
2. Consider validation data of the form $\left(x^{(4)}, y^{(4)}\right)$, where $x^{(4)} = 2$, $y^{(4)} = 1$. Choose the regularisation parameter α such that it solves the minimisation problem for validation error, i.e.

$$\hat{\alpha} = \arg\min_{\alpha\geq 0} \left\{ \left|\hat{w_0} + \hat{w_0}x^{(4)} - y^{(4)}\right|^2 \right\}.$$

3. Repeat the same exercise for the perturbed data samples, i.e. $\mathbf{y}_\delta$ that reads $y_\delta^{(1)} = 2+\varepsilon$, $y_\delta^{(2)} = 2+\varepsilon$ and $y_\delta^{(3)} = 2-\varepsilon$.

Problem 2. For this exercise we consider ridge regression problems of the form

$$w_\alpha = \arg\min_{w\in\mathbb{R}^{d+1}} \left\{ \frac{1}{2}\|Xw-y\|^2 + \frac{\alpha}{2}\|w\|^2 \right\}, \tag{2}$$

for data $y\in\mathbb{R}^s$, a data matrix $X\in\mathbb{R}^{s\times(d+1)}$ and a regularisation parameter $\alpha > 0$.

1. Calculate the gradient of the energy function $E\left(w\right) = \frac{1}{2}\left\|Xw-y\right\|^2 + \frac{\alpha}{2}\left\|w\right\|^2$.
2. Prove that $E\left(w\right)$ is a convex and bounded from below function.
3. Combine the above results to conclude that there is a unique solution w_α of the minimisation problem (2) which also solves the normal equation

$$\left(X^\top X + \alpha I\right) w_\alpha = X^\top y.$$

4. Continuously differentiable function $f:\mathbb{R}^{d+1}\to\mathbb{R}$ is called L-smooth if

$$\left\|\nabla f\left(u\right) - \nabla f\left(v\right)\right\| \leq L\left\|u-v\right\|,$$

   for any vectors $u, v\in\mathbb{R}^{d+1}$. Prove that the energy function E is L-smooth for some value of L. Try to identify the smallest possible such a value L.

Problem 3. Suppose we are given s data vectors $\{\mathbf{y}^{(j)}\}_{j=1}^{s}$ where each $\mathbf{y}^{(j)} \in \mathbb{R}^n$ is of the form $\mathbf{y}^{(j)} = \mathbf{X}\mathbf{w}^\dagger + \varepsilon^{(j)}$, for weights $\mathbf{w}^\dagger$ and outcomes $\varepsilon^{(1)}, \ldots, \varepsilon^{(s)}$ of a random vectors ε with its expectation being zero, i.e. $\mathbb{E}\left[\varepsilon^{(\mathbf{j})}\right] = 0$.

1. Show that the expected value of a random variable is linear.
2. Show that the expected value of a constant value is that constant value itself.
3. Compute the expectation of $\hat{\mathbf{w}}^{(j)}$, where $\hat{\mathbf{w}}^{(j)}$ are the solutions of $\mathbf{X}^\top\mathbf{X}\hat{\mathbf{w}}^{(j)} = \mathbf{X}^\top\mathbf{y}^{(j)}$.
4. Compute the expectation of $\hat{\mathbf{w}}_\alpha^{(j)}$, where $\hat{\mathbf{w}}_\alpha^{(j)}$ are the solutions of (2) for data $\mathbf{y}^{(j)}$.
5. Can the two previous expectations match for any value other than $\alpha = 0$?

Problem 4. This week marks the first coding exercise you are supposed to work on. Working on this assignment is not mandatory, but **highly advisable**. This would allow you to practice the submission algorithm we will use during the final project submission.