Assignment 3

Statistical Machine Learning

Moritz Haas / Ulrike von Luxburg

Summer term 2022 — due on **May 9th at 12:00.**

Exercise 1 (Linear Regression, 1 + 1 + 0.5 + 0.5 + 0.5 + 2 + 1.5 = 7 points)
Let $X \in \mathbb{R}^{n\times d}$ be a data matrix. In this exercise, we are going to prove that the least-squares optimization problem

$$\min_{w\in\mathbb{R}^d} \| Y - Xw \|_2^2 \tag{1}$$

is convex (compare Proposition 5 in the lecture slides). To warm up, we recap multidimensional derivatives.

(a) Let $f(X) = a^T X$, where $X \in \mathbb{R}^3$ is a column vector, and $a^T = [2, -1, 5]$. Compute

$$\frac{\partial f}{\partial X}$$

(b) Let $f(X) = X^T A X$, where $X \in \mathbb{R}^2$ is a column vector of two elements and

$$A = \begin{pmatrix} 1 & 2 \\ 5 & 3 \end{pmatrix}$$

compute

$$\frac{\partial f}{\partial X}$$

(c) Argue that the optimization problem (1) is defined over a convex domain.

(d) Does the optimization problem (1) have any constraints? If not, does this mean that this is an unconstrained optimization problem?

(e) Is the objective function in (1) differentiable? If yes, how many times?

(f) Show that $X^T X$ is positive semi-definite and that $\text{rank}(X^T X) = \text{rank}(X)$.

(g) Show that the optimization problem (1) is convex.
Hint: What is left to show? Use an appropriate criterion (see lecture slides).

Exercise 2 (Linear Regression by Hand, 2.5+1.5 = 4 points)

(a) Let

$$\begin{array}{llll} X_1 = (1,1,0), & X_2 = (1,-1,2), & X_3 = (2,3,-1), & X_4 = (-1,2,-3), \\ Y_1 = 3, & Y_2 = 1, & Y_3 = 7, & Y_4 = 0. \end{array}$$

Prove that the linear regression problem

$$\min_{\omega} \|Y - X\omega\|^2$$

does not have an unique solution and find all ω that solve this problem.

(b) Predict $\hat{Y}$ for $X_5 = (0, 2, -2)$ and $X_6 = (1, 0, 0)$ using two different optimal ω from above. Can you explain why the prediction matches in one case but not in the other?

Exercise 3 (Linear, Ridge, and Lasso in Python, 1+3+2+2+2 = 10 points)

In this exercise you will implement linear regression, ridge regression and the lasso. If you do not manage to solve part (b), you may continue to write the code for subsequent parts as if you had functioning ridge and lasso regression.

(a) Let

$$x \sim \text{Unif}([0,2))$$
$$y(x) \sim 2\sin 2x + 0.1 * \varepsilon$$

where $\varepsilon \sim \mathcal{N}(0,2)$. In the notebook you will find the code that samples $n = 100$ points from this distribution and saves them in $(\texttt{xs}, \texttt{ys})$. Do a scatter plot of the sampled points.

(b) Given n D-dimensional samples `xs` with 1-dimensional `ys`. Write two functions

```
def ridge_regression(xs, ys, lam=1): ...
```

and

```
def lasso_regression(xs, ys, lam=1): ...
```

that, given an $n \times D$ matrix `xs` and an $n \times 1$ vector `ys`, return the ridge regression resp. lasso weights ω as a $D \times 1$ vector. Make sure you output arrays of the correct shape.

(c) For $\lambda \in \{0.1, 1, 10\}$, compute the ridge regression for $(\texttt{xs}, \texttt{ys})$ as in a) and compare the mean squared error (MSE). In three different plots, plot the predictions for the different λ. The plot should contain the scatter plot of the points and the predicted line. Repeat the MSE calculations and the plotting for lasso regression.

(d) Compute the weights for all datasets in the provided list using linear, ridge ($\lambda = 1$) and lasso ($\lambda = 1$) regression. Scatter plot the absolute weight vectors, such that the weight dimensions correspond to the x- and y-axis. Use logarithmic axis scale and appropriate limits.

(e) Extend the design matrix `xs` from (a) by adding the features $f_1(x) = 1$ and $f_2(x) = x^2$ for each data point x. Then perform ridge regression on the new design matrix of shape $n \times 3$. Use $\lambda \in \{0.001, 0.01, 0.1, 1, 10\}$ and plot the MSE as a function of λ. Plot the best prediction. Set $\lambda = 0$, to see what happens with the linear regression.

Exercise 4 (Feedback Poll, 2 bonus points)
Every week there will be a feedback survey on Ilias about the lecture, tutorials and exercise sheets. If you participate, it helps us in understanding what we can improve.