General Regulations.

- Please hand in your solutions in groups of two (preferably from the same tutorial group). Submissions by a single person alone will not be corrected.

- Your solutions to theoretical exercises can be either handwritten notes (scanned), or typeset using LaTeX. For scanned handwritten notes please make sure that they are legible and not too blurry.

- For the practical exercises, the data and a skeleton for your jupyter notebook are available at https://github.com/sciai-lab/mlph_w24. Always provide the (commented) code as well as the output, and don't forget to explain/interpret the latter. Please hand in your notebook (`.ipynb`), as well as an exported pdf-version of it.

- Submit all your files in the Übungsgruppenverwaltung, only once for your group of two. Specify all names of your group in the submission.

1 Regularization and Intercept

Consider a regression problem with two explanatory variables x_1, x_2. As introduced in the lecture, the intercept β_0 can be incorporate by considering $y = \boldsymbol{\beta}^T \mathbf{x}$ with $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ and $\mathbf{x} = (1, x_1, x_2)^T$.

(a) In this setting, write down the loss function for ridge regression, penalizing the L^2-norm of $\boldsymbol{\beta}$, in components. What is the influence of the regularization strength on the intercept β_0? (1 pt)

(b) Oftentimes, a regularization of the intercept term is unwanted. How would you modify the loss function to account for this? (1 pt)

(c) Which shapes in $\mathbb{R}^3$ do the regularization contours (i.e. sets of parameters with equal regularization penalty) of versions (a) and (b) have? (1 pt)

2 σ^2 Estimation

(a) Maximum Likelihood. Focusing on a single point $(y_n, \mathbf{x}_n)$ our linear regression model simplifies to

$$y_n = \boldsymbol{\beta}^T \mathbf{x}_n + \varepsilon_n. \tag{1}$$

If we assume that ε_n follows a normal distribution with zero mean and known variance σ^2, that is $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$, this is equivalent to the assumption that $y_n \sim \mathcal{N}(\boldsymbol{\beta}^T \mathbf{x}_n, \sigma^2)$. The logarithm of $p(y_n | \boldsymbol{\beta}, \sigma^2)$ is known as the log-likelihood. Having observed N data points this formulation generalizes to a sum of log-likelihoods and we can learn $\boldsymbol{\beta}$ by maximizing the logarithm of

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{n=1}^{N} \log \mathcal{N}\left(y_n | \boldsymbol{\beta}^T \mathbf{x}, \sigma^2\right). \tag{2}$$

Show that we are solving the same objective as in the SSQ formulation (just with a different scaling factor) and get the same solution for $\boldsymbol{\beta}$. This formulation is known as the *Maximum Likelihood* approach, as we learn the parameters that maximize the likelihood of the data. (2 pts)

(b) **Estimation of σ^2.** Estimating σ^2 then analogously amounts to finding the $\hat{\sigma}^2$ that maximizes this log-likelihood given the estimates $\hat{\boldsymbol{\beta}}$, i.e.

$$\hat{\sigma}^2 = \arg\max_{\sigma^2} \sum_{n=1}^{N} \log \mathcal{N}\left(y_n | \hat{\boldsymbol{\beta}}^T \mathbf{x}, \sigma^2\right). \tag{3}$$

Solve this and relate the result to the SSQ residual formulation from the lecture.

(2 pts)

3 Visualize Regularization Contours

For two dimensional parameter vectors β we can visualize the error/loss surface of linear regression using contour plots. In this exercise you will create a set of such plots in order to familiarize yourself further with the influence of regularization. You can visualize the contours for example via `plt.contour` or `plt.contourf`.[1]

(a) Plot the Ridge regression regularization term as well as the Lasso[2] regularization term for $\beta_1, \beta_2 \in [-1, 3]$.
(2 pts)

(b) For the data set `linreg.npz` plot the sum of squares (SSQ) of a linear regression as a function of $\boldsymbol{\beta}$ over the same range as in **i)**, i.e. over the grid $[-1, 3] \times [-1, 3]$.
(2 pts)

(c) Plot the ridge and Lasso loss functions, i.e. $\mathrm{SSQ}(\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_2^2$ and $\mathrm{SSQ}(\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_1$, for $\lambda \in \{0, 10, 50, 100, 200, 300\}$ in the same $\boldsymbol{\beta}$ grid as before and *discuss your observations!*
(2 pts)

4 CT Reconstruction

One application of linear regression is the reconstruction of CT-Scans. In this task, you will do this on simulated data in the 2D case. You are given a sinogram $Y \in \mathbb{R}^{ar}$, a matrix where each row corresponds to a (1D) projection of the image consisting of r detector readouts along one of a distinct, evenly spaced angles. Additionally, you are given the design matrix $\mathbf{X} \in \mathbb{R}^{p \times ar}$. Excluding noise, one has $Y = I\mathbf{X}$, with the image $I \in \mathbb{R}^p$ which should be reconstructed.

(a) What is the interpretation of a column of $\mathbf{X}$? Visualize a choice of four columns as images. (2 pts)

(b) What is the interpretation of a row of $\mathbf{X}$? Visualize a choice of four rows as images. (2 pts)

(c) Solve the reconstruction problem with linear regression without any regularization and with ridge regression. What do you observe?

(3 pts)

5 Bonus: X-Ray Free-Electron Lasers

Imagine the reconstruction problem from task 4, but without the knowledge about which detections correspond to which orientations of the sample. This scenario actually happens in the analysis of X-ray free-electron laser data: The laser is aimed at a sample (a tiny protein crystal), and the diffraction pattern of a short high-energy pulse is recorded. This is repeated many times, but never is the orientation of the sample known. Nevertheless, it's possible to reconstruct the structure of proteins from such data[3].

How would you approach this problem? Try to reconstruct the image from the data of task 4, after shuffling the sinogram along the angle axis. (5 pts)

---

[1] See https://matplotlib.org/stable/gallery/images_contours_and_fields/contour_demo.html for an example.

[2] The abbreviation comes from *least absolute shrinkage and selection operator*.

[3] There are more complications in the analysis of diffraction patterns, such as solving the phase problem. Here, we merely focus on the reshuffling problem, reusing the data from the previous exercise.