MAD 2024-25, Assignment 2

Bulat Ibragimov, Sune Darkner

hand in until: 2.12.2024 at 9:00

General comments: The assignments in MAD must be completed and written individually. You are allowed (and encouraged) to discuss the exercises in small groups. If you do so, you are required to list your group partners in the submission. The report must be written completely by yourself. The data needed for the assignment can be found in the assignment folder that you download from Absalon.

Submission instructions: Submit your report as a PDF, not zipped up with the rest. Please add your source code to the submission, both as executable files and as part of your report in appendix. To include it in your report, you can use the `lstlisting` environment in LaTeX, or you can include a "print to pdf" output in your pdf report. In some exercises we will ask you to include a code snippet as part of your solution text - a code snippet is only the most essential lines of code needed for solving the problem, this does not include import statements, other forms of boiler plate code, as well as plotting code.

Exercise 1 (**Weighted Average Loss, 4 points, based on Exercise 1.11 in Rogers & Girolami).** The following expression is known as the **weighted** average loss:

$$\mathcal{L} = \frac{1}{N}\sum_{n=1}^{N}\alpha_n\left(f(\mathbf{x}_n;\mathbf{w}) - t_n\right)^2 = \frac{1}{N}\sum_{n=1}^{N}\alpha_n\left(\mathbf{w}^T\mathbf{x}_n - t_n\right)^2$$

where the influence of each data point is controlled by its associated weight $\alpha_n > 0$ parameter.

a) (2 points): Assuming that each α_n is given a fixed value, show by mathematical derivation that the optimal least squares parameter value is $\hat{\mathbf{w}} = \left(\mathbf{X}^T\mathbf{A}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{A}\mathbf{t}$, where $\mathbf{A}$ is a diagonal matrix that contains the weights $\alpha_1, \ldots, \alpha_N$ on the diagonal. Hint: Start by rephrasing the loss function in matrix-vector form (as done in Rogers & Girolami Sec. 1.3) using the weight matrix $\mathbf{A}$, then compute the gradient of the loss function with respect to the parameters $\mathbf{w}$.

b) (2 points): Similar to Exercises 3 and 4 of Assignment 1, implement a corresponding regression model in Python (e.g., implement a new `linweighreg.py` module by making a copy of `linreg.py` and adding the code you find relevant). Afterwards, fit a model on all the features in the training set `boston_train.csv` using $\alpha_n = t_n^2$. Compute corresponding predictions for the test instances given in `boston_test.csv` and generate a scatter plot as in Exercises 3 and 4 of Assignment 1. What do you expect to happen? What do you observe? Do the additional weights have an influence on the outcome?

Deliverables. a) Derivation for the optimal solution (similarly to the lecture). Note that you do not have to show that the computed solution is a global minimum; just providing the solution and the corresponding derivation is enough. (b) The scatter plot as well as short answers to the questions raised (2-3 lines each). Add your source code to your submission.

Exercise 2 (**Polynomial Fitting with Regularized Linear Regression and Cross-Validation, 4 points**). In this exercise, you will apply leave-one-out cross-validation to study the influence of the regularization parameter λ on the predictive performance of regularized linear regression with penalty term $\lambda\mathbf{w}^T\mathbf{w}$. In the `men-olympic-100.txt` file, you will find data on the men's Olympic 100m running times (Hint: You can read the file into a numpy array using this numpy function, `raw = np.genfromtxt('men-olympics-100.txt', delimiter=' ')`). Take the first place running times (second column of the `raw` table) as target values $t_1, \ldots, t_{27}$. The input variables $x_1, \ldots, x_{27}$ are given in the first column (of the `raw` table).

a) (3 points): Apply polynomial fitting with regularized linear regression to fit a **first order polynomial** to the data. Plot the leave-one-out cross- validation error (y-axis) as a function of λ for some values $\lambda \in [0, 1]$ (x-axis); use `numpy.logspace(-8, 0, 100, base=10)` to generate the λ values to be tested. Report the best value of λ and the regression coefficients $\mathbf{w}$ corresponding to both $\lambda = 0$ (no regularization) and the best value of λ.

Hint: Be aware of the fact that the values outputted via `print` *are usually rounded. Make use of a higher precision when printing, e.g., via* `print("lam=%.10f and loss=%.10f" % (lam, loss))`.

b) (1 point): Repeat the same for fitting a **fourth order polynomial** to the data.

Deliverables. a) The plot, best value of λ, and the two sets of regression coefficients; b) the plot, best value of λ, and the two sets of regression coefficients. Add your source code to your submission.

Exercise 3 (Pdf and cdf, 4 points). We model the life span x of a chip (in years) with a distribution that has the following cumulative distribution function (cdf).

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - exp(-\beta x^{\alpha}) & x > 0 \end{cases}$$

with $\alpha > 0$ and $\beta > 0$ being parameters.

a) (1 point): Determine the probability density function (pdf) of the distribution.
b) (2 points): Suppose we fix the parameters to $\alpha = 2$ and $\beta = \frac{1}{4}$. What is the probability that the chip works longer than four years? What is the probability that the chip stops working in the time interval [5; 10] years?
c) (1 point): How large is the median of a life span (for general choices of α, β)?

Deliverables. a) Besides the pdf also Include the derivation steps needed to go from the cdf to the pdf, b) besides the results also include the steps showing how you compute the probabilities, c) include both result and derivation steps.

Exercise 4 (Conditional probability and expectations, 4 points). Professor James Duane of Regent University, USA, strongly advises exercising the right to remain silent and immediately contacting a lawyer when questioned by police on any matter. His views are presented in a popular YouTube lecture titled "Don?t Talk to the Police" and a book called "You Have the Right to Remain Innocent." For this assignment, we will use a probabilistic model to evaluate his claims. Consider the following scenario and assumptions:

We categorize individuals as either: NC: An innocent person with no history of convictions C: An innocent person with a history of convictions

Assumptions:

1) The risk that a case goes to court if a person (either NC or C) remains silent is 0.001.
2) The risk that a case goes to court if a person NC talks to the police (due to the police using their words against them) is 0.0015.
3) The risk that a case goes to court if a person C talks to the police (due to the police using their words against them) is 0.005.
4) If a case goes to court, the probability of acquittal is 0.8 for NC and 0.2 for C if the person talked to the police.
5) The probability of acquittal drops to 0.2 for NC and 0.05 for C if the person remained silent during the investigation, as it may be perceived by the jury that the person has something to hide.
6) If a person is convicted but had talked to the police during the investigation, their sentence is reduced by 50% due to perceived cooperation.

Scenario: The police have arrested a person suspected of involvement in a crime punishable by a 5-year prison sentence:

a) (2 points): A person has no history of convictions (NC). He is considering whether he should remain silent or answer the police questions. Please calculate the outcomes of both actions in terms of the average number of days in prison.
b) (2 points): A person has a history of convictions (C). He is considering whether he should remain silent or answer the police questions. Please calculate the outcomes of both actions in terms of the average number of days in prison.

Deliverables. For both questions include your answer and derivation steps.