MAD 2021-22, Assignment 2

Bulat Ibragimov, Kim Steenstrup Pedersen

hand in until: 06.12.2021 at 23:59

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. In order to pass the assignment, you will need to get at least 40% of the available points. 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:

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

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 $\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{A}\mathbf{t}$, 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 from Regent University, USA advocates for constant execution of the right to remain silent and immediately calling the lawyer every time a person is questioned by the police on any topic. He has a popular YouTube lecture "Don't talk to the Police" and the book "You Have the Right to Remain Innocent". Let's try to probabilistically evaluate his claims using the following model.

We have individuals with no history of convictions (person NC) and with a history of convictions (person C). Suppose:

1) If any person (either NC or C) remains silent, there is a 0.001 risk of his case ending up in the court.
2) If a person NC talks to the police, there is a 0.002 risk of the police using his words against him and his case ending up in the court.
3) If a person C talks to the police, there is a 0.005 risk of the police using his words against him and his case ending up in the court.
4) If case goes to court, the probability of not being convicted are 0.5 for person NC and 0.1 for person C.
5) The probability of not being convicted drops 4 times if a person did not talk to the police during the investigation, as the jury thinks he has something to hide.
6) If a person is convicted and talked to the police during the investigation, his sentence is reduced by a 0.75 factor as he is considered cooperative.

Probability theory and expectations can be used to make rational choices by computing expected result from different actions.

Peter has no prior convictions (NC) and is arrested by the police. The expected sentence if convicted is 5 years. His goal is to spend as little time in prison as possible, so the question is: Which action should he take?

When speaking, he does not know if he will end up in prison or not, so he can make a random variable X_{speak} modeling the sentence duration. It can take on two values: Either the duration of time in prison if he gets convicted, or 0 if he does not end up in prison.

a) Compute Peter's expected sentence duration when speaking, by computing the expectation of X_{speak}.

b) Model the sentence duration if Peter stays silent by the r.v. X_{silent} and compute the expected sentence duration if taking that action. Which action is the most rational for Peter to take, given his goal?

c) Brian has prior convictions. The expected sentence if convicted is also 5 years. Thus for each the two choices Brian can make we can likewise define Y_{speak} and Y_{silent} to model sentence durations. Compute the two expectations for Y_{speak} and Y_{silent}. Which action is most rational for Brian, if he also has the goal of spending as little time in prison as possible?

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