MAD 2021-22, Assignment 4

Bulat Ibragimov, Kim Steenstrup Pedersen

hand in until: 20.12.2021 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.

Coin Game

Exercise 1 (**Based on Exercises 3.1, 3.2 and 3.3 in the book, 4 points).** Consider the coin game discussed in the book and mentioned in lecture L7. You will compute the posterior distribution $p(r|y_N)$ for three different priors:

a) For $\alpha = \beta = 1$, the beta distribution becomes uniform between 0 and 1. In particular, if the probability of a coin landing heads is given by r and a beta prior is placed over r, with parameters $\alpha = 1 = \beta$, then this prior can be written as:

$$p(r) = 1 \quad (0 \leq r \leq 1).$$

Using this prior, compute the posterior density for r if y heads are observed in N tosses (i.e. multiply this prior by th ebinomial likelihood and manipulate the result to obtain something that looks like a beta density).

b) Repeat exercise a) for the following prior, also a particular form of the beta density:

$$p(r) = \begin{cases} 2r & 0 \leq r \leq 1 \\ 0 & o.w. \end{cases}$$

What are the values of the prior parameters α and β that result in $p(r) = 2r$?

c) Repeat exercise a) for the following prior, again a particular form of the beta density:

$$p(r) = \begin{cases} 3r^2 & 0 \leq r \leq 1 \\ 0 & o.w. \end{cases}$$

What are the prior parameters here?

Deliverables. a) The posterior expression, b) the posterior expression and the parameter values, c) the posterior expression and the parameter values.

Probabilistic regression

Exercise 2 (**Bayesian Regression, 4 points**). In this exercise, we will revisit linear regression from a Bayesian perspective. You will imitate the Bayesian regression from Lecture L7, but apply it to the Olympic 100m dataset used in the book (found in the file `men-olympics-100.txt`). Here, you should use the years (found in the first column) as your input x_n and the winner times (second column) as your output t_n.

a) Assume that the noise in your model is normally distributed with zero mean and variance $\sigma^2 = 10$. What is the likelihood $p(\mathbf{t}|\mathbf{w}, \mathbf{X}, \sigma^2)$ of observing the given output $\mathbf{t}$ given the model defined by $\mathbf{w}$ when the input data matrix is $\mathbf{X}$?

b) Assume that the prior distribution for your model parameters $\mathbf{w}$ is normally distributed $p(\mathbf{w}) = \mathcal{N}(\mu_0, \mathbf{\Sigma}_0)$, and let the likelihood $p(\mathbf{t}|\mathbf{w}, \mathbf{X}, \sigma^2)$ of observing the data matrix $\mathbf{X}$ given model parameters $\mathbf{w}$ be given by another normal distribution $p(\mathbf{t}|\mathbf{w}, \mathbf{X}, \sigma^2) = \mathcal{N}(\mu_l, \mathbf{\Sigma}_l)$. What is the corresponding posterior distribution $p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \sigma^2)$?

c) Set your prior parameters to be $\mu_0 = [0, 0]^T$ and $\mathbf{\Sigma}_0 = \begin{bmatrix} 100 & 0 \\ 0 & 5 \end{bmatrix}$. Assume that your prior and likelihood are both normally distributed as in b). Implement a function that computes the corresponding posterior probability density.

d) What is the mean and covariance of the posterior probability density after seeing the entire dataset? Hint: See `linreg.py` from assignment 1 exercise 4 for reading the data file.

Deliverables. a) A formula, b) a formula, c) your source code, d) the computed mean and covariance.

Principal Components Analysis

Exercise 3 (Implement PCA, 8 points). See the Jupyter notebook file `pca_StudentVersion.ipynb` for the detailed questions and hints.

a) Implement PCA on the diatoms database. Please output the proportion of variance explained by each of the first 10 components (5 points)

b) Visualize fourth component of the PCA (3 points)