Machine Learning 2016/2017, Assignment 1:

Yevgeny Seldin, Christian Igel
Department of Computer Science, University of Copenhagen

The deadline for this assignment is **12:00 pm (noon, not midnight) 28/11/2016**. You must submit your individual solution electronically via the Absalon home page.

A solution consists of:

- A PDF file with detailed answers to the questions, which may include graphs and tables if needed. Do *not* include your source code in this PDF file.
- Your solution source code (Matlab / R / Python scripts or C / C++ / Java code) with comments about the major steps involved in each question (see below). Source code must be submitted in the original file format, not as PDF.
- Your code should be structured such that there is one main file that we can run to reproduce all the results presented in your report. This main file can, if you like, call other files with functions, classes, etc.
- Your code should also include a README text file describing how to compile and run your program, as well as list of all relevant libraries needed for compiling or using your code.

Math and ML

You need basic linear algebra and calculus for understanding machine learning. To recall some of your math knowledge, answer the following questions. Do the calculations by hand.

If you feel unsure about what a question means and how to answer it, this indicates that you are not fully comfortable with mathematical skills that are assumed in this course. No worries. In this case, just take your time and to go back to your notes from school or your first study years – or grab one of the numerous textbooks that are around.

Feel free to ask questions about the mathematical background in the exercise classes!

1 Vectors and Matrices

Consider the two vectors

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \text{ and } \boldsymbol{b} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

and the matrix

$$\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix} .$$

Question 1.1. Calculate the *inner product* (also known as *scalar product* or *dot product*) denoted by $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$, $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$, or $\boldsymbol{a} \cdot \boldsymbol{b}$.

Question 1.2. Calculate the length (also known as Euclidean norm) $\|\boldsymbol{a}\|$ of the vector $\boldsymbol{a}$.

Question 1.3. Calculate the *outer product* $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$. Is it equal to the inner product $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$ you computed in Question 1.1?

Question 1.4. Calculate $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{a}$. Is it equal to $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$? (Test yourself: the answer to one of the questions 1.3 and 1.4 should be “yes” and to the other “no”.)

Question 1.5. Calculate the inverse of matrix $\boldsymbol{M}$, denoted by $\boldsymbol{M}^{-1}$. We remind that you should get that $\boldsymbol{M}\boldsymbol{M}^{-1} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix.

Question 1.6. Calculate the matrix-vector product $\boldsymbol{M}\boldsymbol{a}$.

Question 1.7. Let $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$. Calculate the transpose of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^{\mathrm{T}}$. Is $\boldsymbol{A}$ symmetric? (A matrix is called symmetric if $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$.)

Question 1.8. What is the rank of $\boldsymbol{A}$? (The rank is the number of linearly independent columns.) Give a short explanation.

Question 1.9. What should be the relation between the number of columns and the rank of a square matrix in order for it to be invertible? Is $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$ invertible?

2 Derivatives

We denote the derivative of a univariate function $f(x)$ with respect to the variable x by $\frac{df(x)}{dx}$. We denote the partial derivative of a multivariate function $f(x_1, \ldots, x_n)$ with respect to the variable x_i, where $1 \leq i \leq n$, by $\frac{\partial f(x_1,\ldots,x_n)}{\partial x_i}$. The partial derivative $\frac{\partial f(x_1,\ldots,x_n)}{\partial x_i}$ is the derivative of $f(x_1, \ldots, x_n)$ with respect to x_i when we treat all other variables x_j for $j \neq i$ in f as constants.

Please recall the basic rules for derivatives, namely the sum rule, the chain rule, and the product rule, see the lecture notes available on Absalon.

Question 2.1. What is the derivative of $f(x) = \frac{1}{1+\exp(-x)}$ with respect to x?

Question 2.2. What is the partial derivative of $f(w, x) = 2(wx+5)^2$ with respect to w?

3 Probability Theory: Sample Space

An urn contains five red, three orange, and one blue ball. Two balls are randomly selected (without replacement).

1. What is the sample space of this experiment?
2. What is the probability of each point in the sample space?
3. Let X represent the number of orange balls selected. What are the possible values of X?
4. Calculate $\mathbb{P}\{X = 0\}$.
5. Calculate $\mathbb{E}[X]$.

4 Probability Theory: Properties of Expectation

Let X and Y be two discrete random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Starting from the definitions, prove the following identities:

1. $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
2. If X and Y are independent then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. (Mark the step where you are using the independence assumption. Note that this assumption was not required in point 1.)

3. Provide an example of two random variables X and Y for which $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$. (Describe how you define the random variables, provide a joint probability distribution table [see comment below], and calculate $\mathbb{E}[XY]$ and $\mathbb{E}[X]\mathbb{E}[Y]$.)

4. $\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$.

5. Variance of a random variable is defined as $\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$. Show that $\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Comment: A convenient way to represent a joint probability distribution of two discrete random variables is a table. For example, if X and Y are Bernoulli random variables with bias $\frac{1}{2}$ (fair coins) then the joint distribution table looks like this:

$X \backslash Y$	0	1
0	1/4	1/4
1	1/4	1/4

. And if Z_1 and Z_2 are Bernoulli random variables and we define $X = Z_1 + Z_2$ and $Y = Z_1 \times Z_2$ then the joint distribution of X and Y is:

$X \backslash Y$	0	1
0	1/4	0
1	1/2	0
2	0	1/4

.

5 Probability Theory: Complements of Events

1. The complement of event A is denoted by $\bar{A}$ and defined by $\bar{A} = \Omega \setminus A$. Starting from probability axioms prove that $\mathbb{P}\{A\} = 1 - \mathbb{P}\{\bar{A}\}$.

2. In many cases it is easier to calculate the probability of a complement of an event than to calculate the probability of the event itself. Use this to solve the following question. We flip a fair coin 10 times.

   - What is the probability to observe at least one tail?
   - What is the probability to observe at least two tails?

6 Probability Theory: Coin Flips

We flip a fair coin ten times. Find the probability of the following events:

1. The number of heads and the number of tails are equal.

2. There are more heads than tails.

3. The i-th flip and the $(11 - i)$-th flip are the same for $i = 1, \ldots, 5$.