

# Machine Learning - Homework 2

Spring 2018

## Exercise 1 (2 pts)

Compute the mean and the sample standard deviation of  $x_1, x_2, x_3$ , respectively, and normalize the following data (you can do this by hand or by using R):

	$D_1$	$D_2$	$D_3$	$D_4$
$x_1$	12	21	17	10
$x_2$	-2	-1	1.2	0
$x_3$	-1	2	-2	1

## Exercise 2 (6 pts)

For the solution of this exercise it is enough to write down the overall algorithm, incorporating all of the following points. Please store your algorithm in the .R or .Rdata format and send it to [anna-lena.horlemann@unisg.ch](mailto:anna-lena.horlemann@unisg.ch).

1. Write your own perceptron algorithm in R for the following data samples (with attributes  $x_1, x_2$  and target  $y$ ):

	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$
$x_1$	2	3	1	3	5	-1
$x_2$	3	-1	0	1	3	-1
$y$	1	-1	-1	1	1	-1

The algorithm should go over each sample once.

2. Include a normalization of the data at the beginning of your algorithm (you can do this step by step or look for some inbuilt R functions).
3. Add an outer loop such that you go over all data samples several times.
4. Plot the data points and the decision boundary that you get at the end of the algorithm.

## Exercise 4 (2 pts)

1. Adapt your algorithm to an ADALINE algorithm. (For the solution just write down the lines of your perceptron algorithm that need to be changed.)
2. Find the minimum number of outer loops that you need in your original Perceptron and the ADALINE algorithm to get a perfect classification.

**Remark:** If you use the sample or some other randomization command, please add `set.seed(1)` before it in your source code. This makes the random choices equal every time you run the algorithm. Without this the last exercise might not have a unique solution.

**Playing around with Prof. Binswanger's algorithm:** Download the algorithm of Prof. Binswanger from StudyNet and let it run. Adjust the line value of "mainDir" (quite at the beginning of the algorithm) to your working directory and store the WDBC data set in a folder called "Data" (this folder needs to be in the same directory as your working directory). You can then change the data from "iris" to "cancer" by changing the lines

```
dataLabel = " iris"
#dataLabel = "WDBC cancer"

into

#dataLabel = " iris"
dataLabel = "WDBC cancer".
```

Similarly you can activate the normalization of the data by changing the line "standardize = TRUE" into "standardize = FALSE". Right below you can also change the values of "eta" and the number of samples to use (the variable is called "nTrain").