FINAL EXAM

Authorized documents: 1 Cheat Sheet (A4, recto only, manuscript, original)

Tip: The marking-scheme is given as an indication. This exam might be a bit long: the marking-scheme will be adapted. Consequently, do not waste time on a blocking question, try to maximize your points by answering everything you can.

## Exercise 1: Running experiments   (Indicative marking-scheme: 6 points)

An experimenter tries to find out for which value of the hyper-parameter of his algorithm performing a binary classification is the best. He has a data set of 200 examples.

Question 1.1

Recall in a few lines the general protocol he has to implement to find the best value of the hyper-parameter.

Question 1.2

On his data set, he obtains the following plot:

Evolution of the train and real error
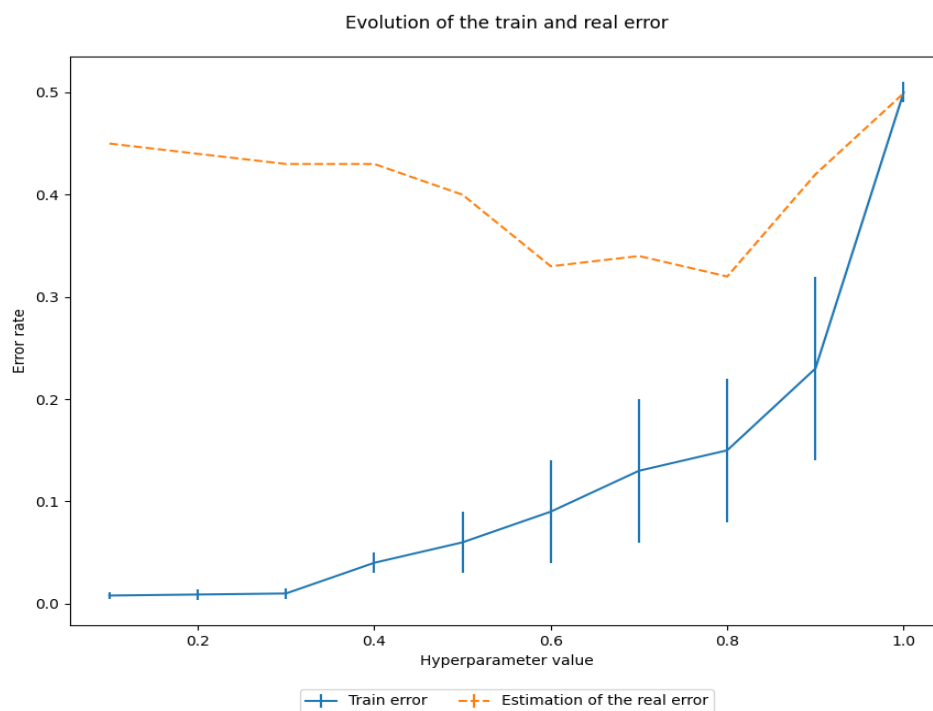
Explain the behavior of the 2 curves: why is the estimation of the real error greater? Why do they behave the same at one point? What happens when their behavior is different? Is there any other information contained in these curves?

Question 1.3

According to this experiment, which value of the hyper-parameter is the best? Justify.

Question 1.4

The experimenter having fixed the hyper-parameter of two different learning algorithms, he tries to compare them on the same data. He obtains the following confusion matrices:

Classifier 1		+	-
	+	15	5
	-	1	20

Classifier 2		+	-
	+	17	3
	-	3	18

Recall what a confusion matrix is. How do we get it?

Question 1.5

Which classifier should be chosen first? Justify.

Exercise 2: k-Nearest Neighbors (Indicative marking-scheme: 4 points)

Here are some data set:

Data	Coordinate 1	Coordinate 2	Label
x1	9	5	0
x2	2	-1	1
x3	5	0	1
x4	11	2	0
x5	9	1	0
x6	-5	7	1

Question 2.1

Compute the pairwise euclidean distances

Question 2.2

Compute the accuracy of the k-nearest neighbors with 6-fold validation for k in {1, 3}. No need to report the computations, we just want a table showing the different results.

Exercise 3: Course Questions (Indicative marking-scheme: 6 points)

Question 3.1

Recall the usual structure of a scientific paper. Describe in few sentences what each section should contain.

Question 3.2

Explain with your own words what is research in Machine Learning and Data Mining (one and a half page max). Do not forget to describe the main steps of the general research process.

Exercise 4 : LaTeX (Indicative marking-scheme: 4 points)

For this exercise, get the following page and answer on it. Do not forget to put it inside your copy.

Question 4.1

Correct the following code so it compiles.

Question 4.2

Draw the pdf outputted by the compilation (once fully debugged) on the verso of the page.

Latex file (main.tex):

```
\documentclass{article}
\usepackage[utf8]{inputenc}
\usepackage{amsmath}

\title{Exam RM}
\author{Rémi Eyraud}
\date{November 2022}

\begin{document}

maketitle

This is just a stupid exercise
But I love LaTeX so much!
\section{Latex}
\subsection{Citation}
Just a small citation \citep{ICML 16}.
\subsection{Some math}
\begin{align}
\sum{i=1}^n i &=  1 + 2 + 3 + \cdots + n \label{eq:line1}\\
& = \frac{n\cdot(n+1)}{2}
\end{align}
Equation \eqref{eq:line1} \textit{equals} Equation \eqref{eq:line2}
""" this has to be a comment """
$1^n = sqrt(2)$

\bibliographystyle{plain}
\bibliography{main}
```

The output at the compilation:

```
<*> main.tex

*** (job aborted, no legal \end found)
```

Bibtex file (main.bib):

```
@inproceedings{ICML16,
  TITLE = {{A New PAC-Bayesian Perspective on Domain Adaptation}},
  AUTHOR = {Germain, Pascal and Habrard, Amaury and Laviolette, Fran{\c c}ois and Morvant,
Emilie},
  BOOKTITLE = {{ICML 16}},
  YEAR = {2016},
}
```