Université de Rennes 1 **année 2016-2017**

M2RI **2 novembre 2016**

Module AS

Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page.

Name and First Name : _____

Duration : 2 hours
Documents and calculators allowed
All grades are indicative subject to formal evaluation.
Number of pages : 9 pages

All answers must be properly and precisely written and detailed. Any unjustified response worth 0 points. A single copy of the slides will be worth 0 points.

Exercise 1 - Learning regular and context-free "reversible" languages (8 pts)

We focus here on learning the classes of 0-reversible regular languages [Angluin 1982] and substitutable context-free languages [Clark and Eyraud 2005, 2007].

Let us recall here that a context-free languages L is substitutable iff for any string $x_1, y_1, z_1, x_2, y_2, z_2$:

$$x_1y_1z_1 \in L \wedge x_1y_2z_1 \in L \wedge x_2y_1z_2 \in L \Rightarrow x_2y_2z_2 \in L$$

while 0-reversible regular languages are the languages that can be represented by a deterministic automaton whose reverse (also called mirror) is deterministic. Following [Yoshinaka 1998], a language is 0-reversible iff for any string x_1, y_1, x_2, y_2 :

$$x_1y_1 \in L \wedge x_1y_2 \in L \wedge x_2y_1 \in L \Rightarrow x_2y_2 \in L$$

As a consequence, the class of 0-reversible regular languages is strictly included in the class of the substitutable context-free languages.

1. What are the most general hypothesis spaces for each of these classes of languages? Compare these hypothesis spaces with respect to variance and bias. (1pt)

2. Being given a positive training sample S_+, could we restrict the search space of 0-reversible regular languages? Why and how? (1pt)

3. We focus now on 0-RI algorithm (k-RI with $k = 0$ [Angluin 1982]).
   a) What would be the automaton returned by 0-RI for the sample $S_+ = \{a, aa, aaa\}$? (1pt)
   b) Does it always exist a characteristic sample for 0-RI? If the answer is positive provide a formal characterization for such a sample, otherwise provide a counter-example. (2pts)

4. a) Is the class of 0-reversible regular languages identifiable in the limit with only positive examples [Gold 67]? Justify your answer. If the answer is positive, explain how the generalization is controlled to ensure the identification. If the answer is negative, explain why the generalization can not be controlled to ensure the identification. (1pt)
   b) Answer the same questions for the class of substitutable context-free languages. (0.5pt)

5. We propose now to use the inference of 0-reversible regular languages or substitutable context-free languages to classify sequences of musical notes into 'Rock' or 'Blues' styles. Being given hundred sequences of musical notes for each style, how would you proceed to provide the best classifier? (1.5pts)

Exercise 2 - General questions (9 pts)

Several general questions (could be QCM) that can be about:

- *supervised learning methodology and evaluation (goal, different sets used, etc.),*
- *the different algorithms (decision trees, SVM, NN, Bayes, Boosting..),*
- *comparison of these algorithms*

Examples:

1. In neural networks, which parameters are we trying to learn? Check all that apply.
   - ◯ the network architecture
   - ◯ the connexion weights
   - ◯ the neuron's characteristics (e.g. the activation function)
2. Increasing the number of hidden layers in neural networks leads to two known issues: (i) vanishing gradients, and (ii) overfitting. In few words, describe these problems, explain why they raise, and what could be some solutions.
3. We have seen in the introduction to supervised machine learning, that generalization was the core of learning. Where does the generalization comes from for the inference of nominal decision trees?
4. Why k-nn classification is called lazy learning? What is learned in this case?

Exercise 3 - practical computation (3 pts)

This part is related to formula application in a given simple setup. Examples:
Given a small labelled dataset (in $\mathbb{R}^2$ or $\mathbb{R}^3$),

- *and a NN architecture with specified weights and activation function,*
- *or a SVM with specified kernel,*
- *or a (partially) built decision tree,*
- *...*

Compute the class of a given test example, or the accuracy given a test set, or the confusion matrix, etc.

For a given a dataset :

Question 1.

Let $[83, 2]$ be a new example. What is its predicted class when considering :

a) the majority rule?

b) the maximum likelihood rule?

c) $\cdots$?

For each case, specify the used rule, possibly the hypothesis made, and detail the calculus.

Question 2.

We give the outputs of a SVM classifier (kernel, lagrange multipliers) learned on the same dataset.

1. Is it a soft or hard margin SVM? Explain.
2. Write the formula to classify the new data $[83, 2]$, according to the information given in the output.

Question 3.

We give the results of Adaboost classifier still on the same dataset (hypothesis, weights).
What is the predicted class of the new data $[83, 2]$ in this case? Explain.