University Jean Monnet DSC/MLDM Master Programs

Exam of Machine Learning Fundamentals and Algorithms

Wednesday, 24th March 2021 - **(2 hours)**

2 handwritten A4 sheets allowed

Exercise 1 - Decision trees (5 points)

You are stranded on a deserted island. Mushrooms of various types grow widely all over the island, but no other food is anywhere to be found. Some of the mushrooms have been determined as poisonous and others as not (determined by your former companions' trial and error). You are the only one remaining on the island. You have the following data to consider:

Example	Not Heavy	Smelly	Spotted	Smooth	Edible
A	1	0	0	0	1
B	1	0	1	0	1
C	0	1	0	1	1
D	0	0	0	1	0
E	1	1	1	0	0
F	1	0	1	1	0
G	1	0	0	1	0
H	0	1	0	0	0
U	0	1	1	1	?
V	1	1	0	1	?
W	1	1	0	0	?

For questions 1-3, you should consider only the training data.

1. What is the entropy of Edible?
2. Which attribute should you choose as the root of a decision tree? Hint: You can figure this out by looking at the data without explicitly computing the information gain of all four attributes.
3. What is the information gain of the attribute you chose in the previous question?
4. Propose a decision tree for this problem and use it to classify mushrooms U,V and W.

Exercise 2 - Neural networks (5 points)

1. Draw a perceptron implementing the XNOR(A,B) = (A AND B) OR (NOT A and NOT B) function having the following truth table:

A	B	XNOR
0	0	1
1	0	0
0	1	0
1	1	1

What is the optimal number of the hidden layers needed for it?

2. Calculate the gradient for any weight in the last layer for the squared loss function.

Exercise 3 - SVM (5 points)

1. True or False

   a) For a given learning sample, SVM can output different classifiers depending on the order of the examples considered.

   b) The polynomial kernel is related to data projection involving the monomials of the original features.

   c) The Gaussian kernel can be considered as the most expressive kernel.

   d) The bias-variance tradeoff cannot be modeled with SVM.

   e) The representer theorem allows one to kernelize a learning procedure using empirical risk minimization with a L2 regularization applied on the parameters of the model.

2. Assuming that it is possible to encode some information on movies into feature vectors of $\mathbb{R}^d$, we would like to use SVM to build a movie recommendation model. Explain how to learn an SVM for this objective.

Exercise 4 - HMM (5 points)

We would like to create a tool capable of simulating the sounds of a drum like: *Boum Boum Bam Tam Tam Bim Boum*. To simplify, we assume that this tool can only produce 4 sounds denoted by the following symbols: `Bam`, `Bim`, `Boum`, `Tam`. With this tool, we would like to be able to produce mainly 3 types of patterns corresponding to the following descriptions:

a) generating sequences with the sound `Tam` only,

b) alternating series of `Boum`, `Bam` such that the sound `Boum` occurs two times more often than the sound `Bam`,

c) generating the 3 sounds `Bam`, `Bim`, `Boum` uniformly.

We would like the tool to be designed such that the probability of continuing the generation of a given pattern is always significantly higher than moving to a new one. We also would like that the tool has twice the probability to start generating the pattern *b)* than the others.

1. Propose a way to model this tool by a HMM. Draw the HMM graphically (states, transitions and probabilities).

2. Compute the probability to generate the sequence `Boum Tam Bim`.

3. What is the most likely sequence of states that explains the previous sequence?