ADVANCED MACHINE LEARNING

Question set

May 2024

Remark 1: Deadline for delivering the questions and the report of practical assignment: **23.59, 9 Jun 2024**. Upload the solutions in a pdf file to a task in poliformat.

Remark 2: The marks will depend on the depth and comments of the solutions. The presentation of the exercises will be also taken into account.

Remark 3: These exercises are individual exercises. The professor could ask for some clarifications in oral sessions upon request.

Theoretical exercises

Question 1 (1 point)

Compute the Kullback-Leibler divergence between two distributions, $D(p||q)$, such that $p(x) = N(\mu_1, \sigma)$ and $q(x) = N(\mu_2, \sigma)$. Provide a graphical representation of this divergence.

Question 2 (2 points)

Define a PFA $\mathcal{A}$ with 10 states that includes loops and cycles, with the vocabulary $\{a, b\}$. Then, apply theorem in slide 17 to compute $H_{\mathcal{A}}(\Theta)$.

Question 3 (1 point)

Apply the algorithm in slide 21 with PFA in slide 22 to the string "ababaa".

Question 4 (2 points)

Repeat experiment in page 61 of the slides but with a training sample of size 1000. Explain your results and the conclusions. Be clear and concise.

Question 5 (2 points)

Solve just one of the exercies with four stars in page 67 of the slides.

Question 6 (2 points)

Choose a public dataset and perform a classification experiment with mallet. The mark will depend on the difficulty of the task, the achived results in comparision with the state of the art (if available), and the depth of the exercise (max. extension of this exercise: one page). You can look for open data at `https://datos.gob.es/es/`.