Assignment 1

Computer Science Department, University of Crete

MACHINE LEARNING - CS 577, Fall 2024

Assignment 1

Deadline: Sunday, 13/10/2024, 23:59 on e-learn (`https://elearn.uoc.gr`).

Deliverable files: Submit a zip file containing a report in PDF with the answers **AND all** Python files (.py) written by you in the scope of the assignment. The final grade will be the result of the quality of your submitted results in your report, together with the correctness of your submitted code.

Python Version: Use python **3.6**, a clean approach to define the python version is via Miniconda, i.e., creating a virtual environment. You can find more on `https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html`

Exercise 1 - Probabilities (Theoretical)

Solve the following problems:

a. In a population of n couples, where each couple has exactly 2 kids. What is the probability that all couples have 1 daughter and 1 son? (assume binary equiprobable genders)

b. You flip a fair (50-50 chance of each side) coin n times. What is the probability that all n times you get Heads?

c. In a box, there are beads of 3 colours: red, green, and blue at percentages 30%, 50% and 20%. Half of the red, 2/3 of the blue, and 2/3 of the green beads are hollow. Which is the probability of drawing a hollow bead by picking a random bead from the box?

Exercise 2 - Bayesian Theorem (Programming)

Include in your submission the python files that implement this exercise and **show the plots in your report**. Do not forget to put labels on the axes.

a. A photon package emitted by a star has a probability of 1e-7 to pass through the Earth's atmosphere and reach a detector placed on the ground.

   The detector has the following properties:

   i) A false positive rate (= probability of incorrectly reporting the detection of a photon package, when <u>no</u> photon package was actually received) of 10%.

   ii) A true positive rate (= probability of correctly reporting the detection of a photon package, when a photon package was actually received) of 85%.

   Which is the probability that – if the detector reported a detection – a photon package was actually received?

b. Assume that the photon package is composed of 100 photons. Each photon has an equal probability [i.e. 25%] to carry an energy of either 10, 20, 30, or 40 electron-volts. That means that each package has photon energies roughly distributed according to a uniform distribution in energy.

   Which distribution the <u>*total*</u> energy of a package (i.e. the sum of the energies of the photons composing a package) follows? Plot it. Does it look like a familiar distribution? Why?

   ***Hint**: You can visualize such distribution as follows. Uniformly sample energies for the 100 photons in the photon package. Sum their energies. Then repeat the experiment for, say, 2 000 packages. Now plot the histogram (i.e. the distribution) of total energies.*

c. We will now assume that the photon package doesn't simply have a defined probability of 1e-7 to reach the ground. Instead, due to stochastic emission from the star and interaction with the atmosphere, the probability density function for this event is described by a normal distribution $\mathcal{N}(\mu, \sigma)$ centered at $\mu =$ 1e-7, and with $\sigma =$ 9e-8.

   The resulting probability that – if the detector reported a detection – a photon package was actually received will now be a distribution (instead of a single value). Plot such distribution. Does it make sense?

   ***Hint**: Randomly sample from $\mathcal{N}(\mu, \sigma)$, say, 10 000 values. Then, for each sampled value, calculate the probability of package reception as in 2.a and store them in an array. Finally, create a histogram of these probabilities.*

   ***Note**: Probabilities cannot be negative! Reject negative values when sampling from $\mathcal{N}(\mu, \sigma)$.*