Domaine Sciences et Technologies
MASTER 1 INFORMATIQUE

Aix*Marseille université

Aspects probabilistes pour l'informatique : TP 1

Code UE: SINB19AL

Année 2018-19

Monte Carlo and PAC learning

The goal of this TP is to implement the Monte Carlo and the PAC learning methods and to test them.

Question 1 (Estimation de $\pi = 3{,}141592653589793\ldots$)

(a) Write a program estimating π using the fact that the area of a circle of radius r is $\pi \cdot r^2$ and drawing randomly and independently points inside the square of side $2r$.

(b) How many points should be drawn to get an approximation of π with 4 decimals ? Try it.

(c) What will happen is now the circle is of radius r but the square is of side $10r$? Compare the results.

(d) Do the estimation of π using the Buffon needle method. The method consists in dropping randomly and independently a needle of length ℓ on a floor made of parallel strips of wood, each the same width ℓ, and counting the number of times the needle intersects a line between two strips. How to draw a random segment of length ℓ ?

(e) Compare the two methods.

Question 2 (Estimation de $e = 2{,}7182818284590 45\ldots$)

(a) Write a function which returns the smallest n such that $\sum_{i=1}^{n} r_i > 1$, where r_i are random numbers from $[0{,}1]$ drawn uniformly. Perform N such experiences and return the average $\frac{\sum_{i=1}^{N} n_i}{N}$ as e.

(b) How many experiences must be run to get an approximation of e with 4 decimals ?

(c) Alternatively, write a function which returns the smallest m such that

$$r_1 > r_2 > \cdots > r_{m-1} \leq r_m.$$

Perform M such experiences and return the average $\frac{\sum_{i=1}^{M} m_i}{M}$ as e.

(d) Compare the two methods.

Question 3 (Estimation de l'aire d'un polygone P).

(a) Draw a simple polygon P inside a square and organize its list of vertices as a circular list.

(b) Write a function which given a point $p = (x_p, y_p)$ computes if p is inside the polygon P or outside.

(c) Use the function from the previous question to implement the Monte Carlo method for approximate computation of the area of P.

(d) Use the *shoelace formula*

$$A(P) = \frac{1}{2}[x_1y_2 + x_2y_3 + \ldots + x_{n-1}y_n + x_ny_1 - x_2y_1 - x_3y_2 - \ldots - x_ny_{n-1} - x_1y_n]$$

to compute the exact area of a simple polygon P with n vertices $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2), \ldots, p_{n-1} = (x_{n-1}, y_{n-1})$, $p_n = (x_n, y_n)$. Compare the two results.

(e) (optionel) Do the Monte Carlo method for polygons with holes. The previous shoelace formula does not longer works in this case. How to compute exactly the area of a polygon with holes ?

Question 4 (PAC Learning).

(a) Draw a hidden axis-parallel rectangle $\mathcal{R}$. Draw randomly and independently N points (based on the estimation given in the lecture) depending of the VC-dimension and the value of ϵ. For each of the drawn points p, give its label "+" or "-" depending if p belongs to $\mathcal{R}$ or not (training set of examples). Based on the obtained results, return an axis-parallel rectangle $\mathcal{R}^*$ consistent with all positive and negative examples. Compute the error between $\mathcal{R}$ and $\mathcal{R}^*$ and compare it with ϵ. Repeat the same experience several times (depending of the chosen value of the confidence δ) and return the best $\mathcal{R}^*$.

Do the same experiences for the case where the hidden concept is a union of two disjoint rectangles.

(b) Do the same experience for a circle $\mathcal{C} = \{p = (x_p, y_p) \in \mathbb{R}^2 : (x_p - x_0)^2 + (y_p - y_0)^2 \leq r^2\}$ of center $p_0 = (x_0, y_0)$ and radius r and for a halfplane $\mathcal{H} = \{p = (x_p, y_p) \in \mathbb{R}^2 : y_p \leq ax_p + b\}$. Return a circle $\mathcal{C}^*$ and a halfplane $\mathcal{H}^*$ consistent with positive and negative examples.

Do the same experiences for the case where the hidden concept is (a) the union of two disjoint circles ; (b) the union of two halfplanes ; (c) the intersection of two halfplanes.