SCC461 – Programming for Data Scientists

Leandro Marcolino

Week 10

Assignment

Deadline: Monday, 18/12, 5pm
Upload on Moodle your code, your test cases (with the output), and your short reflection.

1. Logistic Regression (3%)

   In this assignment, you must implement and test a Logistic Regression class. You can consider a binary classification, using only two features.

   A logistic regression is given by:

   $$\sigma(\mathbf{x}) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}},$$

   where $t = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \ldots + \beta_n \times x_n$. x_i is the feature i of the item $\mathbf{x}$, and β_i is a weight that must be learned.

   Given an item $\mathbf{x}$, we can update the weights of the logistic function by using the following equations:

   $\beta_i := \beta_i - \alpha \times (\sigma(\mathbf{x}) - \text{label of } \mathbf{x}) \times \sigma(x) \times (1 - \sigma(x)) \times x_i$

   $\beta_0 := \beta_0 - \alpha \times (\sigma(\mathbf{x}) - \text{label of } \mathbf{x}) \times \sigma(x) \times (1 - \sigma(x)) \times 1$

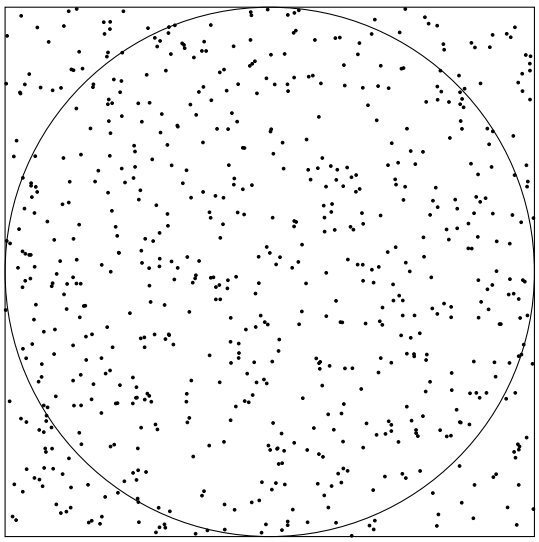
   In the Stochastic Gradient Descent, we order the items randomly, and use one item at a time when adjusting the weights. You can do multiple passes over your training set (using a different random orderings at each time). You can find more details in the slides of SCC403 – Week 6.

2. Monte Carlo method (2%)

   It is possible to use statistics in problem solving. For instance, if we have a circle of radius R, inside a square of length $2R$, and we place N points randomly inside the square, we know that $\frac{N \times \pi}{4}$ points are going to be inside the circle (approximatelly).

   Implement a program that uses the Monte Carlo method to approximate the value of π. Use a large N for a high-quality approximation.

As mentioned in class, you must write a short text reflecting how you approached these problems. You must also report who you discussed with, what you searched online, who you helped, etc. Discussions are allowed, and looking for online materials, books, etc, is allowed. However, directly copying full Python code is not allowed.



N random points
inside the square

Inside circle:

$$\frac{N*\pi}{4}$$

Figure 1: Approximating π by Monte Carlo method.