University Côte d'Azur
Master

Data Mining for Networks
Year

TP n°

Data Mining for Networks

For all exercises, try to :

— See them as some kind of research project. When coding an algorithm, think about its algorithmic complexity, other potential algorithms, why this algorithm is good or not, in which context. Test it on adequate example, potentially of different sizes.

— Use a python notebook with examples instead of just python functions.

— If asked to write a classic function (e.g. `isIsomorphic`) do not use the one from existing library. The goal for you is to understand how work the function and how to code it efficiently.

1 Assignment 1

Exercise 1 [Gradient descent algorithm]

$$f(x) = 2x^4 - 4x^3 + 3x^2 + 4x - 3$$

Apply gradient descent to x.

1. Write the derivative
2. Write the algorithm (take $\alpha = 1/10$)
3. Do 2 steps of gradient descent starting from $x = 0$ and from $x = 10$.
4. Comment.

Exercise 2 [Linear regression and Gradient descent]

1. Install scikit-learn
2. Write a function Linear-Regression (Model and solution using gradient descent)
3. Use mini-batch gradient descent. Compare with Batch gradient descent and stochastic gradient descent.
4. Compare simultaneous and non-simultaneous updates.
5. Compare the different solutions (e.g. Compare the path of intermediate solutions on an example.).

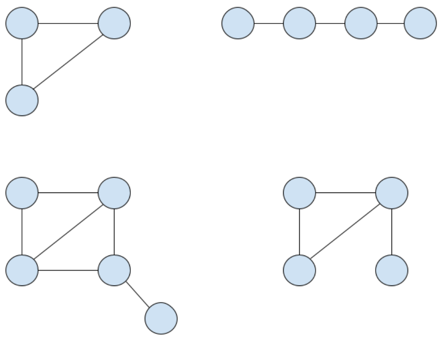
2 Assignment 2

Exercise 3 [Graph comparison. Graph edit distance]

1. Install networkX for manipulating graphs in Python.
2. Write a function to compute the graph edit distance between two graphs.
3. Provide examples to test the function.

Exercise 4 [Ideal Graph Kernel]

1. Compute the feature values of the 4 following graphs for the ideal graph kernel.



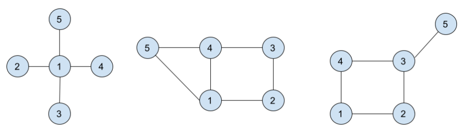
3 Assignment 3

Exercise 5 [Graph comparison. Graph kernels] The goal of the homework is to get an intuition of what are similar and not similar graphs with respect to the Random Walk Kernel. In the following, we consider a kernel function using only walks of lengths $k = 0, 1, 2, 3, 4$, that is we use the formula :

$$k_X(G, G') = \sum_{i,j=1}^{|V_X|} [\sum_{k=0}^{4} A_X^k]_{ij},$$

where A_X is the adjacency matrix of the direct product graph $G \times G'$ and $|V_X|$ its number of vertices. For a matrix M, M_{ij} is the element at the intersection of line i and column j.

1. Consider the following graphs G, G' and G". Compute the random walk kernel value,

   G G' G"

   $k_X(G, G')$, of graphs G and G' and $k_X(G', G'')$, the random walk kernel value of graphs G' and G''.
2. Discuss the results.
3. Play with other graphs and exhibit other triplets (at least 3 triplets) of graphs (G_1, G_2, G_3) for which G_1 and G_2 are similar and G_1 and G_3 are not.