CS Games 2015 - Université de Sherbrooke

Machine Learning Competition

Vincent Picard

The city of Beerbrooke recently decided to install a smart camera system. The system, has many features such as movement detection. One interesting feature is path detection. Which automatically retrieves the 2D trajectory of pedestrians. The city has had many complains from the downtown residents about drunk people having a bad behavior. This led them to think about a system which would automatically detect which pedestrians had too much alcool and which didn't from their path.

You need to elaborate a system that will recognize inbriated people from sober people. You have a sample dataset of 500 paths for each kind of path (drunk, sober). Your system should be able to generalize the given data. Once the system is in place you can expect to have frequent requests. The client (Beerbrooke) also wants to have minimal running costs.

Instructions:

- You need to implement a **machine learning** algorithm that will be able to identify the path of drunk and sober people.
- Solutions that do not present a **machine learning** algorithm will be penalized (but still looked at). See the **machine learning** definition :

> "*Machine learning is a scientific discipline that explores the construction and study of algorithms that can learn from data. Such algorithms operate by building a model based on inputs and using that to make predictions or decisions, rather than following only explicitly programmed instructions."* - Wikipedia, January 2015

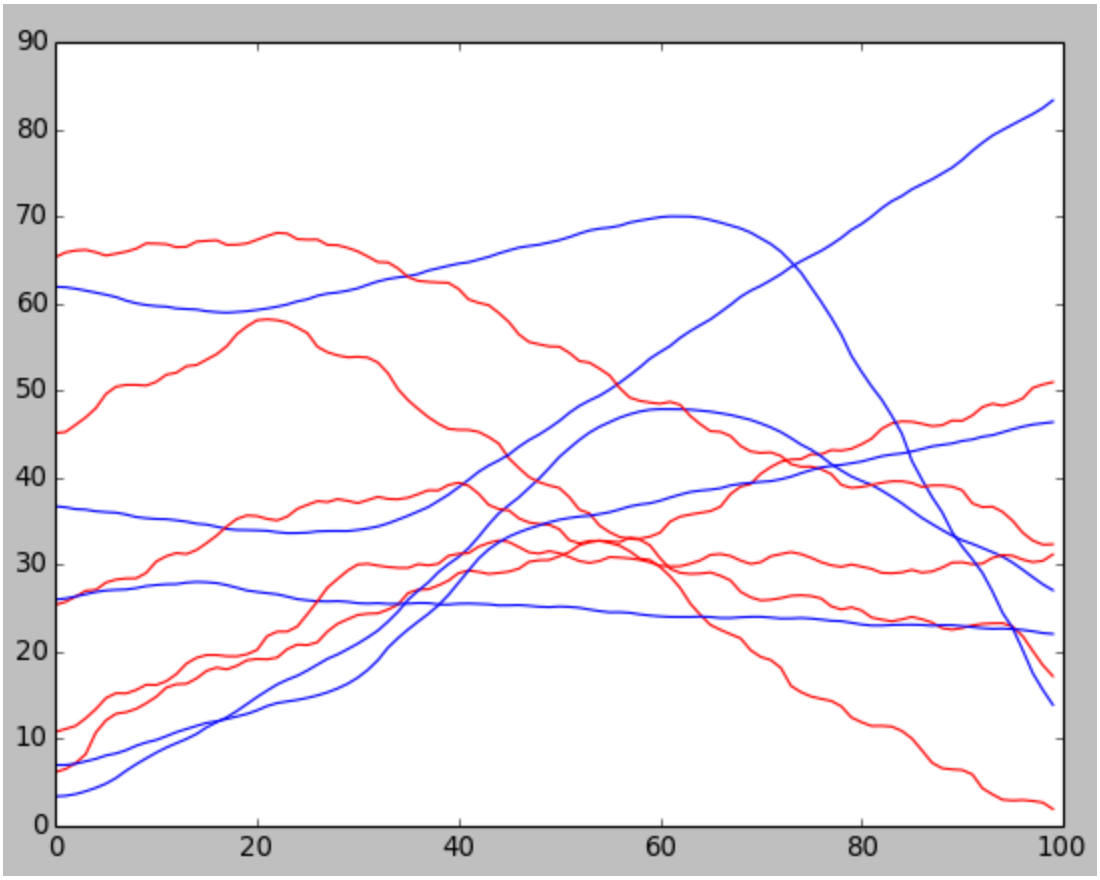
- You can make your solution using Java or Python.
- You must complete the *train* and *evaluate* functions in the *Solution* class. See Python/Solution.py OR Java/src/MachineLearningChallenge/Solution.java. The *evaluate* function will be called with new data to evaluate the efficienty of your solution during the correction.
- Answer the questions on the following pages. If you need more space, extra sheets will be given.
- Questions you may have will not be answered.
- You have 3 hours to complete the competition.

Name (1): _____

Name (2): _____

Team: _____

Here is an example of how the data looks :



Red paths are the paths of inhebriated people. Blue paths are the paths of the sober one. Your algorithm will receive two arrays of 500 paths in the *train* function. Each path is a list of 100 floating point number. The number is the y coordinate and the position in the list is the x coordinate.

The *evaluate* function of your algorithm should be able to categorize the given path (one list of 100 floating point numbers) using machine learning.

1. ALGORITHM CHOICE. Which machine learning algorithm did you choose for this challenge, and why ?

What would have been your choice if you had more time. Would you have tried other machine learning algorithms? which one(s)? why?

2. DATASET USAGE. Explain how you used your data for training and for testing your algorithm precisely. Your answer should explain exactly how you used the 1000 samples.

3. Any transformation made to the data should also be explained, what is the final input of your algorithm?

4. HYPER PARAMETERS AND INITIALIZATION. Explain how exactly you setted up your algorithm. More precisely, if your algorithm has hyper parameters, which values did you use and why?

5. Does your algorithm need initialization? Which values did you choose? Why did you choose those values?