Data Analysis & Pattern Recognition

Supervised Machine Learning

Laboratory exercise

Gerard Escudero

Lab: overview

You are proposed 2 problems: classification, regression. It is advisable that you solve all two problems but your group should only deliver one of them (your choice).

Instructions:

- Work in groups of 2
- Choose **one** of the three mini-projects
- One delivery per group (on Atenea). Remember to include both names
- Deliver executed Jupyter notebook
- Explain why you do what you do; analyze the results you obtain

Lab: guidelines

1. Download and examine the data (attributes, classes, balances)
2. Prepare the data: missing, errors, encodings, normalization
3. Choose an evaluation protocol
4. Train, evaluate and compare next algorithms:
   - Distances: kNN, centroids*
   - Probabilistics*: gnb, lda, logistic regression
   - Rules: dts, bagging with knn, random forest, adaboost, gradient boosting
   - Hyperplanes: SVMs (linear, poly & rbf), perceptron*, MLP
5. Analyze the results obtained, propose improvements

Play and explore: reports with just the basics will not have max grade, you are expected to try different things, see what happens and compare the results.

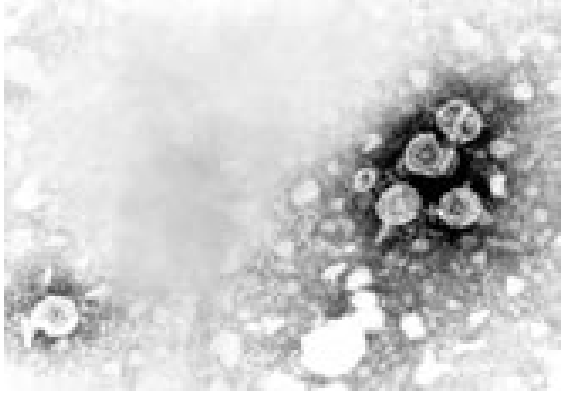
* Only for classification problem.

Classification

Hepatitis Data Set

Abstract: binary classification problem with mostly boolean or numeric-valued attribute types.

Hepatitis* (names / data)



Regression

Forest Fires Data Set

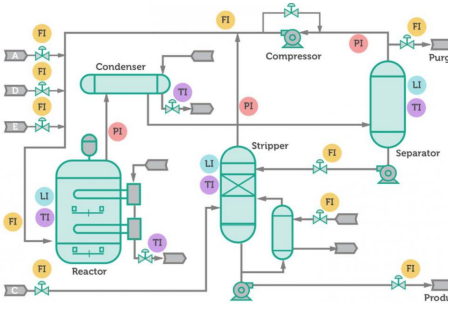
Abstract: This is a difficult regression task, where the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meteorological and other data.

Forest Fires* (names / data)

Fault Detection or Diagnosis

Tenessee Eastman

Abstract: This is a chemical engineering dataset, where the aim is to design the problem and apply the learning algorithms to solve it.



- Tenessee Eastman Data
- (Downs & Vogel, 93)
- How to read RData in pandas

- You can also use this data for the deep learning project.
- You should present the experiment design to the teacher before proceeding.