Big Data Mining
Apache Spark Session Lab 1

Albert Bifet and Jacob Montiel

October 6, 2017

Apache Spark Session Lab 1

- Login in the Community Edition of Databricks:
  - `http://community.cloud.databricks.com/`
- Create a cluster
- Read and run these two notebooks:
  - Introduction to Apache Spark on Databricks
  - Quick Start DataFrames

- Import from Workspace this notebook on ML Pipelines
  - `https://drive.google.com/file/d/0Bz5RPwpp2VWxTTJGNEJfZ3dJNnc/view?usp=sharing`
- Read and run the notebook
- Try to use another classifier instead of Naive Bayes