University of Rome "La Sapienza"

Master in Artificial Intelligence and Robotics

Machine Learning

A.Y. 2018/2019

Prof. Luca Iocchi

Sapienza University of Rome, Italy
Master in Artificial Intelligence and Robotics
Machine Learning (2018/19)

Exercise 2

Luca Iocchi

Exercise 2. Text Classification with Naive Bayes

Learn target function $f : SMS \mapsto \{ham, spam\}$

SMS Spam Collection data set
http://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection

ham some text ...
spam some other text ...

Implement a Naive Bayes Classifier and evaluate performance.

Exercise 2. Text Classification with Naive Bayes

Implementation hints

- Use either Bernoulli or multinomial distributions
- Use a proper evaluation method: either repeated random splits or k-fold cross validation
- Design a modular program (some parts can be reused for Homework 1)
- Use appropriate data structures (e.g., dictionaries or maps instead of feature vectors)
- Perform incremental tests (start with a small dataset to test your program)
- Comment the results by analyzing confusion matrix, precision and recall.

Exercise 2. Text Classification with Naive Bayes

Learning goals

- Understanding the concept of Naive Bayes assumption
- Solve a moderately complex classification problem
- Implement a medium-difficulty learning algorithm