IFT6390 Final Project

Professor: Ioannis Mitliagkas October 16th, 2019

1 Project Summary

The goal of this project is develop experimental skills and intuition by running different classifiers on different datasets and comparing their performance and characteristics. You will find a team, write a project proposal, a short paper summarizing your results, make a poster, and present your results to your peers.

2 Instructions

- Find a team of 3 people (the forum on Studium can be very helpful for this).
- Choose 3 or more classifiers (or regression algorithms) and 2 or more datasets for your experiments. Some suggested classifiers and datasets are provided in Section 3, but you are free to use others. We encourage you to choose data and/or methods that you are interested in and motivated to explore.
- Write a one-page project proposal, **to be submitted November 12th**. The proposal should contain:
  1. A title (your choice), team member names, the course code, and date
  2. A short abstract motivating your choices and describing the phenomena and problems you want to investigate
  3. The list of classifiers you have chosen, with a brief description of each
  4. The list of datasets you have chosen, with a brief description of each
- Write a 4-page paper in the NeurIPS style, **to hand in at the end of term** (DATE TBA), summarizing your project, results, and conclusions:
  1. Describe your approach (chosen datasets and classifiers, with motivation - similar to the project proposal)
  2. Run each of the classifiers on each of the datasets, i.e. at least 6 experiments. We strongly encourage you to run more experiments than this, e.g. to tune hyperparameters or try different regularizers. Include all of these experiments in your final report (in an appendix, if appropriate).
  3. Report results, in terms of accuracy (in a table) and visually (learning curves, decision boundaries, or other figures which are helpful for explaining your results).
  4. Compare the performance and characteristics of each classifier on the different datasets; make recommendations about choosing classifiers, and observations about the properties of your datasets. See section 4 for research questions and other suggestions.
  5. Include in the "Acknowledgements" section a description of each team member's contribution to the project
- Produce a poster in A0 format (or a set of 9-12 printed A4/letter-size slides), summarizing your results, recommendations, and observations. **You will present this poster to your peers at the end of term** (DATE TBA).

3 Suggested classifiers and datasets

3.1 Classifiers

- Naive Bayes (density estimator of your choice)
- Random Forest
- Support Vector Machine
- Multilayer perceptron
- Convnet

3.2 Datasets

- Fashion MNIST (https://github.com/zalandoresearch/fashion-mnist). MNIST is not allowed.
- CIFAR-10 (`https://www.cs.toronto.edu/~kriz/cifar.html`)
- Bank telemarketing (`https://archive.ics.uci.edu/ml/datasets/Bank+Marketing`)
- Wine quality (`http://www3.dsi.uminho.pt/pcortez/wine/`)
- Income prediction (`http://archive.ics.uci.edu/ml/datasets/Adult`)
- Sentiment analysis (`http://ai.stanford.edu/~amaas/data/sentiment/`)

4 Research questions and suggestions

This is a by-no-means-exhaustive list of some of the questions you may want to explore in your paper.

- Which algorithm is theoretically more likely to overfit? What happens in practice?
- What prior assumptions are encoded by your chosen algorithms? Are these 'good' assumptions?
- Compare the performance (accuracy, log-likelihood...) of the different algorithms on each dataset. Is one consistently better than others? Why?
- Explore your data - provide simple statistics and/or analysis like the mean, variance, (possibly over different features), visualizations of how features vary over classes, etc. Do any of these properties help explain different algorithms' performance? Compare statistics between your datasets - is one more balanced than another? Is this a good thing?
- What properties lead to different algorithms' performance? How robust is their performance; what would cause it to degrade?
- What have you learned about your data? What real-world problems could your insights be useful for? In what ways is your data representative of a real-world problem, and it what ways is it not?
- Do you have any recommendations for someone faced with this real-world problem? (Suggested algorithm/hyperparameters, data collection methods, etc.)
- Provide qualitative analysis of performance - show examples that the algorithm fails on, and examples of successes. Are the 'hard' and 'easy' examples in the dataset the same for different algorithms?
- What are some other phenomena or interesting questions you would recommend for further work/experiments?