Model Evaluation and Selection with Scikit-Learn - Jaidev Deshpande

About the speaker

I am a data scientist based in New Delhi. I currently work as the Practice Lead, Data Science at Juxt Smartmandate Analytical Solutions Pvt Ltd. I am an active member of the Python community. I've spoken at various conferences about my FOSS work. My research interests are in signal processing and machine learning. In my spare time I like to dabble with applications of machine learning in personal productivity.

Abstract

It is relatively easy to fit models on a dataset and try them out on another dataset. This is partly because of the wide availability of machine learning packages like scikit-learn, Weka, Spark, etc. Not only is machine learning valuable for its applications in predictive modeling and automation, but it also proves insightful when it explains data or the processes that generated data. Therefore, while the usability of machine learning is an important target, the explainability of machine learning is more rewarding one.

This workshop focuses on just that. By the end of the workshop, you will be ready to answer questions like, given a dataset set of models to choose from:
1. How to pick the best?
2. How to define "best"?
3. How to benchmark models?
4. How to figure out if a model isn't overfitting (learning things it is not supposed to learn)?
And so on…

The entire workshop is based only on the model selection and evaluation techniques available within scikit-learn. The audience is expected to have a basic understanding of the terminology of machine learning. If you can tell the difference between supervised and unsupervised learning or between classification and regression, you will be able to consume the presented material.

This will be a fully hands-on workshop, containing demos in Jupyter notebooks, interspersed with exercises.

Software prerequisites:
1. Scikit-learn (and dependencies)
2. Matplotlib
3. Pandas