CMSC 360: Intelligent Robotics and Perception
Lab 4: Camels – Machine Learning
May 24, 2016

Overview

In the last unit of the course we are studying robots that rely on more complex perceptual and learning algorithms. In this lab, we will explore different machine learning algorithms. Machine learning has found applications in many different areas, but it's potential in robotics is particularly exciting. A robot's sensors provide rich streams of noisy data over extended periods of time. Algorithms that learn from experience, and that can adapt to changing environments, unburden us from pre-programming every little detail.

In this lab we will use the `scikit-learn`[1] open source machine learning library. Scikit-learn is a Python framework for machine learning that implements many standard learning algorithms.

Learning Objectives

- Evaluate a supervised machine learning algorithm
- Evaluate an unsupervised machine learning algorithm
- Use an open-source machine learning package
- Access and process online datasets

Part One - Pick a Dataset

First, you will have to locate an interesting dataset. The more attributes and instances, the better. In your lab report, be sure to justify why you thought this dataset was interesting. The scikit-learn website has links to datasets, but you can find other sources like Google Fusion Tables[2]. You are encouraged to use a dataset you collect and create (e.g. based on your semester project).

Part Two - Evaluate an Unsupervised Learner

Choose one the clustering or pre-processing algorithms (unsupervised learners) implemented in scikit-learn (e.g. principal components analysis, k-means) that looks interesting and seems reasonable for your dataset. Write a page or so describing how this particular algorithm works. Report the results of running this algorithm on the dataset you chose in part one. You should explore the algorithm by varying at least one of its critical parameters. For example, how the k-means clustering changes as you add more cluster centers.

Part Three - Evaluate a Supervised Learner

Choose one of the supervised learners (e.g. decision tree (J48), neural network) that looks interesting and seems reasonable for your dataset. Write a page or so describing how this particular algorithm works. Report the results of running this algorithm on the dataset you chose in part one. You should explore the algorithm by varying at least one of its critical parameters. For example, how the classification performance of the neural network changes with the number of hidden nodes.

---

[1] http://scikit-learn.org/
[2] https://support.google.com/fusiontables/answer/2573812?hl=en&ref_topic=1652595