CS4881 Artificial Intelligence Jay Urbain

Problem Set 4 – Machine Learning

The idea behind learning is that percepts should be used not only for acting, but also for improving the agent's ability to act in the future. Learning takes place as the agent observes its interactions with the world and its own decision-making processes. We focus on inductive learning from observations.

The type of feedback for learning is usually the most important factor in determining the nature of the learning problem that the agent faces and the field of machine learning usually distinguishes three cases: supervised, unsupervised, and reinforcement learning.

Implement at least one of the following supervised, unsupervised, or reinforcement machine learning problems as described below; or use an existing machine learning tool kit to analyze the performance of multiple machine learners for a specific problem or across a range of data sets Note: If you would prefer to use a different algorithm from what is listed below, that is probably ok, but please see me for approval first.

You are required to let me know (email) which problem you've selected by Lab of week 9. Be prepared to discuss your algorithm on the last day of class. *All projects must be demonstrated*. **Projects, reports, and demonstrations are due Friday of week 10.**

1) Supervised Learning

Implement a supervised learning algorithm, e.g., Naïve Bayes, Bayes Net, ID3 Decision Tree Algorithm using the information theoretic selection function (information gain), Collaborative Filtering, or Neural Network). To *develop* (not test!) your algorithm, you can use the data set below. At a minimum, you should also execute your algorithm on the "adult" data set from UC Irvine linked to our course outline (there are many datasets available). In your report include a complete analysis of your algorithm, and include a trace of how your machine learner is constructed, how you dealt with (or should deal with) missing values, over fitting, and ideas for how you might boost the performance of your algorithm. You might want to also explore how you can use supervised learning on a number of randomly generated versions of Wumpus World or similar game.

2) Unsupervised Learning

Implement an unsupervised learning algorithm such as K-Means clustering using the "adult" data set from UC Irvine on the course web site. At a minimum evaluate your algorithm over a range of different numbers of iterations and different numbers of clusters. In your report include a complete analysis of your algorithm, and include a trace of your algorithm's clustering steps, how many iteration steps are necessary, issues with unsupervised learning, and ideas for how you might boost the performance of your algorithm. Consider evaluating other clustering algorithms and evaluating your algorithm on a text collection, if so see me.

3) Reinforcement Learning

Implement the temporal difference learning algorithm. At a minimum execute your algorithm on a game such as tic-tac-toe. If you have time, try a more significant game (mine sweeper, Wumpus, Soduko, checkers). In your report include a complete analysis of your algorithm, and include a trace of your algorithms steps, issues with reinforcement learning, and ideas for how you might boost the performance of your algorithm.

4) Analysis Option

If you choose to use a machine learning toolkit, you are required to select a significant problem that has datasets available for testing, evaluate multiple machine learners for the problem, and analyze the results. There is a toolkit based on Java called Weka, and I can provide another toolkit that includes algorithms that I have developed for Naïve Bayes, Decision Trees, and a Neural Network.

Submission:

Quality of reports and analysis will be rewarded. Submit project as a zipped archive including PDF report including results and analysis, and source with your name, class, and problem set number along with your project archive by email. **Project demonstration is required.**

Credit History Data

Number	Credit History	Debt	Collateral	Income	Risk
1	bad	high	none	$0 to 15K	high
2	unknown	high	none	$15 to 35K	high
3	unknown	low	none	$15 to 35K	moderate
4	unknown	low	none	$0 to 15K	high
5	unknown	low	none	Over $35K	low
6	unknown	low	adequate	Over $35K	low
7	bad	low	none	$0 to 15K	high
8	bad	low	adequate	Over $35K	moderate
9	good	low	none	Over $35K	low
10	good	high	adequate	Over $35K	low
11	good	high	none	$0 to 15K	high
12	good	high	none	$15 to 35K	moderate
13	good	high	none	Over $35K	low
14	bad	high	none	$15 to 35K	high

Regards,
Jay Urbain