Bayesian Model Averaging (BMA) for Naive Bayes Classification Models

Ga Wu

u5411311@anu.edu.au

Supervisors: Dr Scott Sanner
COMP 8740, AI Project, Semester 1, 2014
Australian National University
OUTLINE

- Naïve Bayesian Classifier
- Bayesian Model Averaging
- Research Platform
- Challenges
Naïve Bayesian classifier

- What is classifier?
  - Example: why bank refuse to lend to some applicants?
  - That is a classification problem which make a decision based on the applicant’s bank records, Wage’s level and even check criminal records related.

- Bayes’ theorem

\[ P(C_k | D) = \frac{P(D | C_k) P(C_k)}{P(D)} \]
Bayesian Model Averaging

- What is Bayesian Model Averaging?
  
  $$P(c|x, D) = \sum_M P(c|x, m, D)P(m|x, D) = \sum_M P(c|x, m)P(m|D)$$

- Why is it interesting?
  
  - Reduces variance in predictions vs. using a single model.
  - In asymptotic limit of data, converges to the best model.

- can we average over an exponential number of Naive Bayes models in *linear* time? Yes!
  
  - We know one dataset with n attributes can generate $2^n$ models. How to make it linear? Formula deduction!
Research Platform

- What is weka? And why?
  - **Weka** (Waikato Environment for Knowledge Analysis) is a popular suite of [machine learning](#) software written in [Java](#), developed at the [University of Waikato, New Zealand](#).
  - Provide lots of classification algorithm that can be used as Reference.

- What is UCI data?
  - Be widely used as data source for machine learning. (Finding proper data resources is not easy!)
  - Easy to be understand, we can easily know how many attributes a dataset has, how many records, and even the types of attributes!

- Eclipse! JAVA IDE
Challenges

As the goal of this project is to find an high performance classifier algorithm which can be used in general cases without doing feature selection, it raises the following questions.

- Can this get higher performance than Naïve bayes method?
- How to make it work in practice due to numerical precision issues?
  - Underflow/Overflow is almost Predictable!
- How to weight those models?
  - Uniform?
  - Exponential Distribution? (How to define hyper-parameters?)