Credal Classification

Ben Willis

Durham University

2017

Problem

- Classification is the problem of identifying which class ($c \in \mathcal{C}$) an object with attributes ($\mathbf{a} \in \mathcal{A}$) belongs to.
- We will tackle the problem of classifying the risk to an insurer of an auto mobile.
- The data set we will be analysing contains technical information about 205 vehicles and an expert's assessment of their risk.

Naive Bayes Classifier

Probabilistic Interpretation

Bayes theorem and the naivety assumption gives us:

$$P(c \mid \mathbf{a}) \propto P(c) \prod_{i=1}^{k} P(a_i \mid c) \tag{1}$$

Choosing the Class

We introduce the *0-1 loss function* then to minimize the loss we choose the class:

$$\hat{c} = \arg\max_{c \in \mathcal{C}} P(c) \prod_{i=1}^{k} P(a_i \mid c) \tag{2}$$

This is known as the maximum a posteriori (MAP) estimate.

Estimating Chances

To estimate these probabilities we parametrise them.

The Likelihood Function

Parametrise the unknown probability $P(\cdot)$ by $\theta_{\cdot}$. Also denote the number of objects in class c by $n(c)$ and the number of objects in class c with attribute a_i by $n(a_i, c)$. Then:

$$l(\theta \mid \mathbf{n}) \propto \prod_{c \in \mathcal{C}} \left[\theta_c^{n(c)} \prod_{i=1}^{k} \prod_{a_i \in \mathcal{A}_i} \theta_{a_i|c}^{n(a_i,c)} \right] \tag{3}$$

Maximum Likelihood Estimates

A simple estimate for each parameter is the maximum likelihood estimate (MLE). These are:

$$\hat{\theta}_c = \frac{n(c)}{N}, \hat{\theta}_{a_i|c} = \frac{n(a_i, c)}{n(c)} \tag{4}$$

Application and Results

- We discretize our continuous variables and discard objects with missing values.
- We use a technique called *k-fold cross validation* to estimate key metrics.
- Our classifier has an accuracy of 59.95% on the auto mobile data set.
- In addition 22.45% of the objects in our data set weren't classified.

Introducing the Prior

Prior Distribution

We introduce a conjugate prior for the likelihood function eq. (3):

$$f(\theta \mid \mathbf{t}, s) \propto \prod_{c \in \mathcal{C}} \left[\theta_c^{st(c)-1} \prod_{i=1}^{k} \prod_{a_i \in \mathcal{A}_i} \theta_{a_i|c}^{st(c,a_i)-1} \right] \tag{5}$$

with hyper parameters $s > 0$ and $\mathbf{t}$ such that:

$$\sum_{c \in \mathcal{C}} t(c) = 1, \sum_{a_i \in \mathcal{A}_i} t(a_i \mid c) = t(c) \text{ and } t(a_i \mid c) > 0 \tag{6}$$

To comply with the constraints prior constraints let us set:

$$s = 1, t(c) = \frac{1}{|C|} \text{ and } t(a_i, c) = \frac{1}{|A_i||C|} \tag{7}$$

Diagnostics

Introduce the mean squared error

	Accuracy	MSE	Failed Classifications
NBC	59.95%	2.823	22.45%
Corrected NBC	68.17%	0.689	0%

Credal Classifier

Not sure what I want to put here. Something explaining the theory behind how this can be extended to create a credal classifier (Sets of priors, credal dominance).

Future Work

Summary of findings so far and future work