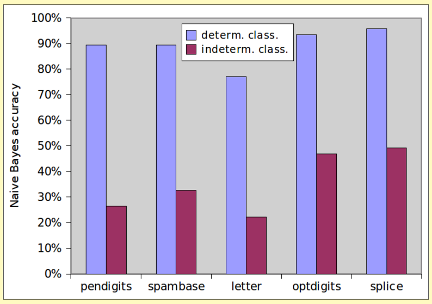
Lazy Credal Classifier and how to compare credal classifiers

G. Corani and M. Zaffalon, IDSIA, Switzerland giorgio{zaffalon}@idsia.ch

Introduction

- The *naive credal classifier* (NCC) (Corani and Zaffalon, 2008) extends naive Bayes to imprecise probabilities.
- NCC returns *indeterminate* classifications (more than one class) on instances that would be classified in a prior-dependent way by naive Bayes (*hard* instances).
- NCC remains reliable even on the *hard* instances thanks to indeterminate classifications.
- Instead, Naive Bayes becomes unreliable on the hard instances.



Accuracy of naive Bayes on instances classified determinately and indeterminately by NCC.

Two issues of NCC

- The simplicistic naive assumption (statistical independence of the features given the class).
- A sometimes excessive indeterminacy.

Lazy Learning

Lazy classifiers do *not* learn until there is an instance to classify (*query*); at that point:

1. the instances of the training set are ranked according to the distance from the query;
2. a local classifier is trained on the k closest instances (k is named *bandwidth*) and issues the classification;
3. the local classifier is discarded, while the training set is kept in memory to answer future queries.

Open question: how to select the bandwidth? We will use *imprecise probabilities*.

Why a Lazy NCC (LNCC)?

- working locally *reduces the bias* due to the naive assumption (Friedman, 1997; Frank et al., 2003);
- to *increase determinacy*, thanks the design of the *banwidth selector*.

Bandwidth selection

The bandwidth is increased until the collected evidence smooths the effect of the choice of the prior.

Pseudo-code for bandwidth selection

- `k=25;`
- `lncc.train(k);`
- **while** `(lncc is indeterminate OR k=training set size)`
  - `k=k+20;`
  - `lncc.train(k);`
- **end**

Comparing Credal Classifiers

Remark:
a credal classifier is *accurate* on a certain instance if its output includes the correct class (regardless the number of returned classes).

1) Discounted accuracy (from multi-label classification)

$$\text{d-acc} = \frac{1}{N}\sum_{i=1}^{N}\frac{(accurate)_i}{|Z_i|}$$

where $|Z_i|$ is the number of classes returned on the i-th instance.

- The discount function entails some arbitrariness: why not discounting on $|Z_i|^2$?
- The d-acc of two credal classifiers can be compared via t-test.

2) Rank test (new)

On each instance we rank two credal classifiers CR_1 and CR_2 as follows:

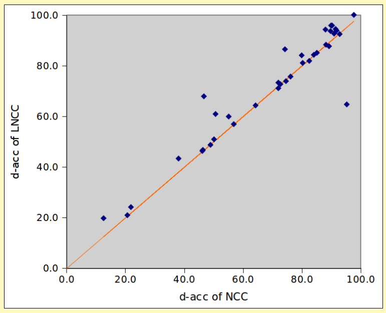
CR1	CR2	winner
accurate	not accurate	CR1
accurate	accurate $\lvert Z_i\rvert_{CR2} < \lvert Z_i\rvert_{CR1}$	CR2
accurate	accurate $\lvert Z_i\rvert_{CR2} = \lvert Z_i\rvert_{CR1}$	tie
inaccurate	inaccurate	tie

- Wins, ties and losses are transformed into ranks and are then analyzed via a non-parametric test.
- The rank test avoids the arbitrariness of d-acc but, using less pieces of information, can be less sensitive.

We cross-check the outcome of both tests.

Experiments

- Comparison of LNCC and NCC on 36 data sets.
- 10 runs of 10-folds cross-validation.
- Supervised discretizazion of numerical features.



Each point refers to a different data set.

	LNCC wins	*ties*	*NCC wins*
d-acc	19	11	6
rank test	15	19	2
cross-check	15	20	1

Why LNCC does outperform NCC?

- On large data sets the improvement is due to the reduced bias.

Data set	instances	d-acc (NCC)	ΔLNCC
letter	20000	86.5	+12.1
nursery	12960	95.8	+5.6
optdigits	5620	93.9	+1.9
pendigits	10992	94.3	+6.3
waveform	5000	84.0	+4.1

- Otherwise, LNCC discovers new instances which can determinately classified.
- *Determinacy*: % of determinately classified instances; *single-acc*: accuracy achieved when determinate.

Data set	LNCC variations over NCC Δdet.	ΔsingleAcc	Δd-acc
anneal	+40.8	0.0	+21.3
post-op.	+18.9	+12.0	+10.3
vowel	+9.4	n.a.	+7.3

References

G. Corani, M. Zaffalon, *Learning reliable classifiers from small or incomplete data sets: The naive credal classifier 2* , J. Mach. Learning Research, (9), 581–621, 2008.

JH Friedman, *On bias, variance, 0-1 loss, and the curse-of-dimensionality*, Data mining and knowledge discovery, (1), 55-77, 1997

E Frank, MA Hall, B Pfahringer, *Locally weighted naive Bayes*, Proc. UAI 2003, 249–256