RISK (2001) (1)

Empirical Study of Naive Bayes

Independee is generally a poor assumption - Still NB Competes well

Bayes Classifiers assign most likely class based on a data instance feature vector

Learning is much simplified by assuming that features are indep per class

$$P(X|C) = \prod_{i=1}^{n} P(X_i|C)$$

$X = (X_1, \ldots, X_n)$, C = Class

Sucess of NB despite poor assump = Optimal Classification error is not necess related to Quality of fit of distribution Approp of independace assumption

Rather the optimal ↘

Rather, optimal classifier =
both actual & estimated distributions agree on the most probable class

This is vague & the paper wants to understand the data characteristic that affect naive Bayes

Most studies of NB uses Benchmark comparison to other classifiers

This paper uses monte carlo sim to allow more systematic study of classif errors
- ~~of~~ on parametric families of rando ~~gener~~ generated problems

- focuses on bias not variance
- by assuming infinite data
- Allows seperation of approx error (bias) & error from sample sample (variance)

"Distribution entrophy"
Quantifies the amount of uncertainty associated w/ the distrib of a variable

low-entrophy dependancies
- Attribs are related to eachother
- esp strong when you focus on data points that belong to the same class
- these related attributes are therefore ~~related~~ Predictable within each class (relationship)

With naive Bayes, if the dependancies are low-entrophy (Predictable), the classifier works well.

This might be considered to be contradict given the independance assum/structure

error of NB vanishes as entrophy approaches ☺

Bayes works best under two Extremes

- Completely unrelated Independant feature - what we would expect

- Functionally Dependant feature - i.e. one feature can be used to deterministically predict another
  - if X = then Y
  - Convert cm to mm

- it performs worst in between
- This would be where features are related by not perfect
- often conditionally related based on other features

e.g. weather, temp change effect Humid — but same ∇ △ does not always cause same change in Humid as other factors @ play

Also

Another way of stating these finds is that the accuracy of NB is not directly correlated w/ the degree of feature dependancies measured as the class-conditional mutual information between the features

- Mutual info = How attribs together provide info about class

Better pred of NB accuracy can be the loss of info that features contain. given as:

$$I(C; X_i, X_j) - \underbrace{I_{NB}(C; X_i, X_j)}_{\text{mutual info}}$$

2 Definitions and Backgrounds

Discriminant function → select highest class

$h(x) = \text{argmax}_i \in \{0, \ldots, M-1\} f_i(x)$

Bayes Risc : $f_i(x) = \underbrace{P(X=x \mid C_i)}_{\text{Class conditional Probability Distribution}} P(C_i)$

likelihood

$P(X=x \mid C_i)$ is hard to estimate/high-Dims (feature space)

↳ why Approximations are used →
Simplistic assum to indep features

yields the NB:

$f_i^{NB} = \prod_{j=2}^{n} P(X_j = a_j \mid C_i)\, P(C=i)$