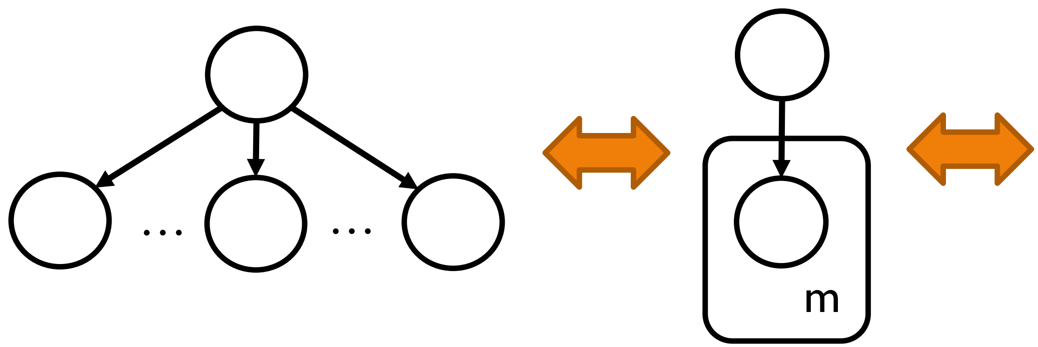
Bayesian Classification: Exemple of Naive Bayes

Naive Bayes:
a very simple Bayesian classification method

(Naive) hypothesis:

Features are independent given class ( )



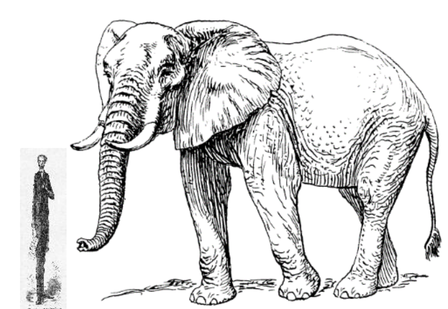
( ) ( )

Ignore information redundancy → requires preprocessing (PCA, etc)

Interpretation: the class mostly determines the characteristics distribution

e.g. : animal species mostly determine size and weight

A small elephant will always be bigger than a tall man



Naive Bayes: parameterization & prediction

Parameters (in the discrete case):

- CPT of :
- CPT of ( ) :

Prediction of given :

Classical Naive Bayes: Maximum likelihood estimation

Estimation of parameters ( ) and (|) by counting

Given observations {( ) ( )} with ( )

( )

(|)

Weka demo: mushroom classification

Naive Bayes:
Example of Spam detection

Given 10 examples of emails from the banking sector:

earn	million	account	password	Class
1	1	0	0	Spam
0	0	1	1	Spam
0	1	1	0	Not spam
1	1	0	0	Spam
0	0	0	0	Not spam
1	0	0	0	Spam
1	0	0	0	Not spam
0	0	0	1	Spam
1	0	1	1	Spam
0	1	1	1	Not Spam

Problem: predict class of messages “earn million”, “million account” and “account password”

Naive Bayes: Solution

Feature		
earn		
million		
account		
password		

Message	Score if spam	Score if not spam	Prediction
earn + million	_ _ _ _ __ __	_ _ _ _ __ ___	Spam
million + account	_ _ _ _ __ __	_ _ _ _ __ ___	Not spam
account + password	_ _ _ _ __ __	_ _ _ _ __ ___	Spam

Naive Bayes:
Extension to continuous features

Learning: estimating ( ) from obs. {( ) ( )}:

1. Assume a distribution shape for each continuous and value of .

e.g. ( ) ( )

2. Estimate parameters for each distribution from

e.g. $\boxed{\frac{(\quad)}{(\quad)}}$ $\boxed{\frac{(\quad)(\quad)}{(\quad(\quad))}}$

Prediction:

$$(\;)\quad\left(\left((\quad)\right)\;(\quad|\;)\right)$$

Naive Bayes: Extension to continuous features

Length in words	Class
30	Spam
40	Spam
100	Not spam
100	Spam
60	Not spam
30	Spam
200	Not spam
90	Spam
70	Spam
40	Not Spam

Naive Bayes:
Extension to continuous features

Feature		
earn		
million		
account		
password		
length		

Message	Score if spam	Score if not spam	Prediction
earn + million			Spam
million + account			Not spam
account + password			Spam

Naive Bayes:
Strong Bayesian version and MAP estimator

Strong Bayesian version introduce priors for each distribution:

- Dirichlet prior for CPT of
- Dirichlet prior for discrete variable
- Gaussian prior for continuous variable with Gaussian distribution

MAP estimator represents prior knowledge with faked examples:

For target class :

$$(\;) \quad \frac{(\qquad)}{}$$

Same for for discrete variable :

$$\frac{\left({}^{(\,)} \qquad \qquad \right)}{(\qquad)}$$

Naive Bayes: A summary

- Most simple model
- Advantages:
  - Robust: low number of parameters $\rightarrow$ low risk of overfitting
  - Fast and simple to compute
  - Deal with mixture of discrete/continuous variables
  - Fully Bayesian (can integrate Dirichlet prior)
- Drawbacks:
  - Independence hypothesis too naive
  - Not accurate for complex classification problems
- Applications:
  - Document classification (e.g. spam detection)