

Machine Learning for Finance

Prepared by Dr. Diaa Salama

FNT633
Naïve Bayes

Lecture 6



Machine Learning for Finance

Lecture 6

FNT633
Navie Bayes

Prepared by
Dr. Diaa Salama

www.midocean.ae

- Classification
- Naïve Bayes
- Classification: Basic Concepts

Naïve Bayes Introduction

Naïve Bayes Basics

Supervised vs. Unsupervised Learning

- Supervised learning (classification)
 - Supervision: The training data (observations, measurements, etc.) are accompanied by **labels** indicating the class of the observations
 - New data is classified based on the training set
- Unsupervised learning (clustering)
 - The class labels of training data is unknown
 - Given a set of measurements, observations, etc. with the aim of establishing the existence of classes or clusters in the data

Prediction Problems: Classification vs. Numeric Prediction

- **Classification**

- predicts categorical class labels (discrete or nominal)
- classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data

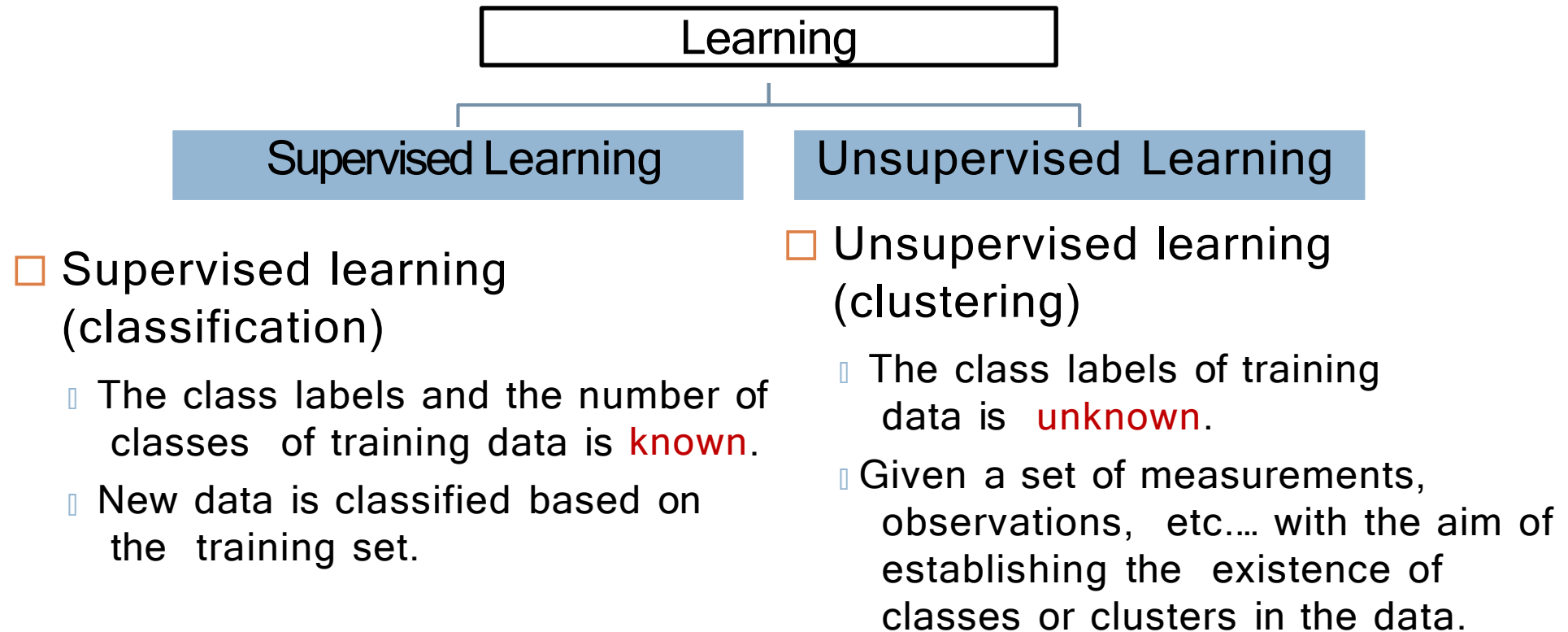
- **Numeric Prediction**

- models continuous-valued functions, i.e., predicts unknown or missing values

- **Typical applications**

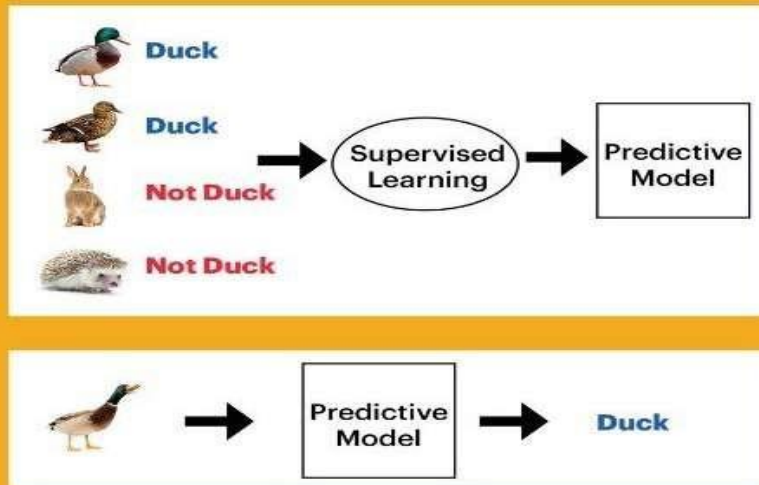
- Credit/loan approval:
- Medical diagnosis: if a tumor is cancerous or benign
- Fraud detection: if a transaction is fraudulent
- Web page categorization: which category it is

Supervised vs. Unsupervised Learning



Example: Supervised vs. Unsupervised

Supervised Learning (Classification Algorithm)



Class labels

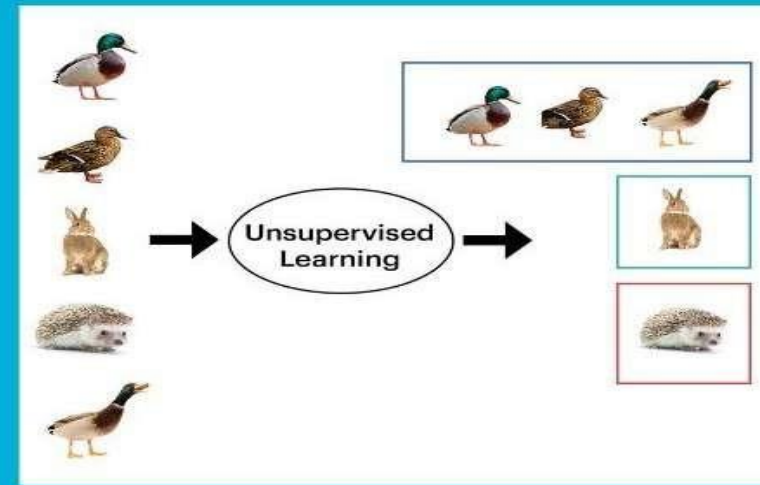
Number of classes

✓✓

X : Unknown in advance

✓ : Known in advance

Unsupervised Learning (Clustering Algorithm)



Class labels X

Number of classes ✓ or

X

- Classification and Regression are two major types of prediction problems.

- Classification:

Predicts categorical class labels (discrete or nominal).

Classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data.

- Regression:

Models continuous-valued functions, i.e., predicts unknown or missing values.

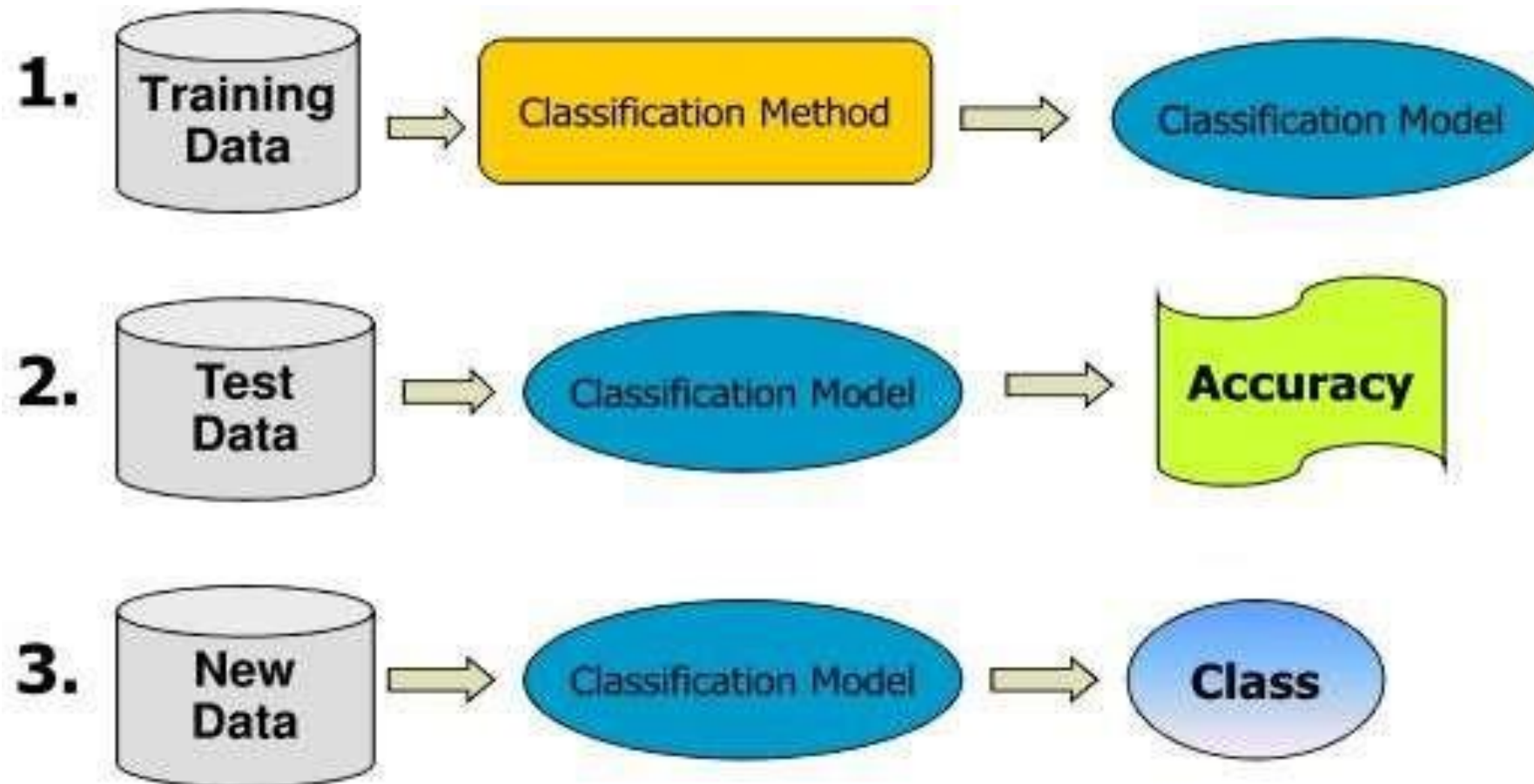


Activity: (classification or regression?)

For some loan application data → predict “safe” or “risky”.

For a marketing manager → predict how much a given customer will spend during a sale.

Classification Process: the 3 steps



Applying the Data Analytics Lifecycle



- In a typical Data Analytics Problem - you would have gone through:
 - Phase 1 – Discovery - have the problem framed
 - Phase 2 – Data Preparation - have the data prepared
- Now you need to plan the model and determine the method to be used.

Phase 3 - Model Planning

How do people generally solve this problem with the kind of data and resources I have?

- Does that work well enough? Or do I have to come up with something new?
- What are related or analogous problems? How are they solved? Can I do that?

Discovery

Data Prep

Model
Planning

Do I have a good idea about the type of model to try? Can I refine the analytic plan?

Validated for sure?

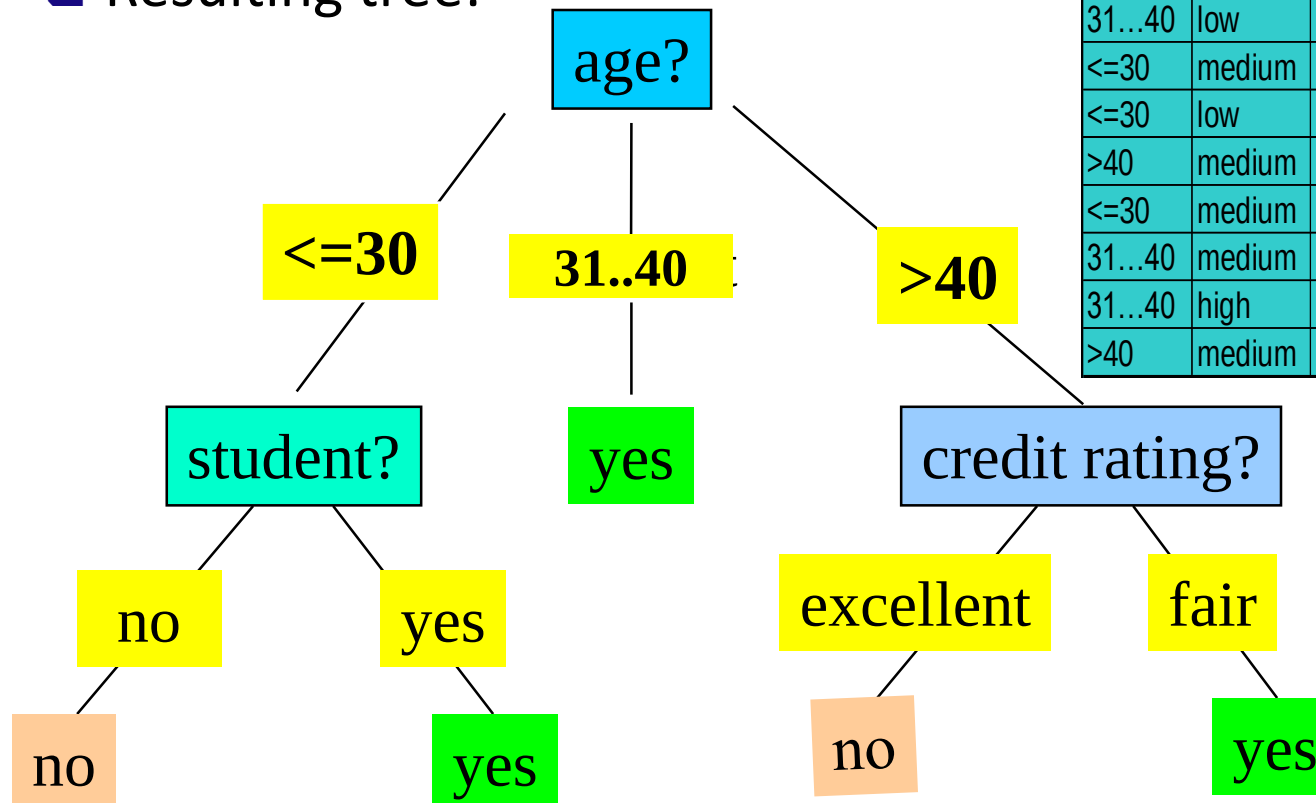
What Kind of Problem do I Need to Solve? How do I Solve it?

The Problem to Solve	The Category of Techniques	Covered in this Course
I want to group items by similarity. I want to find structure (commonalities) in the data	Clustering	K-means clustering
I want to discover relationships between actions or items	Association Rules	Apriori
I want to determine the relationship between the outcome and the input variables	Regression	Linear Regression Logistic Regression
I want to assign (known) labels to objects	Classification	Naïve Bayes Decision Trees
I want to find the structure in a temporal process I want to forecast the behavior of a temporal process	Time Series Analysis	ACF, PACF, ARIMA
I want to analyze my text data	Text Analysis	Regular expressions, Document representation (Bag of Words).

Decision Tree Induction: An Example

- Training data set: Buys_computer
- The data set follows an example of Quinlan's ID3 (Playing Tennis)
- Resulting tree:

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no



Classification: Basic Concepts

- Classification: Basic Concepts
- Decision Tree Induction
- Bayes Classification Methods

Bayesian Classification: Why?

- A statistical classifier: performs *probabilistic prediction*, i.e., predicts class membership probabilities
- Foundation: Based on Bayes' Theorem.
- Performance: A simple Bayesian classifier, *naïve Bayesian classifier*, has comparable performance with decision tree and selected neural network classifiers
- Incremental: Each training example can incrementally increase/decrease the probability that a hypothesis is correct — prior knowledge can be combined with observed data
- Standard: Even when Bayesian methods are computationally intractable, they can provide a standard of optimal decision making against which other methods can be measured

Naïve Bayesian Classifiers

- During this lesson the following topics are covered:
- Naïve Bayesian Classifier
- Theoretical foundations of the classifier
- Use cases
- Evaluating the effectiveness of the classifier
- The Reasons to Choose (+) and Cautions (-) with the use of the classifier

Where in the catalog should I place this product listing?
Is this email spam?

- Classification: assign labels to objects.
- Supervised: training set of pre-classified examples.
- Our examples:
 - 4 Naïve Bayes,
 - 4 Decision Trees

Naïve Bayesian Classifier : What is it?

- Used for classification
 - 4 Actually returns a probability score on class membership:
 - 8 In practice, probabilities generally close to either 0 or 1
- **Input** variables are discrete
 - 4 **Popular for text classification**
- **Output**:
 - 4 Most implementations: log probability for each class
 - 8 You could convert it to a probability, but in practice, we stay in the log space

- Preferred method for many text classification problems.
 - 4 Try this first; if it doesn't work, try something more complicated
- Use cases
 - 4 Spam filtering, other text classification tasks
 - 4 Fraud detection



Building a Training Dataset

- Example : Predicting Good or Bad credit
- Predict the credit behavior of a credit card applicant from applicant's attributes:
 - personal status
 - job type
 - housing type
 - savings account
- These are all categorical variables; better suited to Naïve Bayesian classifier than to linear regression.

personal_status	job	housing	savings_status	credit_class
male single	skilled	own	no known savings	good
female div/dep/mar	skilled	own	<100	bad
male single	unskilled resident	own	<100	good
male single	skilled	for free	<100	good
male single	skilled	for free	<100	bad
male single	unskilled resident	for free	no known savings	good
male single	skilled	own	500<=X<1000	good
male single	high qualif/self emp/mgm	rent	<100	good
male div/sep	unskilled resident	own	>=1000	good
male mar/wid	high qualif/self emp/mgm	own	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	rent	<100	bad
female div/dep/mar	skilled	own	<100	good
male single	unskilled resident	own	<100	bad
female div/dep/mar	skilled	rent	<100	good
female div/dep/mar	unskilled resident	own	100<=X<500	bad
male single	skilled	own	no known savings	good
male single	skilled	own	no known savings	good
female div/dep/mar	high qualif/self emp/mgm	for free	<100	bad
male single	skilled	own	500<=X<1000	good
male single	skilled	own	<100	good
male single	skilled	rent	500<=X<1000	good
male single	unskilled resident	rent	<100	good
male single	skilled	own	100<=X<500	good
male mar/wid	skilled	own	no known savings	good
male single	unskilled resident	own	<100	good
male mar/wid	unskilled resident	own	<100	good

Week

07

Technical Description - Bayes' Law

- B

Week

07

Technical Description - Bayes' Law

- B

Week

07

Technical Description - Bayes' Law

- B

Week

07

Technical Description - Bayes' Law

- B

$$P(B|A) = \frac{P(A \wedge B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$

- B is the class label:
 - 4 $B \in \{b_1, b_2, \dots, b_n\}$
- A is the specific assignment of input variables
 - 4 $A = (a_1, a_2, \dots, a_m)$



Reverend Thomas Bayes

Week

07

Technical Description - Bayes' Law

- B

The "Naïve" Assumption: Conditional Independence

$$P(A|b_j) = P(a_1, a_2, \dots, a_m|b_j)$$

$$= \prod_i^m P(a_i|b_j)$$

so

:

$$P(b_j|a_1, a_2, \dots, a_m) = \frac{\prod_i^m P(a_i|b_j)P(b_j)}{\cancel{P(a_1, a_2, \dots, a_m)}}$$

Independent of class – so it
cancels out

Building a Naïve Bayesian Classifier

- To build a Naïve Bayesian classifier, collect the following statistics from the training data:
 - 4 $P(b_j)$ for all the class labels.
 - 4 $P(a_i | b_j)$ for all possible assignments of the input variables and class labels.

Credit example:

- class labels: {good, bad}
 - 4 $P(\text{good}) = 0.7$
 - 4 $P(\text{bad}) = 0.3$
- aggregates for housing
 - 4 $P(\text{own} | \text{bad}) = 0.62$
 - 4 $P(\text{own} | \text{good}) = 0.75$
 - 4 $P(\text{rent} | \text{bad}) = 0.23$
 - 4 $P(\text{rent} | \text{good}) = 0.14$
 - 4 ... and so on

- Assign the label that maximizes the value

$$\prod_i^m P(a_i|b_j)P(b_j)$$

Credit Example: X

- female
- owns home
- Self-employed
- savings > \$1000

$P(\text{good}|X) > P(\text{bad}|X)$:
Assign X the label "good"

$$P(\text{good}|X) \sim (0.28 \cdot 0.75 \cdot 0.14 \cdot 0.06) \cdot 0.7 = 0.0012$$

$$P(\text{bad}|X) \sim (0.36 \cdot 0.62 \cdot 0.17 \cdot 0.02) \cdot 0.3 = 0.0002$$

a_i	b_j	$P(a_i b_j)$
female	good	0.28
female	bad	0.36
own	good	0.75
own	bad	0.62
self emp	good	0.14
self emp	bad	0.17
savings>1K	good	0.06
savings>1K	bad	0.02



1. Consider the following Training Data Set:

- Apply the Naïve Bayesian Classifier to this data set and compute

$P(y = 1 | X)$ for $X = (1, 0, 0)$

Show your work m

$$\prod_i P(a_i | b_j) P(b_j)$$

Training Data Set

X1	X2	X3	Y
1	1	1	0
1	1	0	0
0	0	0	0
0	1	0	1
1	0	1	1
0	1	1	1

$$P(b_j | a_1, a_2, \dots, a_m) = \frac{\prod_i P(a_i | b_j) P(b_j)}{P(a_1, a_2, \dots, a_m)}$$

Examples – NB Algorithm

?Play Tennis

Day	Outlook	Temperature	Humidity	Wind	Play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$P(\text{PlayTennis} = \text{yes}) = 9/14 = .64$$

$$P(\text{PlayTennis} = \text{no}) = 5/14 = .36$$

Outlook	Y	N		Humidity	Y	N
sunny	2/9	3/5		high	3/9	4/5
overcast	4/9	0		normal	6/9	1/5
rain	3/9	2/5				
Temperature				Wind		
hot	2/9	2/5		Strong	3/9	3/5
mild	4/9	2/5		Weak	6/9	2/5
cool	3/9	1/5				

(Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong)

$$P(\text{yes}|\text{New instance}) = P(\text{yes}) P(\text{sunny}|\text{yes}) P(\text{cool}|\text{yes}) P(\text{high}|\text{yes}) P(\text{strong}|\text{yes}) = .0053$$

$$P(\text{No}|\text{New instance}) = P(\text{no}) P(\text{sunny}|\text{no}) P(\text{cool}|\text{no}) P(\text{high}|\text{no}) P(\text{strong}|\text{no}) = .0206$$

$P(\text{No}|\text{New instance}) > P(\text{yes}|\text{New instance})$

Hence the new instance belongs to play tennis No

Class:

C1:buys_computer = 'yes'

C2:buys_computer = 'no'

Data to be classified:

X = (age <=30,

Income = medium,

Student = yes

Credit_rating = Fair)

age	income	student	credit_rating	com
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

Naïve Bayes Classifier: An Example

- $P(C_i)$: $P(\text{buys_computer} = \text{"yes"}) = 9/14 = 0.643$
 $P(\text{buys_computer} = \text{"no"}) = 5/14 = 0.357$
- Compute $P(X|C_i)$ for each class
 - $P(\text{age} = \text{"<=30"} | \text{buys_computer} = \text{"yes"}) = 2/9 = 0.222$
 - $P(\text{age} = \text{"<= 30"} | \text{buys_computer} = \text{"no"}) = 3/5 = 0.6$
 - $P(\text{income} = \text{"medium"} | \text{buys_computer} = \text{"yes"}) = 4/9 = 0.444$
 - $P(\text{income} = \text{"medium"} | \text{buys_computer} = \text{"no"}) = 2/5 = 0.4$
 - $P(\text{student} = \text{"yes"} | \text{buys_computer} = \text{"yes"}) = 6/9 = 0.667$
 - $P(\text{student} = \text{"yes"} | \text{buys_computer} = \text{"no"}) = 1/5 = 0.2$
 - $P(\text{credit_rating} = \text{"fair"} | \text{buys_computer} = \text{"yes"}) = 6/9 = 0.667$
 - $P(\text{credit_rating} = \text{"fair"} | \text{buys_computer} = \text{"no"}) = 2/5 = 0.4$
- **$X = (\text{age} \leq 30, \text{income} = \text{medium}, \text{student} = \text{yes}, \text{credit_rating} = \text{fair})$**
 - $P(X|C_i) : P(X | \text{buys_computer} = \text{"yes"}) = 0.222 \times 0.444 \times 0.667 \times 0.667 = 0.044$
 - $P(X | \text{buys_computer} = \text{"no"}) = 0.6 \times 0.4 \times 0.2 \times 0.4 = 0.019$
 - $P(X|C_i) * P(C_i) : P(X | \text{buys_computer} = \text{"yes"}) * P(\text{buys_computer} = \text{"yes"}) = 0.028$
 - $P(X | \text{buys_computer} = \text{"no"}) * P(\text{buys_computer} = \text{"no"}) = 0.007$

Therefore, X belongs to class ("buys_computer = yes")

age	income	student	credit_rating	com
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

Prediction using NB

- Estimate conditional probabilities of each attributes {color, legs, height, smelly} for the species classes: {M, H} using the data given in the table.
- Using these probabilities estimate the probability values for the new instance – (Color=Green, legs=2, Height=Tall, and Smelly=No).

No	Color	Legs	Height	Smelly	Species
1	White	3	Short	Yes	M
2	Green	2	Tall	No	M
3	Green	3	Short	Yes	M
4	White	3	Short	Yes	M
5	Green	2	Short	No	H
6	White	2	Tall	No	H
7	White	2	Tall	No	H
8	White	2	Short	Yes	H

Naïve Bayes Classifier: Comments

- Advantages
 - Easy to implement
 - Good results obtained in most of the cases
- Disadvantages
 - Assumption: class conditional independence, therefore loss of accuracy
 - Practically, dependencies exist among variables
 - E.g., hospitals: patients: Profile: age, family history, etc.
Symptoms: fever, cough etc., Disease: lung cancer, diabetes, etc.
 - Dependencies among these cannot be modeled by Naïve Bayes Classifier
- How to deal with these dependencies? Bayesian Belief Networks

Naïve Bayesian Classifier - Reasons to Choose (+) and Cautions (-)

Reasons to Choose (+)	Cautions (-)
Handles missing values quite well	Numeric variables have to be discrete (categorized) Intervals
Robust to irrelevant variables	Sensitive to correlated variables
Easy to implement	Not good for estimating probabilities Stick to class label or yes/no
Easy to score data	
Computationally efficient Handles very high dimensional problems Handles categorical variables with a lot of levels	



Thank You