CS 484
Introduction to Machine Learning

Week 9, March 18, 2021

Spring Semester 2021

ILLINOIS TECH

College of Computing

Week 9 Agenda: Naïve Bayes

DAG

Directed Acyclic Graph

Bayesian Network

$p(B|A)$

naïve

Naïve Bayes

Directed Acyclic Graph

Graph

A visual tool for displaying the assumed relationship among variables

Node

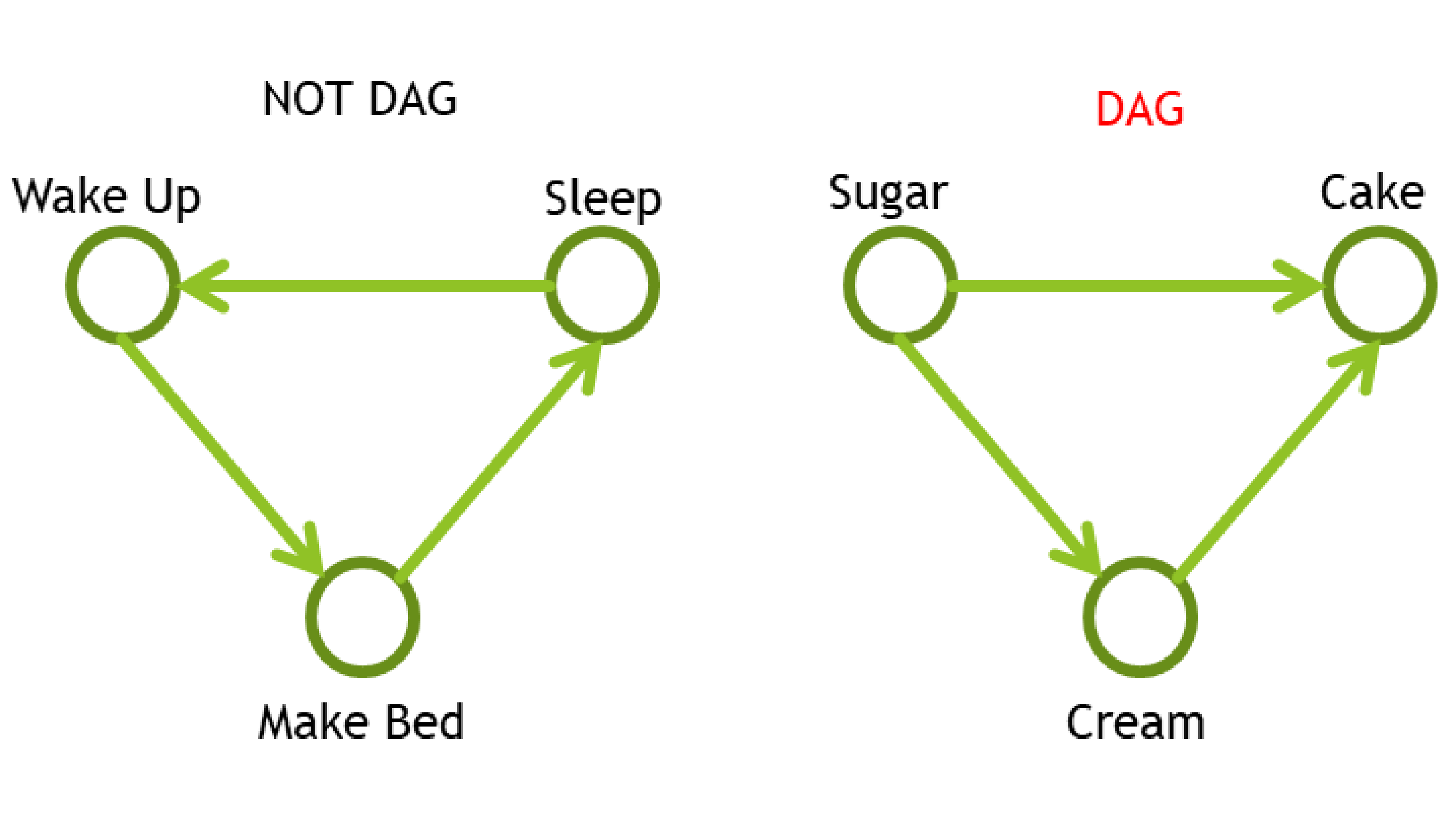
The variables are called nodes in the context of graphs

Edge

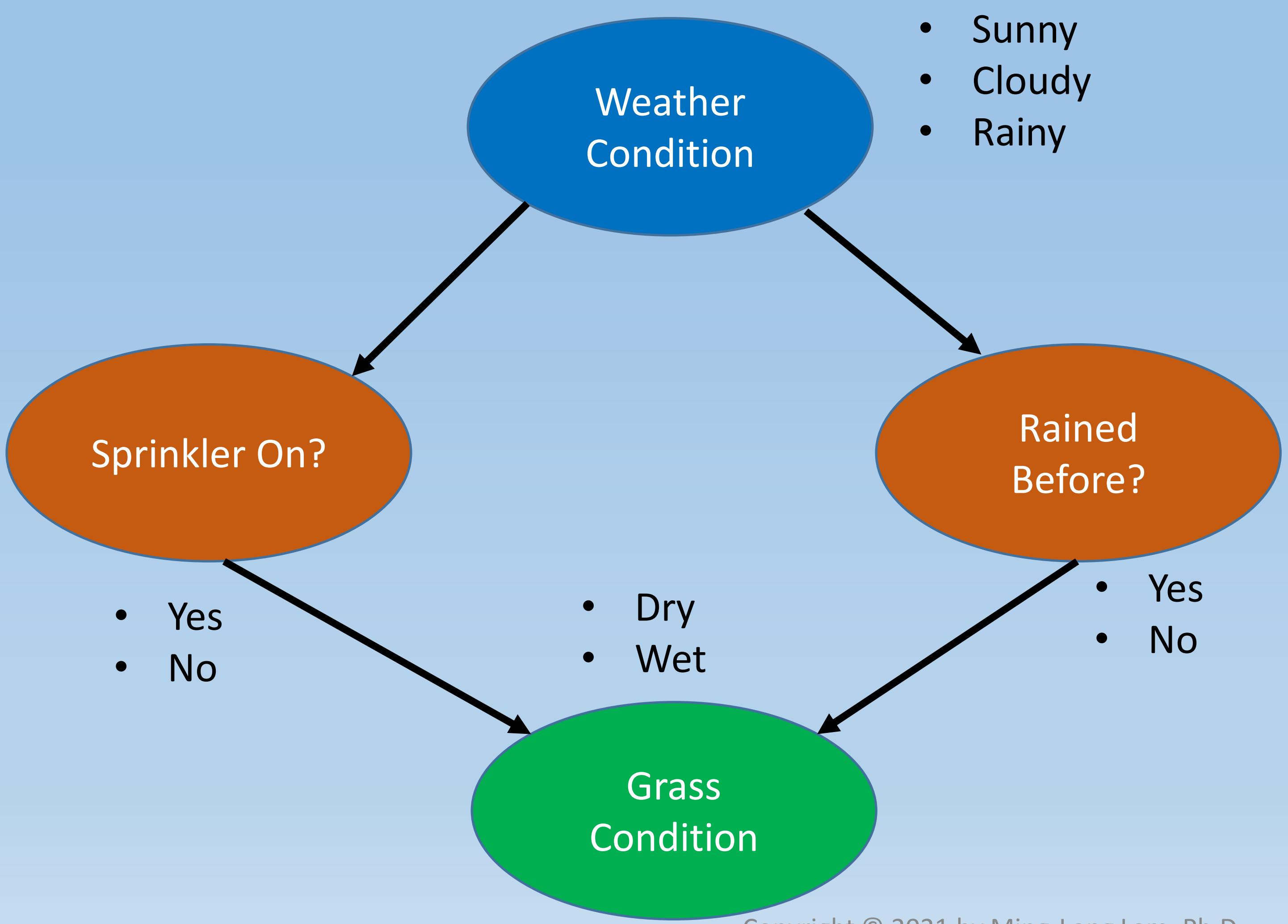
Represent the assumed causal relationships between two variables

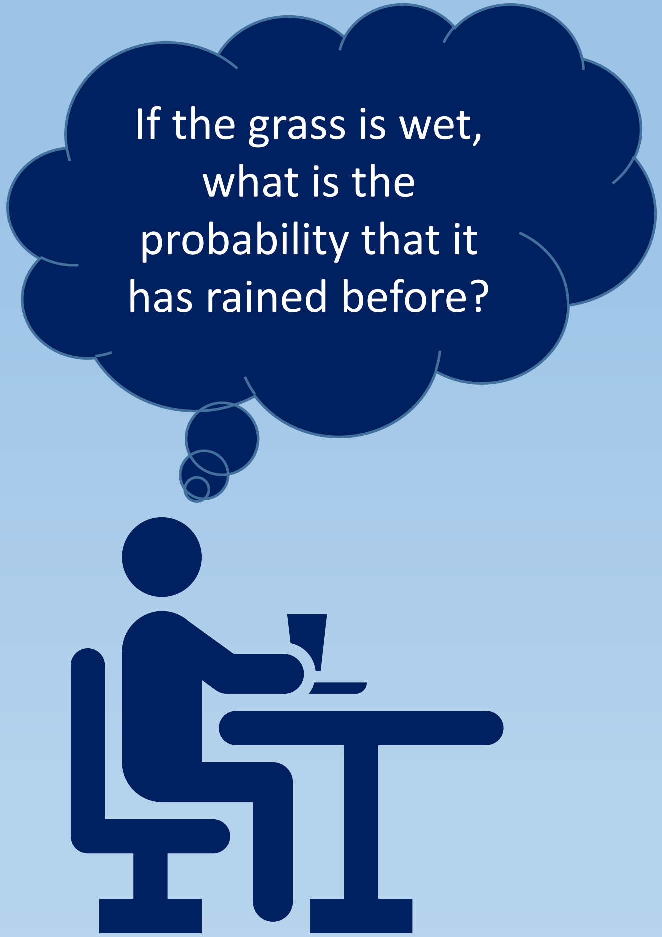
Directed Acyclic Graph

- The edges are acyclic (i.e., not forming part of a cycle)
- The causal relationships are assumed one-directional
- No feedback relationships

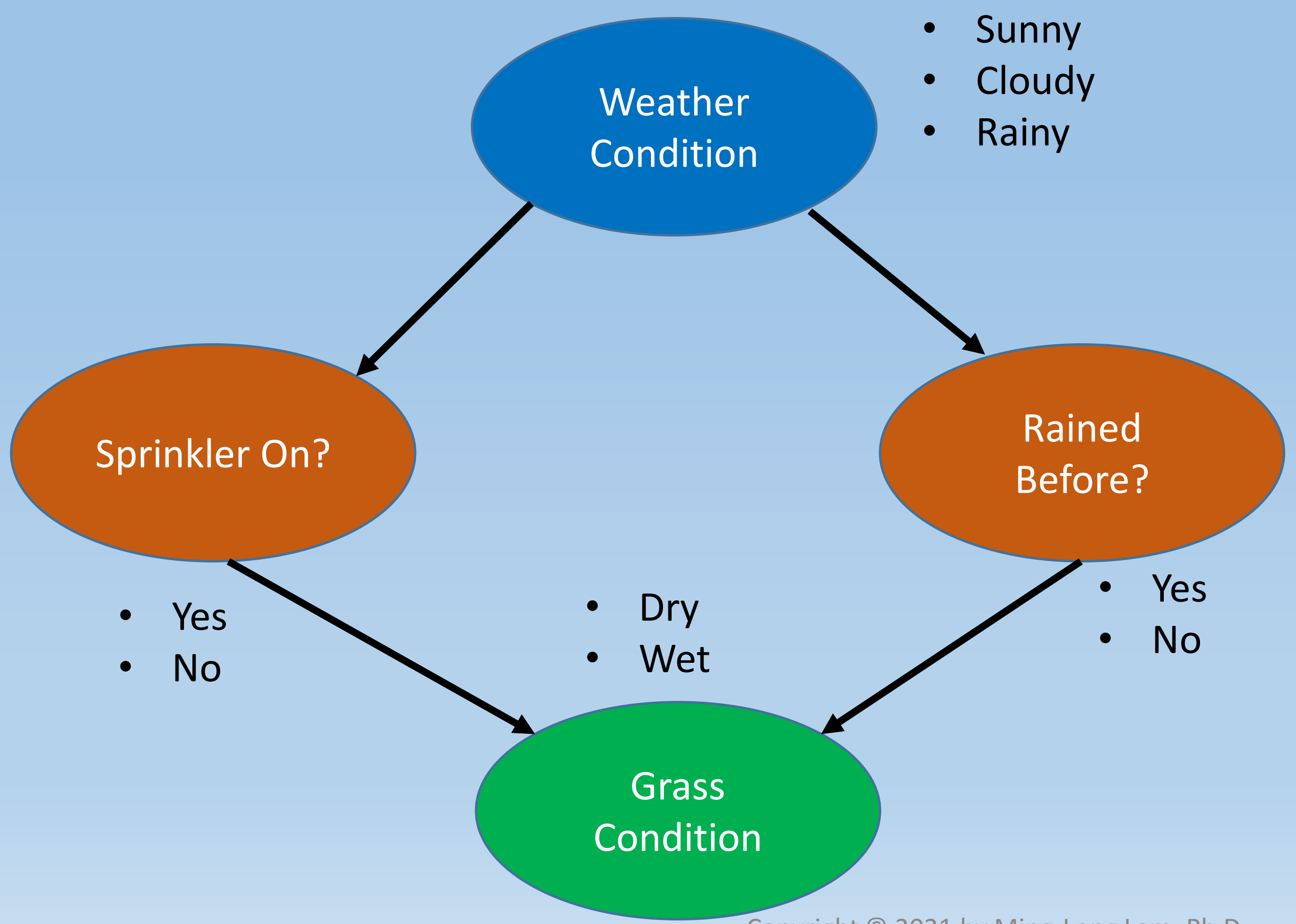


Directed Acyclic Graph





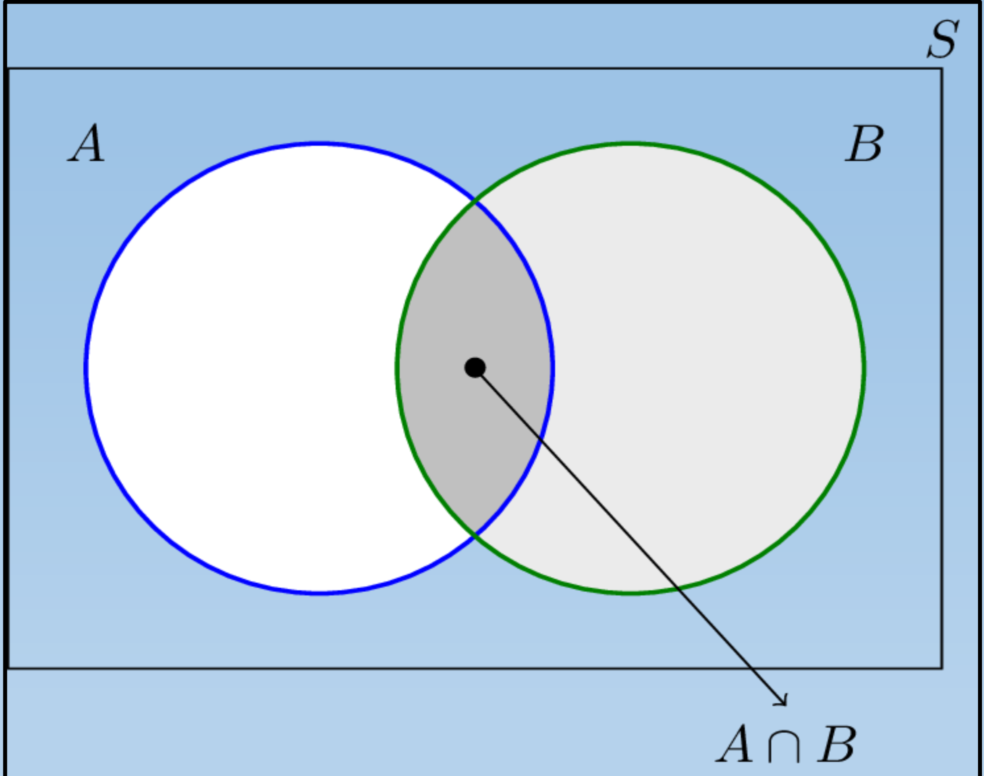
Directed Acyclic Graph



- *Weather Condition* affects *Sprinkler On* and *Rained Before*
- *Sprinkler On* and *Rained Before* each individually affects *Grass Condition*
- There is no relationship between *Sprinkler On* and *Rained Before* given *Weather Condition*

Conditional Probability

- Given two events A and B
- The conditional probability, denoted as Pr(A | B), is the probability that Event A will occur provided that Event B has occurred
- $\Pr(A \cap B)$ is the probability that both events A and B will occur
- $\Pr(B)$ is the probability that event B will occur



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

Conditional Probability

Roll a Fair Dice

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- A = {Odd Value} = {1, 3, 5}
- B = {Value Divisible by 3} = {3, 6}
- $A \cap B = \{3\}$
- $\Pr(A \cap B) = 1 / 6$
- $\Pr(A) = 3 / 6$
- $\Pr(B|A) = (1/6) / (3/6) = 1/3$

Pick a U(0,1) Random Number

- $\Omega = \{x: 0 \leq x \leq 1\}$
- $A = \{2x < 1\} = \{x: 0 \leq x < 0.5\}$
- $B = \{3x > 1\} = \{x: 1/3 < x \leq 1\}$
- $A \cap B = \{x: 1/3 < x < 1/2\}$
- $\Pr(A \cap B) = 1/2 - 1/3 = 1 / 6$
- $\Pr(A) = 1 / 2$
- $\Pr(B|A) = (1/6) / (1 / 2) = 1/3$

Bayes' Theorem

- $Pr(B|A) = Pr(A \cap B) / Pr(A)$

$\rightarrow$ $Pr(A \cap B) = Pr(B|A)\ Pr(A)$

- $Pr(A|B) = Pr(A \cap B) / Pr(B)$

$\rightarrow$ $Pr(A \cap B) = Pr(A|B)\ Pr(B)$

$\rightarrow$ $Pr(A|B)\ Pr(B) = Pr(B|A)\ Pr(A)$

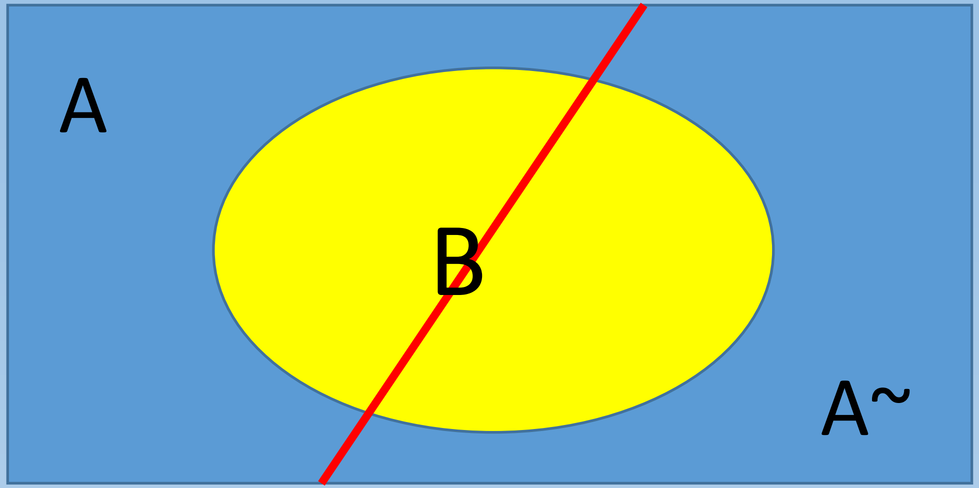
$\rightarrow$ $Pr(A|B) = (Pr(B|A)\ Pr(A)) / Pr(B)$

Roll a Fair Dice

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- A = {Odd Value} = {1, 3, 5}
- B = {Value Divisible by 3} = {3, 6}
- $Pr(B|A) = (1/6) / (3/6) = 1/3$
- $Pr(A|B) = (1/3 * 3/6) / (2/6) = 1/2$

Bayes' Theorem

- $\Pr(A|B) = (\Pr(B|A)\Pr(A)) / \Pr(B)$
- Do we have to calculate Pr(B) explicitly?
- $\Pr(B) = \Pr(B \cap \Omega)$ where Ω is the universal set
- $\Omega = A \cup (\sim A)$ where ~A is the complement set of A (i.e., everything but not in A)
- $B \cap \Omega = B \cap (A \cup \sim A)$



- $B \cap \Omega = B \cap (A \cup \sim A)$
- $B = (B \cap A) \cup (B \cap \sim A)$
- $(B \cap A)$ and $(B \cap \sim A)$ are disjoint
- $\Pr(B) = \Pr(B \cap A) + \Pr(B \cap \sim A)$

Bayes' Theorem

- $\Pr(B) = \Pr(B \cap A) + \Pr(B \cap {\sim}A)$
- $\Pr(B \cap A) = \Pr(B|A)\ P(A)$
- $\Pr(B \cap {\sim}A) = \Pr(B|{\sim}A)\ P({\sim}A)$
- $\Pr(B) = \Pr(B|A)\ P(A) + \Pr(B|{\sim}A)\ P({\sim}A)$

- $\Pr(A|B) = (\ \Pr(B|A)\ \Pr(A)\ )\ /\ (\ \Pr(B|A)\ P(A) + \Pr(B|{\sim}A)\ P({\sim}A)\ )$
- $P({\sim}A) = 1 - P(A)$

Bayesian Network

- A Bayesian network is a directed acyclic graphical model
- It represents (1) probability relationships, and (2) conditional independence structure among the random variables

- A Bayesian network is a family of classification algorithms for
  - Naïve Bayes
  - Tree-augmented Naïve Bayes (TAN)
  - Parent-child Bayesian Network
  - Markov Blanket

Bayesian Network Example

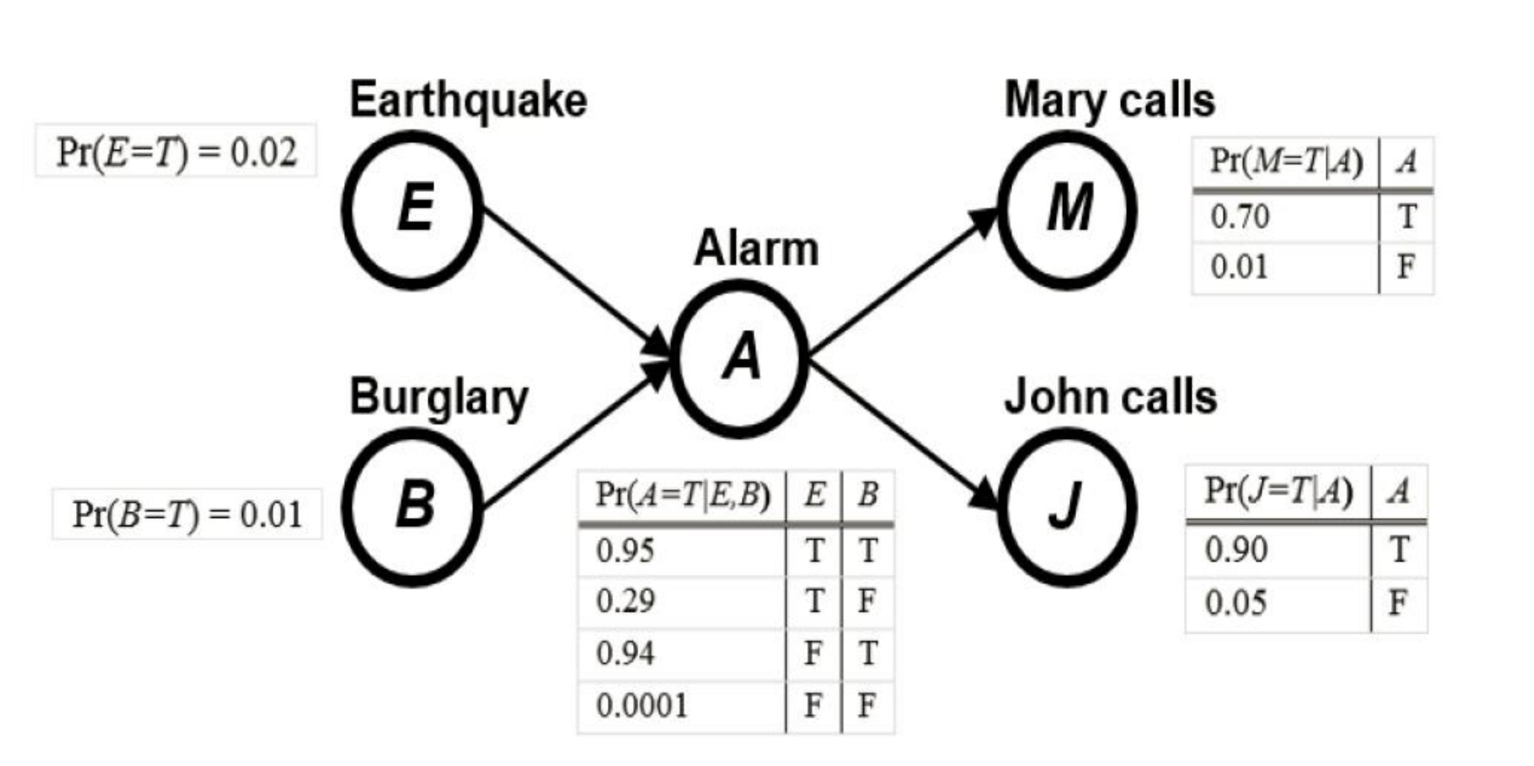
- Russell and Norvig (2010). *Artificial Intelligence: A Modern Approach*, Third Edition. New Jersey: Pearson.
- Suppose you live in an area (e.g., San Francisco Bay) where earthquakes are not uncommon

Bayesian Network Example

Your house has an alarm system against burglary, and the alarm system can be set off occasionally by an earthquake.

You have two neighbors, Mary and John, who do not know each other. If they hear the alarm, they might or might not call you.

Bayesian Network Example

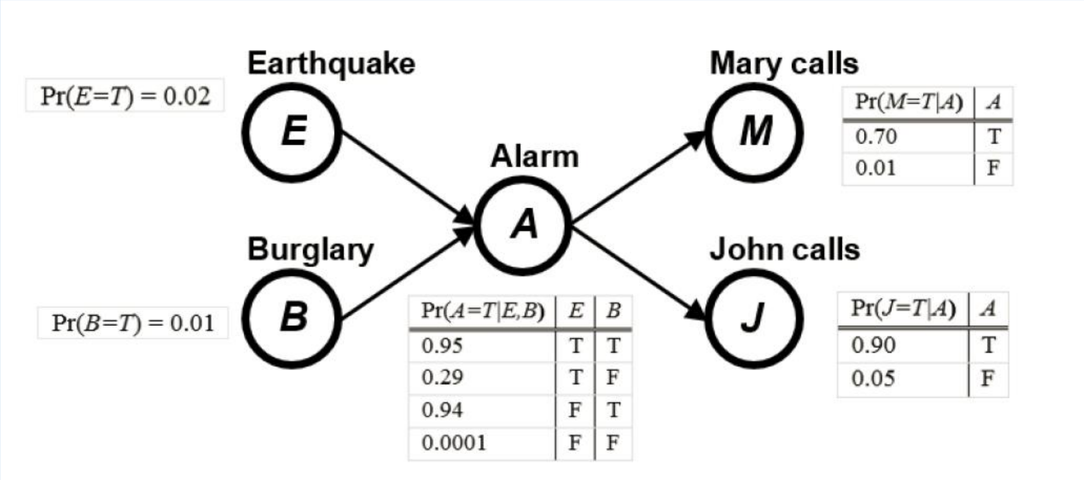


The events are

1. Has Burglary?
2. Has Earthquake?
3. Did Alarm Sound?
4. Did John call?
5. Did Mary call?

Bayesian Network Example

- The probabilities are either assigned or observed.
- Earthquake and Burglary are assumed independent
- Mary and John independently decide whether to call you
- Whether Mary or John calls is conditionally dependent only on the state of the alarm.
  - The purpose of their calls is to inform you that your alarm has sounded, and not to tell you the cause.
  - You are responsible to find out the cause.



Bayesian Network Example

The joint probability of the events (B, E, A, J, and M) is
$\Pr(B, E, A, J, M)$

The Bayes' Theorem:
$\Pr(B, E, A, J, M)$
$= \Pr(J, M | B, E, A) * \Pr(B, E, A)$

Since Mary and John are assumed independent:
$\Pr(J, M | B, E, A)$
$= \Pr(J | B, E, A) \times \Pr(M | B, E, A)$

Bayesian Network:
$\Pr(J | B, E, A) = \Pr(J | A)$
$\Pr(M | B, E, A) = \Pr(M | A)$

Bayesian Network Example

Bayes' Theorem:
$\Pr(B, E, A)$
$= \Pr(A \mid B, E) * \Pr(B, E)$

Burglary and Earthquake are assumed independent:
$\Pr(B, E) = \Pr(B) * \Pr(E)$.

Finally, $\Pr(B, E, A, J, M) = \Pr(J, M \mid B, E, A) \times \Pr(B, E, A)$
$= \Pr(J \mid B, E, A) \times \Pr(M \mid B, E, A) \times \Pr(B, E, A)$
$= \Pr(J \mid A) \times \Pr(M \mid A) \times \Pr(A \mid B, E) \times \Pr(B) \times \Pr(E)$

Bayesian Network Example

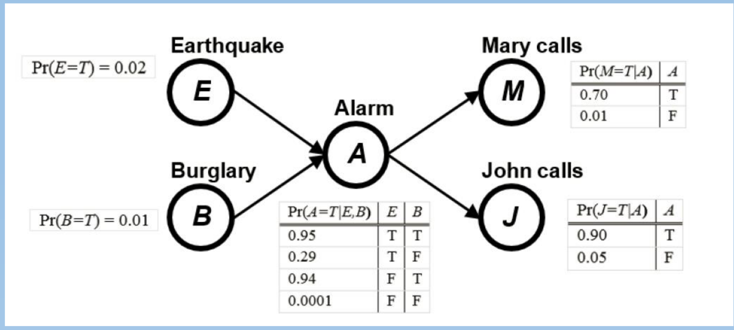
- The network structure together with the conditional probability distributions completely determines the Bayesian network model.

$$\begin{aligned}\text{Finally, } \Pr(B, E, A, J, M) &= \Pr(J, M \mid B, E, A) \times \Pr(B, E, A) \\ &= \Pr(J \mid B, E, A) \times \Pr(M \mid B, E, A) \times \Pr(B, E, A) \\ &= \Pr(J \mid A) \times \Pr(M \mid A) \times \Pr(A \mid B, E) \times \Pr(B) \times \Pr(E)\end{aligned}$$

Bayesian Network Example: Scenario 1

- Suppose you are at work:
  - The house is being burglarized (B = True)
  - There is no earthquake (E = False)
  - Neither John nor Mary calls (J = False and M = False)
- What is the probability that the alarm went off (A = True)?
- The probability is $Pr(A = T | B = T, E = F, J = F, M = F)$
  $= Pr(A | B, \sim E, \sim J, \sim M)$
  $= Pr(B, \sim E, A, \sim J, \sim M) \,/\, Pr(B, \sim E, \sim J, \sim M)$

Bayesian Network Example: Scenario 1



- Pr(B, ~E, A, ~J, ~M)
  = Pr(~J|A) × Pr(~M|A) × Pr(A|B, ~E) × Pr(B) × Pr(~E)
  = (1 – 0.9) × (1 - 0.7) × (0.94) × (0.01) × (1 – 0.02)
  = 0.000276360
- Pr(B, ~E, ~A, ~J, ~M)
  = Pr(~J|~A) × Pr(~M|~A) × Pr(~A|B, ~E) × Pr(B) × Pr(~E)
  = (1 – 0.05) × (1 - 0.01) × (1 – 0.94) × (0.01) × (1 – 0.02)
  = 0.000553014
- Pr(B, ~E, ~J, ~M) = Pr(B, ~E, A, ~J, ~M) + Pr(B, ~E, ~A, ~J, ~M)
  = 0.000276360 + 0.000553014 = 0.000829374

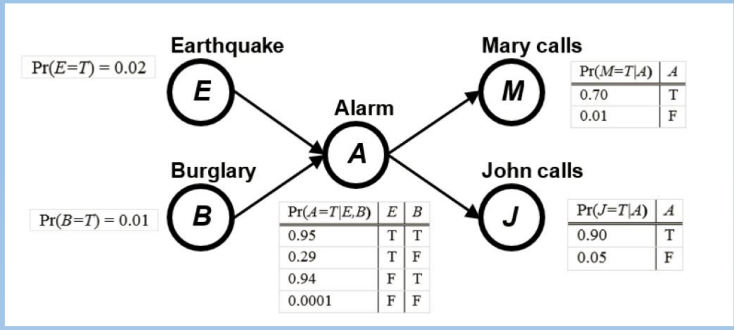
Bayesian Network Example: Scenario 1

- Suppose you are at work, the house is being burglarized (B = True), there is no earthquake (E = False), neither John nor Mary calls to say your alarm is ringing (J = False and M = False).
- The probability that the alarm went off (A = True) is
  $\Pr(A = T \mid B = T, E = F, J = F, M = F)$
  $= \Pr(A \mid B, \sim E, \sim J, \sim M)$
  $= \Pr(B, \sim E, A, \sim J, \sim M) / \Pr(B, \sim E, \sim J, \sim M)$
  $= 0.000276360 / 0.000829374 = 0.333215172$
- In summary, the conditional probability of the alarm having gone off in this situation is about 0.33.

Bayesian Network Example: Scenario 2

- Suppose you are at work:
  - The house is burglarized (B = True)
  - There is no earthquake (E = False)
  - Mary called to say your alarm is ringing (M = True)
  - John did not call (J = False)
- What is the probability that the alarm went off (A = True)?
- The probability is $\Pr(A = T | B = T, E = F, J = F, M = T)$
  $= \Pr(A | B, \sim E, \sim J, M)$
  $= \Pr(B, \sim E, A, \sim J, M) / \Pr(B, \sim E, \sim J, M)$

Bayesian Network Example: Scenario 2



- Pr(B, ~E, A, ~J, M)
  = Pr(~J|A) × Pr(M|A) × Pr(A|B, ~E) × Pr(B) * Pr(~E)
  = (1 – 0.9) × (0.7) × (0.94) × (0.01) × (1 – 0.02)
  = 0.000644840
- Pr(B, ~E, ~A, ~J, M)
  = Pr(~J|~A) × Pr(M|~A) × Pr(~A|B, ~E) × Pr(B) × Pr(~E)
  = (1 – 0.05) × 0.01 × (1 – 0.94) × (0.01) × (1 – 0.02)
  = 0.000005586
- Pr(B, ~E, ~J, M) = Pr(B, ~E, A, ~J, M) + Pr(B, ~E, ~A, ~J, M)
  = 0.00064484 + 0.000005586 = 0.000650426

Bayesian Network Example: Scenario 2

- Suppose you are at work, the house is burglarized (B = True), there is no earthquake (E = False), your neighbor Mary calls to say your alarm is ringing (M = True), but neighbor John doesn't call (J = False). What is the probability that the alarm went off (A = True)?
- The probability is $\Pr(A = T | B = T, E = F, J = F, M = T) = \Pr(A | B, {\sim}E, {\sim}J, M)$
$= \Pr(B, {\sim}E, A, {\sim}J, M) / \Pr(B, {\sim}E, {\sim}J, M)$
$= 0.000644840 / 0.000650426 = 0.991411782$
- In summary, the conditional probability of the alarm having gone off in this situation is about 0.99.

Bayesian Network Example: Pr(A = True)

- What is the overall probability that the alarm went off (A = True)?
- The probability is Pr(A) = P(A,B,E) + P(A,~B,E) + P(A,B,~E) + P(A,~B,~E)
- = Pr(A|B,E) × Pr(B,E) + Pr(A|~B,E) × Pr(~B,E)
  + Pr(A|B,~E) × Pr(B,~E) + Pr(A|~B,~E) × Pr(~B,~E)

  = Pr(A|B,E) × Pr(B) × Pr(E) + Pr(A|~B,E) × Pr(~B) × Pr(E)
  + Pr(A|B,~E) × Pr(B) × Pr(~E) + Pr(A|~B,~E) × Pr(~B) × Pr(~E)

  = 0.95 × 0.01 × 0.02 + 0.29 × (1-0.01) × 0.02
  + 0.94 × 0.01 ×(1-0.02) + 0.0001 × (1-0.01) × (1-0.02) = 0.01524102

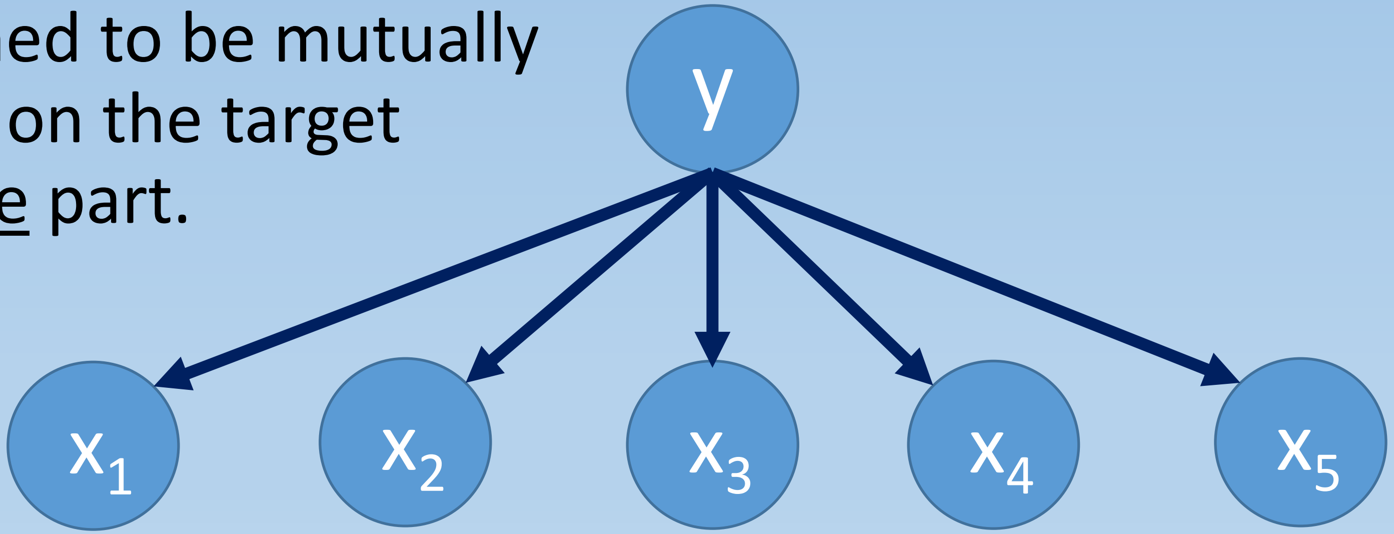
Bayesian Network Example: Summary

- Scenario 1: When Mary calls but John did not:
  - $\Pr(A = T | B = T, E = F, J = F, M = T) = 0.99 > 0.02 = \Pr(A = T)$
  - I can surely classify that the alarm did go off.
  - **Action**: I should then contact the police to check on my house.

- Scenario 2: When both Mary and John did not call:
  - $\Pr(A = T | B = T, E = F, J = F, M = F) = 0.33 > 0.02 = \Pr(A = T)$
  - This is not a negligible probability.
  - **Action**: I may consider subscribing to some monitoring services too.

Naïve Bayes Overview

- A Naïve Bayes is a particular Bayesian Network.
- There is an edge from the nominal target variable to each predictor
- Categorical or interval predictors are allowed.
- The predictors are assumed to be mutually independent conditional on the target variable. This is the Naïve part.



Naïve Bayes: Theory

- Denote the target variable as y.
- Denote the predictors as $x_1, \dots, x_p$.
- Our goal is to calculate the conditional probability of the target variable given the predictors. This is the Bayes part.

$$\Pr(y|x_1, \dots, x_p) = \frac{\Pr(y, x_1, \dots, x_p)}{\Pr(x_1, \dots, x_p)}$$

Naïve Bayes: Theory

- Applying the Bayes' Theorem,

$$\Pr(y, x_1, \ldots, x_p) = \Pr(y) \Pr(x_1, \ldots, x_p | y)$$

- Using the assumption that the predictors are mutually independent conditional on the target variable,

$$\Pr(y, x_1, \ldots, x_p) = \Pr(y) \Pr(x_1, \ldots, x_p | y) = \Pr(y) \prod_{j=1}^{p} \Pr(x_j | y)$$

- In other words, if we already knew the state of the target variable, the states of other predictors do not contribute any additional information about the state of the current predictor.

Naïve Bayes: Theory

- It follows that,

$$\Pr(y|x_1,\dots,x_p) = \frac{\Pr(y,x_1,\dots,x_p)}{\Pr(x_1,\dots,x_p)} = \frac{\Pr(y)\prod_{j=1}^{p}\Pr(x_j|y)}{\Pr(x_1,\dots,x_p)}.$$

- Although $Pr(x_1,\dots,\ x_p)$ is a probability, its value is fixed for a given data. Therefore, $\Pr(y|x_1,\dots,x_p) \propto \Pr(y)\prod_{j=1}^{p}\Pr(x_j|y)$.
- The probability $\Pr(y)$ is the **class probability** because y is categorical.

Naïve Bayes: Classifier

- Given values of $x_1, \dots, x_p$, we calculate this quantity $\Pr(y) \prod_{j=1}^{p} \Pr(x_j|y)$ (not necessary a valid probability value) for all possible categories of the target variable.
- Then divide these quantities by the sum of them to make the resulting values as valid probabilities values.
- Finally, select the category whose corresponding probability is the highest. Alternatively, select the lexically lowest category whose corresponding probability has exceeded a specified threshold.

Naïve Bayes: Representing $\mathbf{Pr}(\boldsymbol{x_j}|\boldsymbol{y})$

- Categorical Predictor
  - $\Pr(x_j|y)$ follows the empirical probability distribution.
- Interval Predictor
  - $\Pr(x_j|y)$ follows a univariate Gaussian (i.e., Normal) distribution
  - The mean and the variance of that distribution is estimated by the sample mean and the sample variance of x_j within each category of y.
- Count Predictor
  - $\Pr(x_j|y)$ follows a multinomial distribution.
  - The parameters of that distribution are estimated by the fractions of observations within each category of y.

Naïve Bayes: Customer Survey

You are working on a marketing campaign to promote the E-Billing service to bank customers.

You need to build the profiles of customers who will register for the E-Billing service.

You have access to a recent Customer Survey Data which contains information about 4,952 customers.

Naïve Bayes: Structure

Behavioral Theory:
A person’s aptitude (e.g., genes) to embrace E-Billing may also exemplifies in the person’s gender, the choice of credit card (spending vs. saving), and the career path (work/life balance).

Ebilling
Yes / No

CreditCard
American Express
Discover
MasterCard
Others
Visa

Gender
Female
Male

JobCategory
Agriculture
Crafts
Labor
Professional
Sales
Service

Naïve Bayes: Customer Survey

```
# Define a function to visualize the percent of a particular target category by a nominal predictor
def RowWithColumn (
   rowVar,          # Row variable
   columnVar,       # Column predictor
   show = 'ROW'):   # Show ROW fraction, COLUMN fraction, or BOTH table

   countTable = pandas.crosstab(index = rowVar, columns = columnVar, margins = False, dropna = True)
   print("Frequency Table: \n", countTable)
   print( )

   if (show == 'ROW' or show == 'BOTH'):
       rowFraction = countTable.div(countTable.sum(1), axis='index')
       print("Row Fraction Table: \n", rowFraction)
       print( )

   if (show == 'COLUMN' or show == 'BOTH'):
       columnFraction = countTable.div(countTable.sum(0), axis='columns')
       print("Column Fraction Table: \n", columnFraction)
       print( )

   return
```

Week 9 EBilling Naive Bayes.py

Customer Survey: Crosstabulation

EBilling	Count	Class Probability
No	3,221	0.6504
Yes	1,731	0.3496

Week 9 EBilling Naive Bayes.py

Count	Credit Card					
Ebilling	American Express	Discover	MasterCard	Others	Visa	Total
No	591	788	815	173	854	3,221
Yes	390	543	369	48	381	1,731

Row Fraction	Credit Card					
EBilling	American Express	Discover	MasterCard	Others	Visa	Total
No	0.1835	0.2446	0.2530	0.0537	0.2651	1.0000
Yes	0.2253	0.3137	0.2132	0.0277	0.2201	1.0000

Customer Survey: Crosstabulation

Count	Gender		
EBilling	**Female**	**Male**	**Total**
No	1,595	1,626	3,221
Yes	895	836	1,731

Row Fraction	Gender		
EBilling	**Female**	**Male**	**Total**
No	0.4952	**0.5048**	1.0000
Yes	**0.5170**	0.4830	1.0000

Week 9 EBilling Naive Bayes.py

Customer Survey: Crosstabulation

Week 9 EBilling Naive Bayes.py

Count	Job Category						
EBilling	**Agriculture**	**Crafts**	**Labor**	**Professional**	**Sales**	**Service**	**Total**
No	134	297	474	859	1,038	419	3,221
Yes	78	152	206	512	588	195	1,731

Row Fraction	Job Category						
EBilling	**Agriculture**	**Crafts**	**Labor**	**Professional**	**Sales**	**Service**	**Total**
No	0.0416	**0.0922**	**0.1472**	0.2667	0.3223	**0.1301**	1.0000
Yes	**0.0451**	0.0878	0.1190	**0.2958**	**0.3397**	0.1127	1.0000

Customer Survey: Conditional Probability

Conditional Probabilities of Ebilling = No given CreditCard, Gender, and JobCategory

$\Pr(\text{EBilling} = \text{No} \mid \text{CreditCard} = \text{American Express}, \text{Gender} = \text{Female}, \text{JobCategory} = \text{Professional})$

$\propto \Pr(\text{EBilling} = \text{No})$
$\times \Pr(\text{CreditCard} = \text{American Express} \mid \text{EBilling} = \text{No})$
$\times \Pr(\text{Gender} = \text{Female} \mid \text{EBilling} = \text{No})$
$\times \Pr(\text{JobCategory} = \text{Professional} \mid \text{EBilling} = \text{No})$

$= (3221/4952) \times (591/3221) \times (1595/3221) \times (859/3221) = 0.015760836$

EBilling	Count	Class Probability
No	3,221	0.6504
Yes	1,731	0.3496

Count	CreditCard					
Ebilling	American Express	Discover	MasterCard	Others	Visa	Total
No	591	788	815	173	854	3,221
Yes	390	543	369	48	381	1,731

Count	Gender		
EBilling	Female	Male	Total
No	1,595	1,626	3,221
Yes	895	836	1,731

Count	JobCategory						
EBilling	Agriculture	Crafts	Labor	Professional	Sales	Service	Total
No	134	297	474	859	1,038	419	3,221
Yes	78	152	206	512	588	195	1,731

Customer Survey: Conditional Probability

Conditional Probabilities of Ebilling = Yes given CreditCard, Gender, and JobCategory

Pr(EBilling = Yes | CreditCard = American Express, Gender = Female, JobCategory = Professional)

$\propto$ Pr(EBilling = Yes)
× Pr(CreditCard = American Express | EBilling = Yes)
× Pr(Gender = Female | EBilling = Yes)
× Pr(JobCategory = Professional | EBilling = Yes)

= (1731/4952) × (390/1731) × (895/1731) × (512/1731) = 0.012044335

EBilling	Count	Class Probability
No	3,221	0.6504
Yes	1,731	0.3496

Count	CreditCard					
Ebilling	American Express	Discover	MasterCard	Others	Visa	Total
No	591	788	815	173	854	3,221
Yes	390	543	369	48	381	1,731

Count	Gender		
EBilling	Female	Male	Total
No	1,595	1,626	3,221
Yes	895	836	1,731

Count	JobCategory						
EBilling	Agriculture	Crafts	Labor	Professional	Sales	Service	Total
No	134	297	474	859	1,038	419	3,221
Yes	78	152	206	512	588	195	1,731

Customer Survey: Conditional Probability

Recap the results

- Pr(EBilling = **No**|CreditCard = American Express, Gender = Female, JobCategory = Professional) $\propto$ 0.015760836

  ➔ Pr(EBilling = **No**|CreditCard = American Express, Gender = Female, JobCategory = Professional) = C * 0.015760836 where C is the proportional constant

- Pr(EBilling = **Yes**|CreditCard = American Express, Gender = Female, JobCategory = Professional) $\propto$ 0.012044335

  ➔ Pr(EBilling = **Yes**|CreditCard = American Express, Gender = Female, JobCategory = Professional) = C * 0.012044335 where C is the proportional constant

Customer Survey: Conditional Probability

Recap the results

- Since Ebilling is either **No** or **Yes**, then

  Pr(EBilling = **No**|CreditCard = American Express, Gender = Female, JobCategory = Professional) +
  Pr(EBilling = **Yes**|CreditCard = American Express, Gender = Female, JobCategory = Professional)
  = 1

- Therefore 1 = C * 0.015760836 + C * 0.01 2044335 = C * 0.027805171
- Hence, C = 1 / 0.027805171.

Customer Survey: Conditional Probability

Convert to Valid Probability Values

- Put C = 1 / 0.027805171
- Pr(EBilling = No|CreditCard = American Express, Gender = Female, JobCategory = Professional)
  = C * 0.015760836 = 0.015760836 / 0.027805171 = 0.566831107
- Pr(EBilling = Yes|CreditCard = American Express, Gender = Female, JobCategory = Professional)
  = C * 0.012044335 = 0.012044335 / 0.027805171 = 0.433168893

Caution About naive_bayes.BernoulliNB

Since the naive_bayes.BernoulliNB function can handle binary features that are coded 0 or 1

↓

Pandas.GetDummies can create 0/1 indicator variables for categorical features

↓

Can we use the BernoulliNB function for categorical features?

- The short answer is **NO**.
- The dummy indicator variables are not functionally independent.
- Since Naïve Bayes will treat each indicator variable as an actual independent variable, it will bring it extraneous probability $\Pr(x_j|y)$ into the calculation.
- The resulting predicted probability from BernoulliNB will be different from that by treating the feature as a categorical variable.

Instead Use naive_bayes. CategoricalNB

Training vectors X

- Assume each feature of X is from a different categorical distribution.
- Require that all categories of each feature are represented by integers 0, …, n - 1, where n is the total number of categories of a feature.
- This can be achieved with the help of OrdinalEncoder.

Target vector y

- Although documentation does not say whether integers 0, 1, … is required, we will use LabelEncoder too.

Customer Survey: Naïve Bayes

```
from sklearn import preprocessing, naive_bayes

labelEnc = preprocessing.LabelEncoder()
yTrain = labelEnc.fit_transform(subData['EBilling'])
yLabel = labelEnc.inverse_transform([0, 1])

uCreditCard = numpy.unique(subData['CreditCard'])
uGender = numpy.unique(subData['Gender'])
uJobCategory = numpy.unique(subData['JobCategory'])

featureCategory = [uCreditCard, uGender, uJobCategory]
featureEnc = preprocessing.OrdinalEncoder(categories = featureCategory)
xTrain = featureEnc.fit_transform(subData[['CreditCard', 'Gender', 'JobCategory']])

_objNB = naive_bayes.CategoricalNB(alpha = 1.0e-10)
thisModel = _objNB.fit(xTrain, yTrain)
```

Customer Survey: Naïve Bayes

```
print('Number of samples encountered for each class during fitting')
print(yLabel)
print(_objNB.class_count_)
print('\n')

print('Probability of each class:')
print(yLabel)
print(numpy.exp(_objNB.class_log_prior_))
print('\n')
```

```
Number of samples encountered for each class during fitting
['No' 'Yes']
[3221. 1731.]

Probability of each class:
['No' 'Yes']
[0.65044426 0.34955574]
```

Customer Survey: Naïve Bayes

```
feature = ['CreditCard', 'Gender', 'JobCategory']
print('Number of samples encountered for each (class, feature) during fitting')
for i in range(3):
   print('Feature: ', feature[i])
   print(featureCategory[i])
   print(_objNB.category_count_[i])
   print('\n')

print('Empirical probability of features given a class, P(x_i|y)')
for i in range(3):
   print('Feature: ', feature[i])
   print(featureCategory[i])
   print(numpy.exp(_objNB.feature_log_prob_[i]))
   print('\n')
```

Customer Survey: Naïve Bayes

```
Number of samples encountered for each (class, feature) during fitting
Feature:  CreditCard
['American Express' 'Discover' 'MasterCard' 'Others' 'Visa']
[[591. 788. 815. 173. 854.]
 [390. 543. 369.  48. 381.]]

Feature:  Gender
['Female' 'Male']
[[1595. 1626.]
 [ 895.  836.]]

Feature:  JobCategory
['Agriculture' 'Crafts' 'Labor' 'Professional' 'Sales' 'Service']
[[ 134.  297.  474.  859. 1038.  419.]
 [  78.  152.  206.  512.  588.  195.]]
```

Customer Survey: Naïve Bayes

```
Empirical probability of features given a class, P(x_i|y)
Feature:  CreditCard
['American Express' 'Discover' 'MasterCard' 'Others' 'Visa']
[[0.18348339 0.24464452 0.25302701 0.05371003 0.26513505]
 [0.22530329 0.31369151 0.21317158 0.02772964 0.22010399]]

Feature:  Gender
['Female' 'Male']
[[0.49518783 0.50481217]
 [0.51704217 0.48295783]]

Feature:  JobCategory
['Agriculture' 'Crafts' 'Labor' 'Professional' 'Sales' 'Service']
[[0.04160199 0.09220739 0.14715927 0.26668736 0.32226017 0.13008382]
 [0.04506066 0.08781051 0.11900635 0.29578278 0.33968804 0.11265165]]
```

Customer Survey: Naïve Bayes

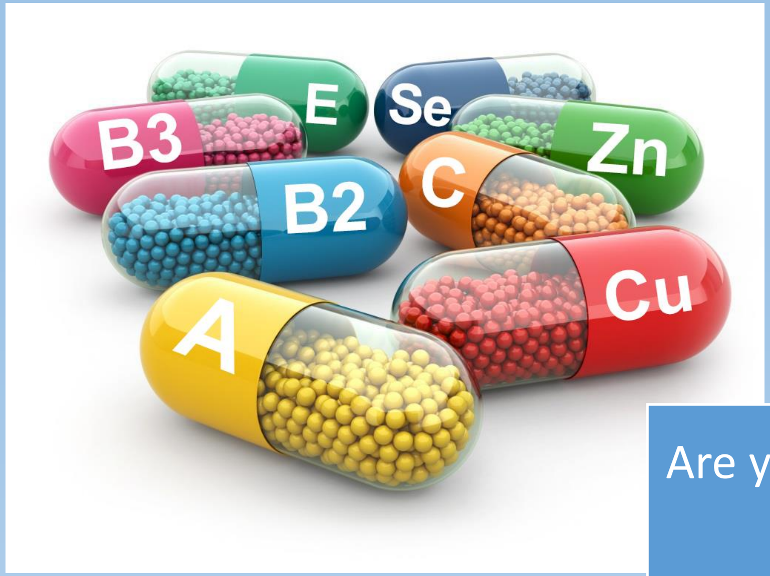
```
# CreditCard = American Express, Gender = Female, JobCategory = Professional
xTest = featureEnc.transform([['American Express', 'Female', 'Professional']])

y_predProb = thisModel.predict_proba(xTest)
print('Predicted Probability: ', yLabel, y_predProb)
```

```
Predicted Probability:  ['No' 'Yes'] [[0.56683111 0.43316889]]
```

- C = 1 / 0.027805171
- Pr(EBilling = No|CreditCard = American Express, Gender = Female, JobCategory = Professional) = C * 0.015760836 = 0.015760836 / 0.027805171 = 0.566831107
- Pr(EBilling = Yes|CreditCard = American Express, Gender = Female, JobCategory = Professional) = C * 0.012044335 = 0.012044335 / 0.027805171 = 0.433168893

Nutrition Information Study



Are you taking any dietary supplements?

1 = Yes
2 = No

Is TV a primary source of information about nutrition?

1 = Yes
2 = No

Are magazines a primary source of information about nutrition?

1 = Yes
2 = No

Are friends a primary source of information about nutrition?

1 = Yes
2 = No

Is your doctor a primary source of information about nutrition?

1 = Yes
2 = No

Source: McKay, D. L., Houser, R. F., Blumberg, J. B., Goldberg, J. P. (2006). Nutrition information sources vary with education level in a population of older adults. *Journal of the American Dietetic Association*, 106, 1108-1111.

Week 9 Nutrition Naive Bayes.py

Nutrition Information: Binary Features

```
# Specify the roles
feature = ['tv', 'magazine', 'friends', 'doctor']
target = 'supps'

# Read the Excel file
nutrition = pandas.read_excel('C:\\IIT\\Machine Learning\\Data\\Nutrition_Information.xls',
                              sheet_name = 'Sheet1',
                              usecols = feature + [target])
nutrition = nutrition.dropna()

# Look at the row distribution
print(nutrition.groupby(target).size())

for pred in feature:
    RowWithColumn(rowVar = nutrition[target], columnVar = nutrition[pred], show = 'ROW')
```

Week 9 Nutrition Naive Bayes.py

Naïve Bayes: Binary Features

supps	Count	Class Probability
Yes	66	0.38150289
No	107	0.61849711
Total	173	1.0

supps: Are you taking any dietary supplements?

tv: Is TV a primary source of information about nutrition?

magazine: Are magazines a primary source of information about nutrition?

friends: Are friends a primary source of information about nutrition?

doctor: Is your doctor a primary source of information about nutrition?

Week 9 Nutrition Naive Bayes.py

Count	tv		
supps	Yes	No	Total
Yes	34	32	66
No	51	56	107

Count	magazine		
supps	Yes	No	Total
Yes	42	24	66
No	62	45	107

Count	friends		
supps	Yes	No	Total
Yes	22	44	66
No	30	77	107

Count	doctor		
supps	Yes	No	Total
Yes	39	27	66
No	68	39	107

Nutrition Information: Conditional Probability

Conditional Probabilities of supps = Yes given tv, magazine, friends, and doctor

$$\Pr(\text{supps} = \text{Yes} \mid \text{tv} = \text{Yes}, \text{magazine} = \text{Yes}, \text{friends} = \text{Yes}, \text{doctor} = \text{Yes})$$

$$\propto \Pr(\text{supps} = \text{Yes}) \times \Pr(\text{tv} = \text{Yes} \mid \text{supps} = \text{Yes}) \times \Pr(\text{magazine} = \text{Yes} \mid \text{supps} = \text{Yes}) \times \Pr(\text{friends} = \text{Yes} \mid \text{supps} = \text{Yes}) \times \Pr(\text{doctor} = \text{Yes} \mid \text{supps} = \text{Yes})$$

$$= (66/173) \times (34/66) \times (42/66) \times (22/66) \times (39/66) = 0.024634150225322$$

supps	Count	Class Probability
Yes	66	0.38150289
No	107	0.61849711
Total	173	1.0

Count	tv		
supps	Yes	No	Total
Yes	34	32	66
No	51	56	107

Count	magazine		
supps	Yes	No	Total
Yes	42	24	66
No	62	45	107

Count	friends		
supps	Yes	No	Total
Yes	22	44	66
No	30	77	107

Count	doctor		
supps	Yes	No	Total
Yes	39	27	66
No	68	39	107

Nutrition Information: Conditional Probability

Conditional Probabilities of supps = No given tv, magazine, friends, and doctor

$\Pr(\text{supps} = \text{No} \mid \text{tv} = \text{Yes}, \text{magazine} = \text{Yes}, \text{friends} = \text{Yes}, \text{doctor} = \text{Yes})$

$$\propto \Pr(\text{supps} = \text{No}) \times \Pr(\text{tv} = \text{Yes} \mid \text{supps} = \text{No}) \times \Pr(\text{magazine} = \text{Yes} \mid \text{supps} = \text{No}) \times \Pr(\text{friends} = \text{Yes} \mid \text{supps} = \text{No}) \times \Pr(\text{doctor} = \text{Yes} \mid \text{supps} = \text{No})$$

$$= (107/173) \times (51/107) \times (62/107) \times (30/107) \times (68/107) = 0.030436492074722$$

supps	Count	Class Probability
Yes	66	0.38150289
No	107	0.61849711
Total	173	1.0

Count	tv		
supps	Yes	No	Total
Yes	34	32	66
No	51	56	107

Count	magazine		
supps	Yes	No	Total
Yes	42	24	66
No	62	45	107

Count	friends		
supps	Yes	No	Total
Yes	22	44	66
No	30	77	107

Count	doctor		
supps	Yes	No	Total
Yes	39	27	66
No	68	39	107

Nutrition Information: Conditional Probability

Recap the results

- Pr(supps = Yes | tv = Yes, magazine = Yes, friends = Yes, doctor = Yes) $\propto$ 0.0246341502253221
- Pr(supps = No | tv = Yes, magazine = Yes, friends = Yes, doctor = Yes) $\propto$ 0.030436492074722
- The sum is 0.0246341502253221 + 0.030436492074722 = 0.0550706423000441

Nutrition Information: Conditional Probability

Convert to Valid Probability Values

- Pr(supps = Yes | tv = Yes, magazine = Yes, friends = Yes, doctor = Yes)
  = 0.0246341502253221 / 0.0550706423000441
  = 0.4473191014
- Pr(supps = No | tv = Yes, magazine = Yes, friends = Yes, doctor = Yes)
  = 0.0304364920747220 / 0.0550706423000441
  = 0.5526808986

Nutrition Information: BernoulliNB

```
# Make the binary features take values 0 and 1 (was 2=No and 1=Yes)
nutrition[feature] = 2 - nutrition[feature]

xTrain = nutrition[feature].astype('category')
yTrain = nutrition[target].astype('category')

_objNB = naive_bayes.BernoulliNB(alpha = 1.e-10)
thisFit = _objNB.fit(xTrain, yTrain)

print('Probability of each class')
print(numpy.exp(thisFit.class_log_prior_))

print('Empirical probability of features given a class, P(x_i|y)')
print(numpy.exp(thisFit.feature_log_prob_))

print('Number of samples encountered for each class during fitting')
print(thisFit.class_count_)

print('Number of samples encountered for each (class, feature) during fitting')
print(thisFit.feature_count_)
```

The features must take only 0 or 1 values

Alpha is the additive (Laplace/Lidstone) smoothing parameter (0 for no smoothing). Ideally, we want alpha = 0, but this is as small as the function allows.

Nutrition Information: BernoulliNB

```
Probability of each class
supps = No  supps = Yes
[0.38150289 0.61849711]

Empirical probability of features given a class, P(x_i|y)
              tv         magazine   friends    doctor
supps = Yes [[0.51515152 0.63636364 0.33333333 0.59090909]
supps = No   [0.47663551 0.57943925 0.28037383 0.63551402]]

Number of samples encountered for each class during fitting
[ 66. 107.]

Number of samples encountered for each (class, feature) during fitting
supps = Yes [[34. 42. 22. 39.]   # tv magazine friends doctor
supps = No   [51. 62. 30. 68.]]
```

Nutrition Information: BernoulliNB

```
# Create the all-possible combinations of the features' values
xTest = pandas.DataFrame(list(itertools.product([0,1],
                                repeat = len(feature))),
                            columns = feature)

# Score the xTest and append the predicted probabilities to the xTest
yTest_predProb = pandas.DataFrame(_objNB.predict_proba(xTest),
                                  columns = ['P_suppsYes', 'P_suppsNo'])

yTest_score = pandas.concat([xTest, yTest_predProb], axis = 1)
```

Nutrition Information: Naïve Bayes

tv	magazine	friends	doctor	P_suppsYes	P_suppsNo
0	0	0	0	0.33938350	0.66061650
0	0	0	1	0.29853915	0.70146085
0	0	1	0	0.39733497	0.60266503
0	0	1	1	0.35324558	0.64675442
0	1	0	0	0.39486710	0.60513290
0	1	0	1	0.35089211	0.64910789
0	1	1	0	0.45575652	0.54424348
0	1	1	1	0.40959031	0.59040969
1	0	0	0	0.37475009	0.62524991
1	0	0	1	0.33178710	0.66821290
1	0	1	0	0.43476616	0.56523384
1	0	1	1	0.38920564	0.61079436
1	1	0	0	0.43223255	0.56776745
1	1	0	1	0.38675587	0.61324413
1	1	1	0	0.49417844	0.50582156
1	1	1	1	0.44731910	0.55268090

Nutrition Information: Naïve Bayes

Listen Only to your Doctor!

tv	magazine	friends	doctor	P_suppsYes	P_suppsNo
0	0	0	1	0.29853915	0.70146085
1	0	0	1	0.33178710	0.66821290
0	0	0	0	0.33938350	0.66061650
0	1	0	1	0.35089211	0.64910789
0	0	1	1	0.35324558	0.64675442
1	0	0	0	0.37475009	0.62524991
1	1	0	1	0.38675587	0.61324413
1	0	1	1	0.38920564	0.61079436
0	1	0	0	0.39486710	0.60513290
0	0	1	0	0.39733497	0.60266503
0	1	1	1	0.40959031	0.59040969
1	1	0	0	0.43223255	0.56776745
1	0	1	0	0.43476616	0.56523384
1	1	1	1	0.44731910	0.55268090
0	1	1	0	0.45575652	0.54424348
1	1	1	0	0.49417844	0.50582156

Decreasing P_suppsYNo

Gaussian Naïve Bayes

- The likelihood $Pr\left(x_j|y\right) = \frac{1}{\sqrt{2\pi\sigma_{y_c}^2}} \exp\left(-\frac{\left(x_i-\mu_{y_c}\right)^2}{2\sigma_{y_c}^2}\right)$
- The mean μ_{y_c} is estimated by the sample mean of x_i within the y_c category of the target variable.
- Likewise, the variable mean $\sigma_{y_c}^2$ is estimated by the sample variance of x_i within the y_c category of the target variable.

Multinomial Naïve Bayes

- Suppose the feature x_j has k categories in the training data.
- Let $n_{rc} \geq 0$ be the number of observations in the r^{th} category of the predictor and the y_c category of the target variable.
- Let $n_c = \sum_{r=1}^{k} n_{rc}$ be the number of observations in the y_c category of the target variable.
- The likelihood $\Pr\left(x_j | y\right) = \frac{n_c!}{\prod_{r=1}^{k} n_{rc}!} \prod_{r=1}^{k} (\theta_{rc})^{n_{rc}}$ where $0 < \theta_{rc} < 1$.

Multinomial Naïve Bayes: Smoothing Alpha

Naturally, we estimate θ_{rc} by the relative frequencies as $\theta_{rc} = n_{rc}/n_c$

If the r^{th} category of the feature is not observed in the y_c category of the target variable, then $n_{rc} = 0$ and the natural estimate $\theta_{rc} = 0$

Any $\theta_{rc} = 0$ will make the likelihood zero, and thus spoil all the calculations

Multinomial Naïve Bayes: Smoothing Alpha

- Therefore, we estimate $\theta_{rc} = \frac{n_{rc} + \alpha}{n_c + \alpha k}$. Note that $\sum_{r=1}^{k} \theta_{rc} = 1$
- The smoothing prior $\alpha \geq 0$ accounts for the categories of the feature x_j that are not observed in the y_c category of the target variable.
- Common choices of α are:
  - No smoothing: $\alpha = 0$ if all categories of the predictor x_j are always observed
  - Laplace smoothing: $\alpha = 1$ (Pierre-Simon Laplace, French, 1749 – 1827)
  - Lidstone smoothing: $\alpha < 1$ (George James Lidstone, British, 1870 – 1952)

Multinomial Naïve Bayes: Text Analysis

	ID	Words in Document	City in China?
Training	1	Chinese Beijing Chinese	Yes
	2	Chinese Chinese Shanghai	Yes
	3	Chinese Macao	Yes
	4	Tokyo Japan Chinese	No
Testing	5	Chinese Chinese Chinese Tokyo Japan	?
	6	Beijing Shanghai Macao	?

- Determine if the document contains the name of a Chinese city
- Reference: https://nlp.stanford.edu/IR-book/html/htmledition/naive-bayes-text-classification-1.html

Week 10 Chinese City Naive Bayes.py

Multinomial Naïve Bayes: Text Analysis

	ID	Words in Document	City in China?
Training	1	Chinese Beijing Chinese	Yes
	2	Chinese Chinese Shanghai	Yes
	3	Chinese Macao	Yes
	4	Tokyo Japan Chinese	No

- Six Features that indicate the number of times a word appeared.
- The features are: How Often <word> appear in the document?
- The word: (1) Chinese, (2) Beijing, (3) Shanghai, (4) Macao, (5) Tokyo, and (6) Japan

Multinomial Naïve Bayes: Text Analysis

(1) Chinese, (2) Beijing, (3) Shanghai, (4) Macao, (5) Tokyo, and (6) Japan

```
import numpy
import pandas
import sklearn.naive_bayes as naive_bayes

X = numpy.array([[2,1,0,0,0,0],
                 [2,0,1,0,0,0],
                 [1,0,0,1,0,0],
                 [1,0,0,0,1,1]])

y = numpy.array([1,1,1,0])

classifier = naive_bayes.MultinomialNB().fit(X, y)

print('Class Count:\n', classifier.class_count_)
print('Log Class Probability:\n', classifier.class_log_prior_ )
print('Feature Count (after adding alpha):\n', classifier.feature_count_)
print('Log Feature Probability:\n', classifier.feature_log_prob_)
```

Week 9 Chinese City Naive Bayes.py

Multinomial Naïve Bayes: Text Analysis

- Since three out of four documents have positive identification, the class probabilities are
  - Negative: Pr(y = 0) = 1/4 = 0.25.
  - Positive: Pr(y = 1) = 3/4 = 0.75
- The natural logarithm of these probabilities are
  - Negative: ln(Pr(y = 0)) = ln(0.25) = -1.38629436
  - Positive: ln(Pr(y = 1)) = ln(0.75) = -0.28768207

```
Class Count:
 [1. 3.]
Log Class Probability:
 [-1.38629436 -0.28768207]
```

Multinomial Naïve Bayes: Text Analysis

- Count the number of occurrences of each word by identification result.
-
```
X = numpy.array([[2,1,0,0,0,0],
                 [2,0,1,0,0,0],
                 [1,0,0,1,0,0],
                 [1,0,0,0,1,1]])
```

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	1	0	0	0	1	1
Positive (y = 1)	5	1	1	1	0	0

Multinomial Naïve Bayes: Text Analysis

- Specify alpha = 1. Thus, add one to each cell of the table
- Before

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	1	0	0	0	1	1
Positive (y = 1)	5	1	1	1	0	0

- After

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	2	1	1	1	2	2
Positive (y = 1)	6	2	2	2	1	1

Multinomial Naïve Bayes: Text Analysis

- Calculate the probability of each word, by Identification Result
- Table

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan	Total
Negative (y = 0)	2	1	1	1	2	2	9
Positive (y = 1)	6	2	2	2	1	1	14

- Result

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	2/9	1/9	1/9	1/9	2/9	2/9
Positive (y = 1)	6/14	2/14	2/14	2/14	1/14	1/14

Multinomial Naïve Bayes: Text Analysis

- Natural logarithm of the probabilities, by Identification Result
- The probability

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	2/9	1/9	1/9	1/9	2/9	2/9
Positive (y = 1)	6/14	2/14	2/14	2/14	1/14	1/14

- The natural logarithm of the probability

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	-1.5041	-2.1972	-2.1972	-2.1972	-1.5041	-1.5041
Positive (y = 1)	-0.8473	-1.9459	-1.9459	-1.9459	-2.6391	-2.6391

Multinomial Naïve Bayes: Text Analysis

- The natural logarithm of the probability

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	-1.5041	-2.1972	-2.1972	-2.1972	-1.5041	-1.5041
Positive (y = 1)	-0.8473	-1.9459	-1.9459	-1.9459	-2.6391	-2.6391

```
Log Feature Probability:
 [[-1.5040774  -2.19722458 -2.19722458 -2.19722458 -1.5040774  -1.5040774 ]
  [-0.84729786 -1.94591015 -1.94591015 -1.94591015 -2.63905733 -2.63905733]]
```

Multinomial Naïve Bayes: Text Analysis

- Given the number of occurrences of the words, what is the likelihood of a positive identification?

	ID	Words in Document	City in China?
Training	1	Chinese Beijing Chinese	Yes
	2	Chinese Chinese Shanghai	Yes
	3	Chinese Macao	Yes
	4	Tokyo Japan Chinese	No

Multinomial Naïve Bayes: Text Analysis

- Pr(Y = Negative | Chinese = 2, Beijing = 1, Shanghai = 0, Macao = 0, Tokyo = 0, Japan = 0)

  $\propto$ Pr(Y = Negative) × Pr(Chinese = 2, Beijing = 1, Shanghai = 0, Macao = 0, Tokyo = 0, Japan = 0 | Y = Negative)

  $$\propto \frac{1}{4} \times \left(\frac{2}{9}\right)^2 \times \left(\frac{1}{9}\right)^1 \times \left(\frac{1}{9}\right)^0 \times \left(\frac{1}{9}\right)^0 \times \left(\frac{2}{9}\right)^0 \times \left(\frac{2}{9}\right)^0 = \frac{1}{729}$$

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	2/9	1/9	1/9	1/9	2/9	2/9
Positive (y = 1)	6/14	2/14	2/14	2/14	1/14	1/14

Score First Document in Training

- Pr(Y = Positive | Chinese = 2, Beijing = 1, Shanghai = 0, Macao = 0, Tokyo = 0, Japan = 0)

  ∝ Pr(Y = Positive) × Pr(Chinese = 2, Beijing = 1, Shanghai = 0, Macao = 0, Tokyo = 0, Japan = 0 | Y = Positive)

$$\propto \frac{3}{4} \times \left(\frac{6}{14}\right)^2 \times \left(\frac{2}{14}\right)^1 \times \left(\frac{2}{14}\right)^0 \times \left(\frac{2}{14}\right)^0 \times \left(\frac{1}{14}\right)^0 \times \left(\frac{1}{14}\right)^0 = \frac{27}{1372}$$

Identification Result	Chinese	Beijing	Shanghai	Macao	Tokyo	Japan
Negative (y = 0)	2/9	1/9	1/9	1/9	2/9	2/9
Positive (y = 1)	6/14	2/14	2/14	2/14	1/14	1/14

Multinomial Naïve Bayes: Text Analysis

- For the first document in Training,
  - Pr(Y = Negative | Chinese = 2, Beijing = 1, Shanghai = 0, Macao = 0, Tokyo = 0, Japan = 0) $\propto$ 1/729
  - Pr(Y = Positive | Chinese = 2, Beijing = 1, Shanghai = 0, Macao = 0, Tokyo = 0, Japan = 0) $\propto$ 27/1372
- Final step is to rescale these two values such that the resulting values add up to one.
  - Pr(Y = Negative | Chinese = 2, Beijing = 1, Shanghai = 0, Macao = 0, Tokyo = 0, Japan = 0) = (1/729) / (1/729 + 27/1372) = 0.06516267
  - Pr(Y = Positive | Chinese = 2, Beijing = 1, Shanghai = 0, Macao = 0, Tokyo = 0, Japan = 0) = (27/1372) / (1/729 + 27/1372) = 0.93483733

Multinomial Naïve Bayes: Text Analysis

- Predicted Probabilities for all the documents

	ID	Words in Document	City in China?	Pr(Positive|Dcoument)	Pr(Negative|Document)
Training	1	Chinese Beijing Chinese	Yes	0.93483733	0.06516267
	2	Chinese Chinese Shanghai	Yes	0.93483733	0.06516267
	3	Chinese Macao	Yes	0.88149940	0.11850060
	4	Tokyo Japan Chinese	No	0.37412328	0.62587672
Testing	5	**Chinese Chinese Chinese** Tokyo Japan	Yes	0.68975861	0.31024139
	6	Beijing Shanghai Macao	No	0.33878380	0.66121620

```
Predicted Conditional Probability (Training):
[[0.06516267 0.93483733]
 [0.06516267 0.93483733]
 [0.1185006  0.8814994 ]
 [0.62587672 0.37412328]]
Predicted Conditional Probability (Testing):
 [[0.31024139 0.68975861]
  [0.6612162  0.3387838 ]]
```

Lecture Recap

- Introduced to Directed Acyclic Graph (DAG) and Bayesian Network
- Understood Naïve Bayes' Algorithms
- Target is always Categorical
- Features can be
  - Categorical and Binary in particular
  - Continuous
  - Multinomial