Introduction to Artificial Intelligence

Unit # 7

Acknowledgement

- The slides of this lecture have been taken from the lecture slides of CS307 – "Introduction to Artificial Intelligence" and CSE652 – "Knowledge Discovery and Data mining" by Dr. Sajjad Haider.

Popular Machine Learning Techniques

- Classification
  - Classification Trees ✔
  - **Naïve Bayes**
  - Neural Networks
- Clustering
  - K-Means
  - Associative Memory

- In this course, the focus is on the classification techniques

Conditional Probability Example

	B		~B	
	C	~C	C	~C
A	12	5	9	2
~A	4	8	20	4

- P(A | B, C) = 12/16
- P(A, B | ~C) = 5 / 19
- P(B | ~A, C) = 4 / 24

Conditional Independence

- Two events A and B are independent if knowing that A has happened does not say anything about B happening.

  P(A B) = P(A) P(B)

  P(A | B) = P(A)

- Two events A and B are conditionally independent given a third event C precisely if the occurrence or non-occurrence of A and B are independent events in their conditional probability distribution given C.

  P(A B | C) = P(A | C) P(B | C)

  P(A | B C) = P(A | C)

Bayes Theorem

- $P(A \mid B) = \dfrac{P(B \mid A)\, P(A)}{P(B)}$

  $= \dfrac{P(B \mid A)\, P(A)}{P(B|A)P(A)+P(B|\neg A)P(\neg A)}$

- P(A) is the prior probability and P(A | B) is the posterior probability.

- Suppose events $A_1, A_2, \ldots\ldots, A_k$ are mutually exclusive and exhaustive; i.e., exactly one of the events must occur. Then for any event B:

  $P(A_i \mid B) = \dfrac{P(B \mid A_i)\, P(A_i)}{\sum P(B \mid A_i)\, P(A_i)}$

Example I

- According to American Lung Association, 7% of the population has lung cancer. Of these people having lung disease, 90% are smokers; and of those not having lung disease, 25.3% are smokers.
- Determine the probability that a randomly selected smoker has lung cancer.

Example I Solution

- Let L = Lung Cancer, S = Smoker
- Given that
  - P(L) = 0.07
  - P(S | L) = 0.90   P(~S | L) = 0.10
  - P(S | ~L) = 0.253   P(~S | ~L) = 0.747
- Find probability, P(L | S)

$$P(L \mid S) = \frac{P(S \cap L)}{P(S)} = \frac{P(S \mid L)P(L)}{P(S|L)P(L) + P(S|{\sim}L)P({\sim}L)}$$

$$P(L|S) = \frac{0.9 \times 0.07}{0.9 \times 0.07 + 0.253 \times 0.93}$$

Example II

- Assume that about 1 in 1000 individuals in a given organization have committed a security violation.
- Assume that the **sensitivity** of a routine screening polygraph is about 85%. That is, the probability that the polygraph report will indicate a concern is about 85% if the individual has committed a security violation.
- Assume the **specificity** of the polygraph is about 80%. That is, if the individual has not committed a security violation, there is about an 80% chance that the polygraph report will not indicate a concern.
- What is the posterior probability that an individual whose polygraph report indicates a concern has committed a security violation?

Example II Solution

- Let
  - S = Security Violation Committed,
  - T = Test Positive
- Given that
  - P(S) = 0.001
  - P(T | S) = 0.85 P(~T | S) = 0.15
  - P(T | ~S) = 0.20 P(~T | ~S) = 0.80
- Find probability, P(S | T)

$$P(S \mid T) = \frac{P(T \mid S)P(S)}{P(T \mid S)P(S) + P(T \mid \sim S)P(\sim S)}$$

$$P(T|S) = \frac{0.85 \ x \ 0.001}{0.85 \ x \ 0.001 + 0.20 \ x \ 0.999}$$

How to Estimate Probabilities from Data?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Class: $P(C) = N_c/N$
  - e.g., P(No) = 7/10, P(Yes) = 3/10
- For discrete attributes:

  $P(A_i \mid C_k) = |A_{ik}| / N_c$

  - where $|A_{ik}|$ is number of instances having attribute A_i and belongs to class C_k
  - Examples:

    P(Status=Married|No) = 4/7
    P(Refund=Yes|Yes)=0

Naïve Bayes
Classification: Mammals vs. Non-mammals

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

- Train the model (learn the parameters) using the given data set.
- Apply the learned model on new cases.

Naïve Bayes
Classification: Mammals vs. Non-mammals

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

A: attributes

M: mammals

N: non-mammals

$$P(A|M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A|N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A|M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A|N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

P(A|M)P(M) > P(A|N)P(N)

=> Mammals

Example: Play Tennis

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

P(P) = 9/14	
P(N) = 5/14	

Outlook	Temperature	Humidity	Windy	Class
rain	hot	high	false	?

outlook	
P(sunny|p) = 2/9	P(sunny|n) = 3/5
P(overcast|p) = 4/9	P(overcast|n) = 0
P(rain|p) = 3/9	P(rain|n) = 2/5
temperature	
P(hot|p) = 2/9	P(hot|n) = 2/5
P(mild|p) = 4/9	P(mild|n) = 2/5
P(cool|p) = 3/9	P(cool|n) = 1/5
humidity	
P(high|p) = 3/9	P(high|n) = 4/5
P(normal|p) = 6/9	P(normal|n) = 2/5
windy	
P(true|p) = 3/9	P(true|n) = 3/5
P(false|p) = 6/9	P(false|n) = 2/5

Accuracy or Error Rates

- Partition: Training-and-testing
  - use two independent data sets, e.g., training set (2/3), test set(1/3)
  - used for data set with large number of examples

Metrics for Performance Evaluation

- Focus on the predictive capability of a model
  - Rather than how fast it takes to classify or build models, scalability, etc.
- Confusion Matrix:

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a	b
	Class=No	c	d

a: TP (true positive)
b: FN (false negative)
c: FP (false positive)
d: TN (true negative)

Metrics for Performance Evaluation…

	PREDICTED CLASS		
ACTUAL CLASS		Class=Yes	Class=No
	Class=Yes	a (TP)	b (FN)
	Class=No	c (FP)	d (TN)

- Most widely-used metric:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

Limitation of Accuracy

- Consider a 2-class problem
  - Number of Class 0 examples = 9990
  - Number of Class 1 examples = 10
- If model predicts everything to be class 0, accuracy is 9990/10000 = 99.9 %
  - Accuracy is misleading because model does not detect any class 1 example

Cost Matrix

	PREDICTED CLASS		
ACTUAL CLASS	C(i|j)	**Class=Yes**	**Class=No**
	Class=Yes	C(Yes|Yes)	C(No|Yes)
	Class=No	C(Yes|No)	C(No|No)

C(i|j): Cost of misclassifying class j example as class i

Cost Matrix (Cont'd)

	PREDICTED CLASS		
ACTUAL CLASS		True	False
	True	10	5
	False	1	14

	PREDICTED CLASS		
ACTUAL CLASS		True	False
	True	10	3
	False	3	14

	PREDICTED CLASS		
ACTUAL CLASS		True	False
	True	10	6
	False	0	14

All three confusion matrices have the same accuracy value, i.e., **24 / 30**

What if the cost of misclassification is not the same for both type of errors?

Cost Matrix (Cont'd)

	PREDICTED CLASS	
	True	False
ACTUAL CLASS True	10	5x5
ACTUAL CLASS False	1	14

	PREDICTED CLASS	
	True	False
ACTUAL CLASS True	10	3x5
ACTUAL CLASS False	3	14

	PREDICTED CLASS	
	True	False
ACTUAL CLASS True	10	6x5
ACTUAL CLASS False	0	14

Suppose the cost of misclassifying True as False is 5 while the cost of misclassifying False as True is 1.

Accuracy values are:
24/50, 24/42, 24/54

Cost Matrix (Cont'd)

	PREDICTED CLASS	
	True	False
ACTUAL CLASS True	10	5x4
ACTUAL CLASS False	1	14

	PREDICTED CLASS	
	True	False
ACTUAL CLASS True	10	3x4
ACTUAL CLASS False	3	14

	PREDICTED CLASS	
	True	False
ACTUAL CLASS True	10	6x4
ACTUAL CLASS False	0	14

Suppose the cost of misclassifying True as False is **4** while the cost of misclassifying False as True is 1.

Accuracy values are:
24/45, 24/39, 24/48

Cost-Sensitive Measures

$$\text{Precision (p)} = \frac{a}{a+c}$$

$$\text{Recall (r)} = \frac{a}{a+b}$$

$$\text{F-measure (F)} = \frac{2rp}{r+p} = \frac{2a}{2a+b+c}$$

- Precision is biased towards C(Yes|Yes) & C(Yes|No)
- Recall is biased towards C(Yes|Yes) & C(No|Yes)
- F-measure is biased towards all except C(No|No)

$$\text{Weighted Accuracy} = \frac{w_1 a + w_4 d}{w_1 a + w_2 b + w_3 c + w_4 d}$$

Recall and Precision

Actual	Prediction
T	T
T	F
F	T
F	F
F	T
T	T
T	T
T	F
F	T
T	T

Recall and Precision

Actual	Prediction
T	T
T	F
F	T
F	F
F	T
T	T
T	T
T	F
F	T
T	T

- Recall = 4 / 6

Recall and Precision

Actual	Prediction
T	T
T	F
F	T
F	F
F	T
T	T
T	T
T	F
F	T
T	T

- Recall = 4 / 6
- Precision = 4 / 7
- F-Measure = 8 / 13