CSCI 347: Introduction to Data Mining

Introduction to Classification

Input Data Matrix:

Class label is always categorical

New Data Instance:

Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes

Weather	Weekend?	Finished HW?	Go Hiking?
Sunny	Yes	No	?

Goal: Predict the class of new data

Input also commonly in the form:

Weather	Weekend?	Finished HW?	"Target/ Label/Class"
Snow	Yes	No	Yes
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes

Weather	Weekend?	Finished HW?	Target/Label/ Class
Sunny	Yes	No	?

Input also commonly in the form:

New Data Instance

Weather	Weekend?	Finished HW?
Snow	Yes	No
Overcast	No	No
Sunny	Yes	No
Overcast	Yes	Yes
Overcast	No	Yes
Snow	No	Yes
Overcast	Yes	No
Sunny	Yes	No
Sunny	No	Yes
Snow	No	Yes
Snow	Yes	No
Overcast	Yes	No
Overcast	No	Yes

У
Yes
No
Yes
Yes
Yes
No
No
No
Yes
Yes
No
No
Yes

Weather	Weekend?	Finished HW?	у
Sunny	Yes	No	y=?



Introduction to Naive Bayes Algorithm

Input Data Matrix:

Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes

Weather	Weekend?	Finished HW?	Go Hiking?
Sunny	Yes	No	?

New Data Instance:

•	Weather	Weekend?	Finished HW?	Go Hiking?
	Sunny	Yes	No	?

Input Data Matrix:

1			
Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes



Weather	Weekend?	Finished HW?	Go Hiking?		
Snow	Yes	No	Yes		
Sunny	Yes	No	Yes		
Overcast	Yes	Yes	Yes		
Overcast	No	Yes	Yes		
Sunny	No	Yes	Yes		
Snow	No	Yes	Yes		
Overcast	No	Yes	Yes		

Weather	Weekend?	Finished HW?	Go Hiking?		
Overcast	No	No	No		
Snow	No	Yes	No		
Overcast	Yes	No	No		
Sunny	Yes	No	No		
Snow	Yes	No	No		
Overcast	Yes	No	No		

New Data Instance:

•	Weather	Weekend?	Finished HW?	Go Hiking?
	Sunny	Yes	No	?

 $n_1 = 7$

Input Data Matrix:

1			
Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes



	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Overcast	No	Yes	Yes

Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	
Snow	No	Yes	No	$n_2 = 6$
Overcast	Yes	No	No	
Sunny	Yes	No	No	
Snow	Yes	No	No	
Overcast	Yes	No	No	

What is the probability of "Yes" and what is the probability of "No" in our data? (these are the *prior* probabilities)

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

We want to now estimate the probabilities of c_1 and c_2 after observing the new data instance and then choose the one with maximum probability

New Data Instance:

Weather	Weekend?	Finished HW?	Go Hiking?
Sunny	Yes	No	?

 $n_1 = 7$

Class 1 (c_1) : "Yes"

000001 (0)// 100				
Weather	Weekend?	Finished HW?	Go Hiking?	
Snow	Yes	No	Yes	
Sunny	Yes	No	Yes	
Overcast	Yes	Yes	Yes	
Overcast	No	Yes	Yes	
Sunny	No	Yes	Yes	
Snow	No	Yes	Yes	
Overcast	No	Yes	Yes	

Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	
Snow	No	Yes	No	$n_2 = 6$
Overcast	Yes	No	No	
Sunny	Yes	No	No	
Snow	Yes	No	No	
Overcast	Yes	No	No	

NAIVE BAYES FINDS THE CLASS WITH HIGHEST PROBABILITY

New Data Instance:

What is the probability of "Yes" and what is the probability of "No" in after observing the new data instance?

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_1 | x) = ?$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(c_2 \mid x) = ?$$

• • •	Weather	Weekend?	Finished HW?	Go Hiking?
\mathcal{X}	Sunny	Yes	No	?

 $n_1 = 7$

Class 1 (c_1) : "Yes"

Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Overcast	No	Yes	Yes

Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	
Snow	No	Yes	No	$n_2 = 6$
Overcast	Yes	No	No	
Sunny	Yes	No	No	
Snow	Yes	No	No	
Overcast	Yes	No	No	

NAIVE BAYES USES BAYES' RULE

What is the probability of "Yes" and what is the probability of "No" in after observing the new data instance? Use Bayes' Rule:

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_1 | x) = \frac{p(x | c_1)p(c_1)}{p(x)}$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

New Data Instance:

•	Weather	Weekend?	Finished HW?	Go Hiking?	
•	Sunny	Yes	No	?	

 $n_1 = 7$

Class 1 (c_1) : "Yes"

	Ctu35 1 (C). 103						
Weather	Weekend?	Finished HW?	Go Hiking?				
Snow	Yes	No	Yes				
Sunny	Yes	No	Yes				
Overcast	Yes	Yes	Yes				
Overcast	No	Yes	Yes				
Sunny	No	Yes	Yes				
Snow	No	Yes	Yes				
Overcast	No	Yes	Yes				

C (C) (T (C							
Weather	Weekend?	Finished HW?	Go Hiking?				
Overcast	No	No	No				
Snow	No	Yes	No	$n_2 = 6$			
Overcast	Yes	No	No				
Sunny	Yes	No	No				
Snow	Yes	No	No				
Overcast	Yes	No	No				

New Data Instance:

Since we are going to choose the c_i x that **maximizes** $p(c_i|x)$, we can ignore p(x) and only need to further compute $p(x|c_i)$ for each class

$$argmax_{c_i} p(c_i | x) = argmax_{c_i} \frac{p(x | c_i)p(c_i)}{p(x)}$$
$$= argmax_{c_i} p(x | c_i)p(c_i)$$

$$p(c_1|x) = \underbrace{\frac{p(x|c_1)p(c_1)}{p(x)}}_{p(x)} \qquad p(c_2|x) = \underbrace{\frac{p(x|c_2)p(c_2)}{p(x)}}_{p(x)}$$
$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54 \qquad p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

•	Weather	Weather Weekend?		Go Hiking?	
x	Sunny	Yes	No	?	

 $n_1 = 7$

Class 1 (c_1) : "Yes" Weather Weekend? Go Hiking? Snow Yes No Yes Yes Sunny Yes No Yes Overcast Yes Yes Overcast No Yes Yes Sunny No Yes Yes Snow No Yes Yes Overcast No Yes Yes

Class 2 (a), "No"

	<u>lass 2 (</u>	$(C_2): I$	10	
Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	
Snow	No	Yes	No	1
Overcast	Yes	No	No	_
Sunny	Yes	No	No	
Snow	Yes	No	No	_
Overcast	Yes	No	No	

New Data Instance:

Since we are going to choose the c_i that maximizes $p(c_i|x)$, we can ignore p(x) and only need to further compute $p(x|c_i)$ for each class

$$argmax_{c_i} p(c_i|x) = argmax_{c_i} \frac{p(x|c_i)p(c_i)}{p(x)}$$
$$= argmax_{c_i} p(x|c_i)p(c_i)$$

$$p(c_1|x) = \underbrace{p(x|c_1)p(c_1)}_{p(x)} \qquad p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$
$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54 \qquad p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$p(x c_1) = p(X_1 = Sunny, X_2 = Yes, X_3 = No c_1$
$= p(X_1 = Sunny c_1)p(X_2 = Yes c_1)p(X_3 = No c_1)$

•	Weather Weekend?		Finished HW?	Go Hiking?	
x	Sunny	Yes	No	?	

 $n_1 = 7$

Class 1 (c_1) : "Yes"

	<u> </u>						
Weather	Weekend?	Finished HW?	Go Hiking?				
Snow	Yes	No	Yes				
Sunny	Yes	No	Yes				
Overcast	Yes	Yes	Yes				
Overcast	No	Yes	Yes				
Sunny	No	Yes	Yes				
Snow	No	Yes	Yes				
Overcast	No	Yes	Yes				

NAIVE BAYES ASSUMES ATTRIBUTES ARE INDEPENDENT

New Data Instance:

We make the **naive** assumption that

$$p(X_1 = \text{sunny}, X_2 = \text{Yes}, X_3 = \text{No} | c_1)$$
 is equivalent to :

$$p(X_1 = \text{Sunny} | c_1)p(X_2 = \text{Yes} | c_1)p(X_3 = \text{No} | c_1)$$

$$argmax_{c_i} p(c_i | x) = argmax_{c_i} \frac{p(x | c_i)p(c_i)}{p(x)}$$
$$= argmax_{c_i} p(x | c_i)p(c_i)$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_2|x) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$p(x \mid c_1) =$	$p(X_1 = Sunny$	$Y, X_2 = Yes, X_3$	$= No \mid c_1$
$= p(X_1 = Su)$	$nny \mid c_1)p(X_2 =$	$Yes \mid c_1 p(X_3 =$	$No c_1)$

•	Weather Weekend?		Finished HW?	Go Hiking?	
x	Sunny	Yes	No	?	

Class 1 (c_1) : "Yes"

	1 1 CC111	$C_1/.$	63
Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Overcast	No	Yes	Yes

..because these probabilities are easer to estimate

New Data Instance:

anama an	12 (a	(مد ا		anana an	10 (20	0)10/	(a)
argmax _c	$p(c_i)$	X) =	argmax	p(x)	$(c_i)p($	(c_i)
· ·	1			C	l		

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54 \qquad p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_2|x) = p(x)$$

$$p(x)$$

•	Weather	Weekend?	HW?	Go Hiking?
X	Sunnv	Yes	No	?

 $n_1 = 7$

Class 1 (c_1) : "Yes"

Ctu33 1 (C1): 1C3					
Weather	Weekend?	Finished HW?	Go Hiking?		
Snow	Yes	No	Yes		
Sunny	Yes	No	Yes		
Overcast	Yes	Yes	Yes		
Overcast	No	Yes	Yes		
Sunny	No	Yes	Yes		
Snow	No	Yes	Yes		
Overcast	No	Yes	Yes		

$$p(x | c_1) = p(X_1 = Sunny | c_1)p(X_2 = Yes | c_1)p(X_3 = No | c_1)$$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$

New Data Instance:

anana an	10(0	(مد ا	\	G14G144 G16	10 (20	0)10	(
argmax _c	$p(c_i)$	(X)) =	argmax	p(x)	$(c_i)p($	(c_i)
	i			C	i		

$$p(c_1|x) = \underbrace{\frac{p(x|c_1)p(c_1)}{p(x)}}_{p(x)}$$

$$p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54 \qquad p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$
Solve the second second

$$p(c_2) = \frac{n_2}{n} = \frac{6}{12} = 0.46$$

• •	Weather	Weekend?	Finished HW?	Go Hiking?
\mathcal{X}	Sunny	Yes	No	?

 $n_1 = 7$

Class 1 (c_1) "Yes"

$Ciu33$ I (C_1) . $IC3$						
Weather	Weekend?	Finished HW?	Go Hiking?			
Snow	Yes	No	Yes			
Sunny	Yes	No	Yes			
Overcast	Yes	Yes	Yes			
Overcast	No	Yes	Yes			
Sunny	No	Yes	Yes			
Snow	No	Yes	Yes			
Overcast	No	Yes	Yes			

$$p(x | c_1) = p(X_1 = Sunny | c_1)p(X_2 = Yes | c_1)p(X_3 = No | c_1)$$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$

$$p(X_2 = Yes | c_1) = \frac{3}{7} = 0.43$$

New Data Instance:

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54 \qquad p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(c_2 | x) = \frac{p(x | c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_2|x) = p(x)$$

$$p(x)$$

• •	Weather	Weekend?	Finished HW?	Go Hiking?
$ \mathcal{X} $	Sunny	Yes	No	?

 $n_1 = 7$

Class 1 (c.). "Voc"

$Ciuss I (C_1)$. Its					
Weather	Weekend?	Finished HW?	Go Hiking?		
Snow	Yes	No	Yes		
Sunny	Yes	No	Yes		
Overcast	Yes	Yes	Yes		
Overcast	No	Yes	Yes		
Sunny	No	Yes	Yes		
Snow	No	Yes	Yes		
Overcast	No	Yes	Yes		

$$p(x | c_1) = p(X_1 = Sunny | c_1)p(X_2 = Yes | c_1)p(X_3 = No | c_1)$$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$

$$p(X_3 = No \mid c_1) = \frac{2}{7} = 0.29$$

New Data Instance:

$argmax_{c_i}$	$p(c_i)$	$ x\rangle$	=	argmax	p(x	$c_i)p$	(c_i)	
9				C	i			

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$
 $p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$

$$p(X_3 = No \mid c_1) = \frac{2}{7} = 0.29$$

$p(c_2 x) =$	$p(x \mid c_2)p(c_2)$
$p(c_2 x)$ –	p(x)

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

•	Weather	Weekend?	Finished HW?	Go Hiking?
\mathcal{X}	Sunny	Yes	No	?

 $n_1 = 7$

Class 1 (c_{\bullet}) . "Yes"

$\underline{\hspace{1cm}}$					
Weather	Weekend?	Finished HW?	Go Hiking?		
Snow	Yes	No	Yes		
Sunny	Yes	No	Yes		
Overcast	Yes	Yes	Yes		
Overcast	No	Yes	Yes		
Sunny	No	Yes	Yes		
Snow	No	Yes	Yes		
Overcast	No	Yes	Yes		

$$p(x \mid c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035$$

New Data Instance:

$argmax_{c_i} p(c_i x) =$	$argmax_{c_i} p(x $	$c_i)p(c_i)$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$
 $p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$

$$p(X_3 = No \mid c_1) = \frac{2}{7} = 0.29$$

$$p(x \mid c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035$$

• •	Weather	Weekend?	Finished HW?	Go Hiking?
\mathcal{X}	Sunny	Yes	No	?

 $n_1 = 7$

Class 1 (c_1) : "Yes"

Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Overcast	No	Yes	Yes

$$p(x | c_1)p(c_1) = 0.035(0.54) = 0.0189$$

New Data Instance:

argmax _c	$p(c_i)$	$ x\rangle$) =	argmax	p(x)	$(c_i)p($	(c_i)
c	$P \setminus l$	1 22)		w.8,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		$(v_l)P$	$(^{\circ}l)$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \underbrace{\frac{p(x|c_2)p(c_2)}{p(x)}}_{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54 \qquad p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$

$$p(x | c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035$$

$$p(x | c_1)p(c_1) = 0.0189$$

ľ	
$n(c \mid r)$ -	$(p(x c_2))p(c_2)$
$p(c_2 x) =$	$-\frac{p(x)}{p(x)}$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

• Weather	Weekend?	Finished HW?	Go Hiking?
x (Sunny)	Yes	No	?
x Sunny	Yes	No	?

Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	
Snow	No	Yes	No	1
Overcast	Yes	No	No	
Sunny	Yes	No	No	
Snow	Yes	No	No	
Overcast	Yes	No	No	

$$p(X_1 = Sunny) c_2) = \frac{1}{6} = 0.17$$

New Data Instance:

• • • • • • • • • • • • • • • • • • • •	Weather	Weekend?	HW?	Go Hiking?
$argmax_{c} p(c_{i} x) = argmax_{c} p(x c_{i})p(c_{i})$	Sunny	Yes	No	?

$$argmax_{c_i} p(c_i | x) = argmax_{c_i} p(x | c_i) p(c_i)$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \underbrace{\frac{p(x|c_2)p(c_2)}{p(x)}}_{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$
 $p(X_1 = Sunny | c_2) = \frac{1}{6} = 0.17$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$

$$p(x \mid c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035$$

$$p(x | c_1)p(c_1) = 0.0189$$

$n(c \mid r)$ -	p(x)	$ c_2 $	$p(c_2)$)
$p(c_2 x) =$		p(x	;)	_

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$
 $p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$

$$p(X_1 = Sunny \mid c_2) = \frac{1}{6} = 0.17$$

	30000 - (
Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	
Snow	No	Yes	No	1
Overcast	Yes	No	No	
Sunny	Yes	No	No	
Snow	Yes	No	No	
Overcast	Yes	No	No	

$$p(X_2 = Yes \mid c_2) = \frac{4}{6} = 0.67$$

New Data Instance:

$argmax_{c_i} p(c_i x) = argmax_{c_i} p(x c_i) p(c_i)$	
--	--

 $p(X_1 = Sunny | c_2) = \frac{1}{6} = 0.17$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \underbrace{p(x|c_2)p(c_2)}_{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$
 $p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$

$$p(X_1 = Sunny \mid c_1) = \frac{2}{7} = 0.29$$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$
 $p(X_2 = Yes \mid c_2) = \frac{4}{6} = 0.67$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$

$$p(x \mid c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035$$

$$p(x | c_1)p(c_1) = 0.0189$$

$p(x \mid c_i)p(c_i)$ X Sunny Yes No	?	

	$\lambda 1005 Z$ ((C_{γ}) : IN	0	
Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	
Snow	No	Yes	No	n
Overcast	Yes	No	No	
Sunny	Yes	No	No	
Snow	Yes	No	No	
Overcast	Yes	No	No	

$$p(X_3 = No \mid c_2) = \frac{5}{6} = 0.83$$

New Data Instance:

• • • • • • • • • • • • • • • • • • • •	• • •	Weather	Weekend?	Finished HW?	Go Hiking?
$argmax_{o} p(c_{i} x) = argmax_{o} p(x c_{i})p(c_{i})$	$\chi[$	Sunnv	Yes	No	?

$argmax_{c_i}p(c_i $	$x) = argmax_0$	$p(x \mid c_i)p(c_i)$
\mathcal{S}_l		

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$
 $p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \underbrace{\frac{p(x|c_2)p(c_2)}{p(x)}}_{p(x)}$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(X_1 = Sunny \mid c_2) = \frac{1}{6} = 0.17$$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$
 $p(X_2 = Yes \mid c_2) = \frac{4}{6} = 0.67$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$
 $p(X_3 = No \mid c_2) = \frac{5}{6} = 0.83$

	(_ / /		
Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	
Snow	No	Yes	No	1
Overcast	Yes	No	No	
Sunny	Yes	No	No	
Snow	Yes	No	No	
Overcast	Yes	No	No	

$$p(x | c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035$$

$$p(x | c_1)p(c_1) = 0.0189$$

$$p(x | c_2) = \left(\frac{1}{6}\right) \left(\frac{4}{6}\right) \left(\frac{5}{6}\right) = 0.093$$

New Data Instance:

• • • • • • • • • • • • • • • • • • • •		Weather	Weekend?	Finished HW?	Go Hiking?
$argmax_{c} p(c_{i} x) = argmax_{c} p(x c_{i})p(c_{i})$	$\mathcal{X}[$	Sunnv	Yes	No	?

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$
 $p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$

$$p(X_1 = Sunny | c_1) = \frac{2}{7} = 0.29$$
 $p(X_1 = Sunny | c_2) = \frac{1}{6} = 0.17$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \underbrace{\frac{p(x|c_2)p(c_2)}{p(x)}}_{p(x)}$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(X_1 = Sunny \mid c_2) = \frac{1}{6} = 0.17$$

$$p(X_2 = \text{Yes} \mid c_1) = \frac{3}{7} = 0.43$$
 $p(X_2 = \text{Yes} \mid c_2) = \frac{4}{6} = 0.67$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$
 $p(X_3 = No \mid c_2) = \frac{5}{6} = 0.83$

$$p(x \mid c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035 \quad p(x \mid c_2) = \left(\frac{1}{6}\right) \left(\frac{4}{6}\right) \left(\frac{5}{6}\right) = 0.093$$

Class 2 (c_a) · "No"

	ω	U)). II	10
Weather	Weekend?	Finished HW?	Go Hiking?
Overcast	No	No	No
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Snow	Yes	No	No
Overcast	Yes	No	No

$$p(x | c_1)p(c_1) = 0.0189$$

$$p(x | c_2)p(c_2) = (0.093)(0.46) = 0.0428$$

New Data Instance:

Finished

Go Hiking?

			HW?
$argmax_c p(c_i x) = argmax_c p(x c_i)p(c_i)$	X Sunny	Yes	No
$c_i g_i c_i f(e_i) f(e_i) f(e_i)$			

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$
 $p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(X_1 = Sunny \mid c_1) = \frac{2}{7} = 0.29$$

$$p(X_1 = Sunny \mid c_2) = \frac{1}{6} = 0.17$$

$$p(X_2 = Yes \mid c_1) = \frac{3}{7} = 0.43$$

$$p(X_2 = \text{Yes} \mid c_1) = \frac{3}{7} = 0.43$$
 $p(X_2 = \text{Yes} \mid c_2) = \frac{4}{6} = 0.67$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$

$$p(X_2 = No \mid c_1) = \frac{2}{7} = 0.29$$
 $p(X_3 = No \mid c_2) = \frac{5}{6} = 0.83$

$$p(x | c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035$$

$$p(x \mid c_1) = \left(\frac{2}{7}\right) \left(\frac{3}{7}\right) \left(\frac{2}{7}\right) = 0.035 \quad p(x \mid c_2) = \left(\frac{1}{6}\right) \left(\frac{4}{6}\right) \left(\frac{5}{6}\right) = 0.093$$

$$p(x \mid c_1)p(c_1) = 0.0189$$

$$p(x | c_1)p(c_1) = 0.0189$$
 $p(x | c_2)p(c_2) = 0.0428$

Class	1 (c_1	: "	Yes	"
-------	-----	-------	-----	-----	---

Weather	Weekend?	Finished HW?	Go Hiking?	
Snow	Yes	No	Yes	_
Sunny	Yes	No	Yes	
Overcast	Yes	Yes	Yes	$n_1 = 1$
Overcast	No	Yes	Yes	
Sunny	No	Yes	Yes	_
Snow	No	Yes	Yes	
Overcast	No	Yes	Yes	

Weather	Weekend?	Finished HW?	Go Hiking?	
Overcast	No	No	No	$n_2 =$
Snow	No	Yes	No	
Overcast	Yes	No	No	_
Sunny	Yes	No	No	
Snow	Yes	No	No	_
Overcast	Yes	No	No	

NAIVE BAYES FOR NUMERICAL ATTRIBUTES

 $argmax_{c_i} p(c_i | x) = argmax_{c_i} p(x | c_i) p(c_i)$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(x \mid c_1) = ?$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(x \mid c_2) = ?$$

	Hours of sleep previous night	Percentage HW Finished	Go Hiking?
0.2	7.5	85	?

Inches of Rain in Past 2 hours	Hours of sleep previous night	Percentage HW Finished	Go Hiking?
0	9	80	Yes
0.5	5	90	No
1	7	95	Yes
5	7	100	Yes
0.3	8	100	Yes
0.4	4	100	No
0.1	9	27	No
0	9	50	No
0	8	100	Yes
3	10	98	Yes
6	8	95	No
2.1	8	70	No
1.02	8.5	98	Yes

NAIVE BAYES FOR NUMERICAL ATTRIBUTES

$argmax_{c_i} p(c_i | x) = argmax_{c_i} p(x | c_i) p(c_i)$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$
$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54 \qquad p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(x \mid c_1) = ?$$

$$p(x \mid c_2) = ?$$

Assume a (multivariate) normal distribution

$$p(x | c_i) = \prod_{j=1}^d p(x_j | c_i) = \prod_{j=1}^d f(x_j | \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2)$$

Where:

$$f(x_j | \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left(\frac{-(x_j - \hat{\mu}_{ij})^2}{2\hat{\sigma}_{ij}^2}\right)$$

	Hours of sleep previous night	Percentage HW Finished	Go Hiking?
0.2	7.5	85	?

	•		•
Inches of Rain in Past 2 hours	Hours of sleep previous night	Percentage HW Finished	Go Hiking?
0	9	80	Yes
0.5	5	90	No
1	7	95	Yes
5	7	100	Yes
0.3	8	100	Yes
0.4	4	100	No
0.1	9	27	No
0	9	50	No
0	8	100	Yes
3	10	98	Yes
6	8	95	No
2.1	8	70	No
1.02	8.5	98	Yes

NAIVE BAYES FOR NUMERICAL ATTRIBUTES

$$argmax_{c_i} p(c_i | x) = argmax_{c_i} p(x | c_i) p(c_i)$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)}$$

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x)} \qquad p(c_2|x) = \frac{p(x|c_2)p(c_2)}{p(x)}$$
$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54 \qquad p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(c_1) = \frac{n_1}{n} = \frac{7}{13} = 0.54$$

$$p(c_2) = \frac{n_2}{n} = \frac{6}{13} = 0.46$$

$$p(x \mid c_1) = ?$$

$$p(x \mid c_2) = ?$$

$$p(x | c_i) = \prod_{j=1}^{d} p(x_j | c_i) = \prod_{j=1}^{d} f(x_i | \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2)$$

$$f(x_j | \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2) = \frac{1}{\sqrt{2\pi\sigma_{ii}}} \exp\left(\frac{-(x_j - \hat{\mu}_{ij})^2}{2\hat{\sigma}_{ij}^2}\right)$$

	Hours of sleep previous night	Percentage HW Finished	Go Hiking?
0.2	7.5	85	?

Inches of Rain in Past 2 hours	Hours of sleep previous night	Percentage HW Finished	Go Hiking?
0	9	80	Yes
0.5	5	90	No
1	7	95	Yes
5	7	100	Yes
0.3	8	100	Yes
0.4	4	100	No
0.1	9	27	No
0	9	50	No
0	8	100	Yes
3	10	98	Yes
6	8	95	No
2.1	8	70	No
1.02	8.5	98	Yes

NAIVE BAYES ALGORITHM (NUMERICAL DATA)

NaiveBayes($D = \{x_j, y_j\}_{j=1}^n$):

1. for
$$i = 1, ..., k$$
:

2.
$$D_i = \{x_j | y_j = c_i, j = 1,..., n\}$$

3.
$$n_i = |D_i|$$

$$4. \ p(c_i) = \frac{n_i}{n}$$

$$5. \ \hat{\mu}_i \leftarrow \frac{1}{n_i} \sum_{x_j \in D_i} x_j$$

6.
$$Z_i \leftarrow D_i - 1.\hat{\mu}_i^T$$

7. For j = 1, ..., d:

7.
$$\hat{\sigma}_{ij}^2 \leftarrow \frac{1}{n_i} Z_{ij}^T Z_{ij}$$

8.
$$\hat{\sigma}_i \leftarrow (\hat{\sigma}_{i1}, ..., \hat{\sigma}_{id})^T$$

9. Return $p(c_i)$, $\hat{\mu}_i$, $\hat{\sigma}_i$ for all $i \in \{1,...,k\}$

PredictClass(x and $p(c_i)$, $\hat{\mu}_i$, $\hat{\sigma}_i$ for all i = 1, ..., k):

1.
$$c \leftarrow \operatorname{argmax}_{c_i} \{ p(c_i) \prod_{j=1}^d f(x_j | \hat{\mu}_{ij}, \sigma_{ij}^2) \}$$

2. Return c

CSCI 347: Introduction to Data Mining

Evaluating Classification Algorithms

Input Data Matrix:

Class label is always categorical

New Data Instance:

Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes

Weather	Weekend?	Finished HW?	Go Hiking?
Sunny	Yes	No	?

Goal: Predict the class of new data

EVALUATION OF CLASSIFICATION

Input Data Matrix:

Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes

Test Data Instance:

Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes

Evaluate the class predictions of test data

TRAINING SET AND TEST SET

Test Data Instance:

Weather Weekend? Finished HW? Go Hiking?

Snow Yes No Yes

Input Data Matrix:

Weather	Weekend?	Finished HW?	Go Hiking?
Snow	Yes	No	Yes
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes

Evaluate the class predictions of test data based an algorithm that built a model using only training data

Weather	Weekend?	Finished HW?	Go Hiking?
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	Yes	Yes	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes

TRAINING SET AND TEST SET:

OFTEN 80/20 SPLIT

Input Data Matrix:

	Weather	Weekend?	Finished HW?	Go Hiking?
i_1	Snow	Yes	No	Yes
	Overcast	No	No	No
	Sunny	Yes	No	Yes
4	Overcast	Yes	Yes	Yes
	Overcast	No	Yes	Yes
	Snow	No	Yes	No
	Overcast	Yes	No	No
	Sunny	Yes	No	No
	Sunny	No	Yes	Yes
1) Snow	No	Yes	Yes
	Snow	Yes	No	No
	Overcast	Yes	No	No
	Overcast	No	Yes	Yes

Test Data:

	Weather	Weekend?	Finished HW?	Go Hiking?
x_1	Snow	Yes	No	Yes
\mathcal{X}_4	Overcast	Yes	Yes	Yes
x_{10}	Snow	No	Yes	Yes

Evaluate the class predictions of test data based on algorithm that built a model using only training data

Training Data:

Weather	Weekend?	Finished HW?	Go Hiking?
Overcast	No	No	No
Sunny	Yes	No	Yes
Overcast	No	Yes	Yes
Snow	No	Yes	No
Overcast	Yes	No	No
Sunny	Yes	No	No
Sunny	No	Yes	Yes
Snow	Yes	No	No
Overcast	Yes	No	No
Overcast	No	Yes	Yes

PIPELINE FOR EVALUATION

Test Data:

	•	•			
100	1111	111	\mathbf{O}	1)	ata:
110					ala.

Weather	Weekend?	Finished HW?	Go Hiking?		
Overcast	No	No	No		
Sunny	Yes	No	Yes		
Overcast	No	Yes	Yes		
Snow	No	Yes	No		
Overcast	Yes	No	No		
Sunny	Yes	No	No		
Sunny	No	Yes	Yes		
Snow	Yes	No	No		
Overcast	Yes	No	No		
Overcast	No	Yes	Yes		



Finished Weather Weekend? Go Hiking? HW? Snow Yes No Yes **Overcast** Yes Yes Yes x_{10} Snow No Yes Yes

Feed the model unlabeled test data to make predictions on

("predict" the test data classes)

	Weather	Weekend?	Finished HW?	Go Hiking?
x_1	Snow	Yes	No	Yes
x_4	Overcast	Yes	Yes	Yes
x_1	₀ Snow	No	Yes	No

Give a classification algorithm training data to use to output a model

("fit" the data)

Compare predicted classes to ground truth classes

EVALUATION METRICS

How to compare predicted classes to ground truth classes?

Predicted Classes:

	Weather	Weekend?	Finished HW?	Go Hiking?
x_1	Snow	Yes	No	Yes
x_4	Overcast	Yes	Yes	Yes
x_{10}	Snow	No	Yes	No

Test Data:

	Weather	Weekend?	Finished HW?	Go Hiking?
x_1	Snow	Yes	No	Yes
x_4	Overcast	Yes	Yes	Yes
x_{10}	Snow	No	Yes	Yes

EVALUATION METRICS

How to compare predicted classes to ground truth classes?

Predicted Classes:

	Weather	Weekend?	Finished HW?	Go Hiking?
\tilde{x}_1	Snow	Yes	No	$\hat{y}_1 = y_{es}$
\tilde{x}_2	Overcast	Yes	Yes	$\hat{y}_2 = y_{es}$
\tilde{x}_3	Snow	No	Yes	$\hat{y}_3 = No$

Test Data:

	Weather	Weekend?	Finished HW?	Go Hikir	ıg?
\tilde{x}_1	Snow	Yes	No	$y_1 =$	Yes
\tilde{x}_2	Overcast	Yes	Yes	$y_2 =$	Yes
\tilde{x}_3	Snow	No	Yes	$y_3 =$	Yes

Accuracy:

$$\frac{1}{n_T} \sum_{i=1}^{n_T} I(y_i = \hat{y}_i)$$

where:

$$I(y_i = \hat{y}_i)$$

is 1 if y_i and \hat{y}_i
have the same
value, and is 0
otherwise

EVALUATION METRICS

How to compare predicted classes to ground truth classes?

Predicted Classes:

	Weather	Weekend?	Finished HW?	Go Hiking?
\tilde{x}_1	Snow	Yes	No	$\hat{y}_1 = y_{es}$
\tilde{x}_2	Overcast	Yes	Yes	$\hat{y}_2 = y_{es}$
$\tilde{\chi}_3$	Snow	No	Yes	$\hat{y}_3 = No$

Test Data:

	Weather	Weekend?	Finished HW?	Go Hikin	g?
\tilde{x}_1	Snow	Yes	No	$y_1 =$	Yes
\tilde{x}_2	Overcast	Yes	Yes	$y_2 =$	Yes
\tilde{x}_3	Snow	No	Yes	$y_3 =$	Yes

Accuracy:

$$\frac{1}{n_T} \sum_{i=1}^{n_T} I(y_i = \hat{y}_i) = \frac{1}{3} (1 + 1 + 0) = \frac{2}{3}$$

OTHER EVALUATION METRICS

How to compare predicted classes to ground truth classes?

Predicted Classes:

	Weather	Weekend?	Finished HW?	Go Hiking?
x_1	Snow	Yes	No	Yes
\mathcal{X}_4	Overcast	Yes	Yes	Yes
x_{10}	Snow	No	Yes	No

Test Data:

	Weather	Weekend?	Finished HW?	Go Hiking?
x_1	Snow	Yes	No	Yes
x_4	Overcast	Yes	Yes	Yes
x_{10}	Snow	No	Yes	Yes

Contingency-based measures:

➤ Precision, recall, F-measure

For binary classification:

- ➤ TP, TN,FP,FN
- > Sensitivity, specificity

Area under ROC Curve