Naïve Bayes - ML

Modified from Weka book notes

Bayes's rule

- Probability of class *C* given evidence *E:*

$$P(C \mid E) = \frac{P(E \mid C)P(C)}{P(E)}$$

- *A priori* probability of *C* :
  - Probability of event *before* evidence is seen
- *A posteriori* probability of *C* :
  - Probability of event *after* evidence is seen

Naïve Bayes for classification

- Classification learning: what's the probability of the class given an instance?
  - Evidence E = instance
  - Event C = class value for instance
- Naïve assumption: evidence splits into parts (i.e. attributes) that are *independent so for s attributes*

$$P[C \mid E] = \frac{P[E_1 \mid C] P[E_2 \mid C] \ldots P[E_s \mid C] P(C)}{P[E]}$$

Probabilities for Weka weather data

Outlook			Temperature			Humidity			Windy			Play	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

Outlook	**Temp**	**Humidity**	**Windy**	**Play**
Sunny	**Hot**	**High**	**False**	**No**
Sunny	**Hot**	**High**	**True**	**No**
Overcast	**Hot**	**High**	**False**	**Yes**
Rainy	**Mild**	**High**	**False**	**Yes**
Rainy	**Cool**	**Normal**	**False**	**Yes**
Rainy	**Cool**	**Normal**	**True**	**No**
Overcast	**Cool**	**Normal**	**True**	**Yes**
Sunny	**Mild**	**High**	**False**	**No**
Sunny	**Cool**	**Normal**	**False**	**Yes**
Rainy	**Mild**	**Normal**	**False**	**Yes**
Sunny	**Mild**	**Normal**	**True**	**Yes**
Overcast	**Mild**	**High**	**True**	**Yes**
Overcast	**Hot**	**Normal**	**False**	**Yes**
Rainy	**Mild**	**High**	**True**	**No**

Probabilities for weather data

Outlook			Temperature			Humidity			Windy			Play	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/ 14	5/ 14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

- A new day:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Likelihood of the two classes

For “yes” =

For “no” =

Probabilities for weather data

Outlook			Temperature			Humidity			Windy			Play	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

- A new day:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Discriminant for the the two classes

For "yes" = log 2/9 + log 3/9 + log 3/9 +log 3/9 + log 9/14= -5.24174

For "no" = log 3/5 + log 1/5 + log 4/5 + log 3/5 + log 5/14= - 3.88385

Choose Which? Highest value (No).

Probabilities for weather data

Outlook			Temperature			Humidity			Windy			Play	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

- A new day:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Likelihood of the two classes

For "yes" = 2/9 × 3/9 × 3/9 × 3/9 × 9/14 = 0.0053

For "no" = 3/5 × 1/5 × 4/5 × 3/5 × 5/14 = 0.0206

Conversion into a probability by normalization:

P("yes") = 0.0053 / (0.0053 + 0.0206) = 0.205

P("no") = 0.0206 / (0.0053 + 0.0206) = 0.795

Weather data example

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← ***Evidence E***

$$P[yes \mid E] = P[Outlook = Sunny \mid yes] \times P[Temperature = Cool \mid yes]$$
$$\times P[Humidity = High \mid yes] \times P[Windy = True \mid yes]$$
$$\times \frac{P[yes]}{P[E]}$$

Probability of class "yes"

$$= \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{P[E]}$$

The “zero-frequency problem”

- What if an attribute value doesn’t occur with every class value?
  (e.g. “Humidity = high” for class “yes”)
  - Probability will be zero!
  - *A posteriori* probability will also be zero!
    (No matter how likely the other values are!)
- Remedy: add 1 to the count for every attribute value-class combination (*Laplace estimator)*
- Result: probabilities will never be zero!
  (also: stabilizes probability estimates)

Modified probability estimates

- In some cases adding a constant different from 1 might be more appropriate
- Example: attribute *outlook* for class *yes*

$(2+\mu p_1)/(9+\mu)$ $\quad$ $(4+\mu p_2)/(9+\mu)$ $\quad$ $(3+\mu p_3)/(9+\mu)$

Sunny ***Overcast*** ***Rainy***

- Weights don't need to be equal (but they must sum to 1) and above they will, if

p_i=1/3 and $\mu = 1$ *or* $\mu=3$.

Modified probability estimates

- Example: attribute *outlook* for class *yes*

$(2+\mu p_1)/(9+\mu)$ $(4+\mu p_2)/(9+\mu)$ $(3+\mu p_3)/(9+\mu)$

Sunny ***Overcast*** ***Rainy***

- If p_i=1/3 and μ=3. *We have*
- *3/12, 5/12 and 4/12 which sum to 1.*
- If p_i=1/3 and μ=1. *We have*
- *2.33/10, 4.33/10 and 3.33/10 which sum to 1*
- *What is effect of* p_1=1/3, p_2=1/2, p_3=1/6, μ=10
- *Differential weights with original value of rainy reduced. Still sum to 1.*
-