Statistical Learning

- **Conditional Probability**
  - Conditional probability is a measure of the probability of an event occurring, given that another event has already occurred.
  - If the occurrence of event B depends on event A, then the event B can be analyzed by a conditional probability with respect to A
  - the conditional probability of B given A:

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

  - From this equation, we can infer:

$$P(A \cap B) = P(A)P(B|A)$$

Statistical Learning

- ✓ Similarly, the conditional probability of A given B:

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

- ✓ From this equation, we can infer:

$P(A \cap B) = P(B)P(A|B)$

- ✓ We can replace $P(A \cap B)$ with $P(B)P(A|B)$ in our first formula:

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

Posterior= (likelihood x prior)/evidence

Bayesian Theory: The conditional probability of B given A

Bayesian Classifier (Naive Bayes)

- ✓ Naive Bayes is a probabilistic algorithm that's typically used for classification problems
- ✓ Simple, intuitive, and yet performs surprisingly well in many cases
  - ✓ Spam filter E-mail app are built on Naïve Bayes
- ✓ Naïve Bayes calculates given a data point X=(x1,x2,…,xn), what the odd of Y being y (C being Ci).

$$P(Ci|X) = \frac{P(X|Ci)P(Ci)}{\mathrm{P(X)}}$$

Bayesian Classifier (Naive Bayes)

- Naïve Bayes assumes that in the attributes set, Xi and Xj are conditionally independent given Y, for all i≠j.
- The fundamental Naive Bayes assumption is that each feature is independent and equal.

- Advantages:
  - Simple
  - Intuitive
  - High Performance

- Disadvantages
  - In real life, features are dependent and have different importance (weight). This limits the applicability of this algorithm in real-world use cases.
  - Its estimations can be wrong

Bayesian Classifier (Naive Bayes)

- **Example**: Probability of raining
  - When we see dark clouds, we assume that it will rain. Our evidence are dark clouds.
  - Modeling this with Bayesian theorem:

$$P(\text{raining} \mid \text{dark cloud}) = \frac{P(\text{dark cloud} \mid \text{raining}) \times P(\text{raining})}{P(\text{dark cloud})}$$

Bayesian Classifier (Naive Bayes)

- Raining depends on other features as well.
  - Wind speed
  - Humidity

$$P(\text{raining} \mid \text{dark cloud, wind speed, humidity}) = \frac{P(\text{dark cloud} \mid \text{raining}) \text{ x } P(\text{wind speed} \mid \text{raining}) \text{ x } P(\text{humidity} \mid \text{raining}) \text{ x } P(\text{raining})}{P(\text{dark cloud, wind speed, humidity})}$$

Bayesian Classifier (Naive Bayes)

- **REAL LIFE EXAMPLE**

Instance	Education	Age	Income	Approved?
1	College	GenX	Upper-Middle	Yes
2	HighSchool	GenZ	Upper-Middle	No
3	GradSchool	GenY	Lower-Middle	No
4	College	GenY	Upper-Middle	Yes
5	HighSchool	GenY	Upper-Middle	Yes
6	GradSchool	GenX	Lower-Middle	Yes
7	HighSchool	GenZ	Lower-Middle	No
8	College	GenY	Lower-Middle	Yes

- ✓ X1:Education=GradSchool
- ✓ X2: Age=GenY
- ✓ X3: Income=Upper-Middle
- ✓ Approved? = ?

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	GenX	Upper-Middle	Yes
2	HighSchool	GenZ	Upper-Middle	No
3	GradSchool	GenY	Lower-Middle	No
4	College	GenY	Upper-Middle	Yes
5	HighSchool	GenY	Upper-Middle	Yes
6	GradSchool	GenX	Lower-Middle	Yes
7	HighSchool	GenZ	Lower-Middle	No
8	College	GenY	Lower-Middle	Yes

- Two possibilities
  - ✓ C1: approved?=yes
  - ✓ C2: approved?=no

Bayesian Classifier (Naive Bayes)

		APPROVED?			
		5 YES		3 NO	
attributes	values	# of instances	probability	# of instances	probability
education	HighSchool	1	1/5	2	2/3
	College	3	3/5	0	0
	GradSchool	1	1/5	1	1/3
age	GenZ	0	0	2	2/3
	GenY	3	3/5	1	1/3
	GenX	2	2/5	0	0
Income	Upper-Middle	3	3/5	1	1/3
	Lower-Middle	2	2/5	2	2/3

- ✓ X1:Education=GradSchool
- ✓ X2: Age=GenY
- ✓ X3: Income=Upper-Middle

Bayesian Classifier (Naive Bayes)

1. P(X|c1)P(c1) c1: approved? = yes
   - P(X{x1,x2,…xn} | approved?=Yes) probability
     - ✓ P(x1|c1) = P(education=GradSchool | approved?=Yes) = 1/5
     - ✓ P(x2|c1) = P(age=GenY | approved?=Yes) = 3/5
     - ✓ P(x3|c1) = P(Income=Upper-Middle| approved?=Yes) = 3/5
     - ✓ P(X|c1) = P(X| approved?=Yes) = 1/5 * 3/5 * 3/5 = 9/125
   - P(c1) = P(approved?=Yes) = 5/8
   - P(X|c1)P(c1) = P(X| approved?=Yes) * P(approved?=Yes)
     - ✓ P(X|c1) * P(c1) = 9/125 * 5/8 = **0.045**

Bayesian Classifier (Naive Bayes)

2. P(X|c2)P(c2)   c2: approved?= no
   - P(x1|c2) = P(education=GradSchool approved?= no) = 1/3
   - P(x2|c2) = P(age=GenY | approved?=no) = 1/3
   - P(x3|c2) = P(Income=Upper-Middle| approved?= no) = 1/3

   - P(X|c2) = P(X| approved?= no) = 1/3 *1/3 * 1/3 = 1/27

   - P(c2) = P(approved?= no) = 3/8

   - P(X|c2)P(c2) = P(X | approved?= no) * P(approved?= no)
     - ✓ P(X|c2) * P(c2) = 1/27 * 3/8 = **0.014**

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	GenX	Upper-Middle	Yes
2	HighSchool	GenZ	Upper-Middle	No
3	GradSchool	GenY	Lower-Middle	No
4	College	GenY	Upper-Middle	Yes
5	HighSchool	GenY	Upper-Middle	Yes
6	GradSchool	GenX	Lower-Middle	Yes
7	HighSchool	GenZ	Lower-Middle	No
8	College	GenY	Lower-Middle	Yes

- ✓ X1:education=GradSchool
- ✓ X2: age=GenY
- ✓ X3: Income=Upper-Middle

- 0.045 > 0.014 result: Approved? =**YES**

Bayesian Classifier (Naive Bayes)

- ✓ Real Life Example 2

	outlook	temperature	humidity	windy	play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rainy	mild	high	false	yes
5	rainy	cool	normal	false	yes
6	rainy	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rainy	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rainy	mild	high	true	no

New instance:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← *Evidence E*

Bayesian Classifier (Naive Bayes)

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	Cool	High	True	?

← *Evidence E*

Bayesian Classifier (Naive Bayes)

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← ***Evidence E***

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

Probability of class "yes"

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

Bayesian Classifier (Naive Bayes)

- Result:

Likelihood of the two classes

For "yes" = $2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.0053$

For "no" = $3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.0206$

Conversion into a probability by normalization:

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

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

Bayesian Classifier (Naive Bayes)

- **Zero Frequency Problem**
  - What if our new instance has the attribute value of "outlook = overcast"?
  - In this case;

    P(outlook=overcast | play=no) = 0/5 = 0

- Without even checking the other features, it's:

  P(X | play=no) = 0

Bayesian Classifier (Naive Bayes)

- zero frequency problem in Bayesian Classifier - *"Laplace smoothing"*

$$P(\text{outlook=sunny} \mid \text{play=no}) = \frac{3 + \mu\, p_1}{5 + \mu}$$

$$P(\text{outlook=overcast} \mid \text{play=no}) = \frac{0 + \mu\, p_2}{5 + \mu}$$

$$P(\text{outlook=raining} \mid \text{play=no}) = \frac{2 + \mu\, p_3}{5 + \mu}$$

$$\text{where } (p_1 + p_2 + p_3) = 1.0$$

Bayesian Classifier (Naive Bayes)

- **zero frequency problem in Bayesian Classifier -** ***"Laplace smoothing"***
  - We could assume they have the same probability

$$p_1 = p_2 = p_3 = 1/3$$

$$P(\text{outlook=sunny} \mid \text{play=no}) = \frac{3 + \mu/3}{5 + \mu} = \frac{3 + 3/3}{5 + 3} = 4/8$$

$$P(\text{outlook=overcast} \mid \text{play=no}) = \frac{0 + \mu/3}{5 + \mu} = \frac{0 + 3/3}{5 + 3} = 1/8$$

$$P(\text{outlook=raining} \mid \text{play=no}) = \frac{2 + \mu/3}{5 + \mu} = \frac{2 + 3/3}{5 + 3} = 3/8$$

Bayesian Classifier (Naive Bayes)

- **Continuous Values**
  - If the attribute value is a continuous value, then our formula is not applicable.
  - For continuous values, we use Gaussian Naïve Bayes.

$$P(X|C) = f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

  - x : attribute value
  - μ : mean
  - σ : standard deviation
  - e = 2.71828

PS: You don’t need to implement the algorithm or formula. It already exists in the scikit-learn library (Python).

Bayesian Classifier (Naive Bayes)

- **Example 3**

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

  - Should this person be approved?
    - education : GradSchool
    - age : 44
    - Income : Lower-Middle

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

P(instance|c1) = P(instance|approved?=yes)

P(instance |c1) = P(education=GradSchool|approved?=yes) * P(age=44| approved?=yes) * P(Income=Lower-Middle| approved?=yes)

P(instance |c1) = 1/5 * P(age=44|approved?=yes) * 2/5

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

- P(age=44|approved?=yes)
  - $\mu_{c1} = 48.4$
  - $\sigma_{c1} = 10.62$

P(age=44 | approved? = yes) $= \dfrac{1}{10{,}63\sqrt{2\pi}} e^{-(44-48{,}4)^2/(2.10{,}62^2)} \cong 0{,}0344$

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

- P(instance|c1) = P(instance|approved?=yes)
  - P(X|C1) = P(X|approved?=yes)
  - P(X|C1) = 1/5 * (0.0344) * 2/5 = **0.0027**

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

P(instance|C2) = P(instance|approved?=no)

- P(instance |C2) = P(education=GradSchool|approved?=no) * P(age=44|approved?=no) * P(Income=Lower-Middle|approved=no)
- P(instance |C2) = 1/3 * P(age=44|approved?=no) * 2/3

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

- P(age=44|approved?=no)
  - $\mu_{c2} = 26.66$
  - $\sigma_{c2} = 9.86$

P(age=44 | approved? = no) $= \frac{1}{9,86\sqrt{2\pi}} e^{-(44-26,66)^2/(2*9,86^2)} \cong 0,0086$

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

- ✓ P(instance|c2) = P(instance|approved?=no)
  - ○ P(X|c2) = P(X|approved?=no)
  - ○ P(X|c2) = 1/3 x (0,0086) x 2/3 = **0.0011**

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

- We have:
  - P(X|C1)=P(X|approved?=yes)=1/5 x (0,0344) x (2/5) = 0.0027
  - P(X|C2)=P(X|approved?=no)=1/3 x (0,0086) x (2/5) = 0.0011

Bayesian Classifier (Naive Bayes)

Instance	Education	Age	Income	Approved?
1	College	60	Upper-Middle	Yes
2	HighSchool	22	Upper-Middle	No
3	GradSchool	38	Lower-Middle	No
4	College	40	Upper-Middle	Yes
5	HighSchool	40	Upper-Middle	Yes
6	GradSchool	60	Lower-Middle	Yes
7	HighSchool	20	Lower-Middle	No
8	College	42	Lower-Middle	Yes

- Result:
  - 0.0027 > 0.0011 **YES**