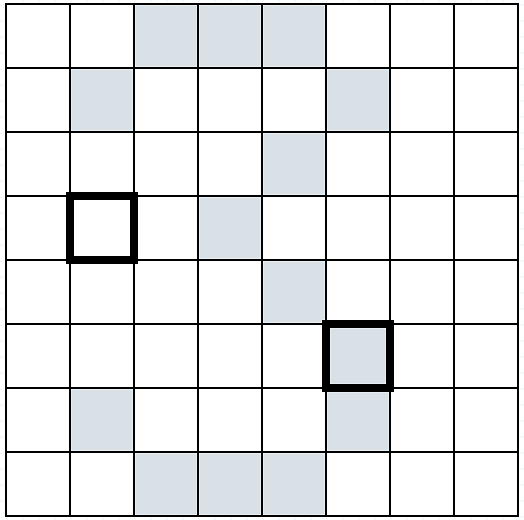
الجامعة السورية الخاصة
SYRIAN PRIVATE UNIVERSITY

Week 12 | كلية الهندسة المعلوماتية | مقرر تعلم الآلة

Naïve Bayes Classifier

د. رياض سنبل

The Idea



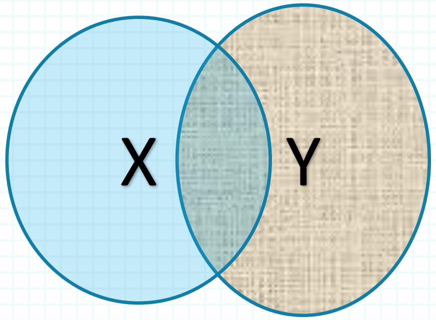
The theory!

- **Our main goal in machine learning:**

  Estimate $P(Y|D)$ Y is the output, D is the data.

- For a specific input X,

  What is the meaning of $P(Y|X)$? And how can we calculate it using D?
  Remember: $P(Y|X)=P(X,Y)/P(X)$



- Is it feasible? Why?
- Btw, Why KNN is a good estimator for the last formula?

The theory!

- Naïve Bayes solution :

Posterior Probability $P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)}$

prior probability
Not a big problem

The likelihood
Generative Probability
(the main challenge)

We can forgot it, it is just a normalizer!

So:

$$\hat{y} = \underset{y}{arg\max}\big(P(Y|X)\big) = \underset{y}{arg\max}\left(\frac{P(X|Y).P(Y)}{P(X)}\right) = \underset{y}{arg\max}(P(X|Y).P(Y))$$

Naïve Bayes solution

$$P(X = x|Y = y) = P(X_1 = x_1, \ldots, X_n = x_n|Y = y) \approx P(X_1 = x_1|Y = y) \ldots . P(X_n = x_n|Y = y)$$

The theory!

Multiplying lots of probabilities, which are between 0 and 1 by definition, can result in floating-point underflow.

Then:

$$\hat{y} = \underset{y}{argmax}(P(X|Y).P(Y)) = \underset{y}{argmax}(P(Y).\prod_i P(X_i|Y)) =$$

$$= \underset{y}{argmax}\left(\log(P(Y)) + \sum_i \log(P(X_i|Y))\right)$$

What should we learn from the data in Naïve Bayes?

Example: Play Tennis

What do we need in order to use Naïve Bayes?

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Example: Play Tennis

P(Play=*Yes*) =

P(Play=*No*) =

$P(X_i|Y)$

Outlook	Play= *Yes*	Play= *No*
Sunny		
Overcast		
Rain		

Temp	Play= *Yes*	Play= *No*
Hot		
Mild		
Cool		

Humidity	Play= *Yes*	Play= *No*
High		
Normal		

Wind	Play= *Yes*	Play= *No*
Strong		
Weak		

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Example: Play Tennis

$P(\text{Play}=Yes) = 9/14 \quad P(\text{Play}=No) = 5/14$

Outlook	Play=Yes	Play=No
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temp	Play=Yes	Play=No
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Humidity	Play=Yes	Play=No
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=Yes	Play=No
Strong	3/9	3/5
Weak	6/9	2/5

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Example

- Test Phase
  - Given a new instance,
    $\mathbf{x}'$=(Outlook=*Sunny*, Temperature=*Cool*, Humidity=*High*, Wind=*Strong*)
  - Prediction

$$P(Yes|\mathbf{x}'): [P(Sunny|Yes)P(Cool|Yes)P(High|Yes)P(Strong|Yes)]P(Play=Yes) = 0.0053$$
$$P(No|\mathbf{x}'): [P(Sunny|No)\,P(Cool|No)P(High|No)P(Strong|No)]P(Play=No) = 0.0206$$

Given the fact $P(Yes|\mathbf{x}') < P(No|\mathbf{x}')$, we label $\mathbf{x}'$ to be "*No*".