Naïve Bayes Classifier

Outline

- Probability Basics
- Naïve Bayes Theorem
- Naïve Bayes Classification
- Example: Play Tennis
- Laplace Smoothing
- Applications

Probability Basics

We all know that when flip a coir

the probability of getting head

p(head) = ½

Probability Basics

Pick a random card,
what is the probability of
getting a **queen**?

$P(\text{queen}) = 4/52 = 1/13$

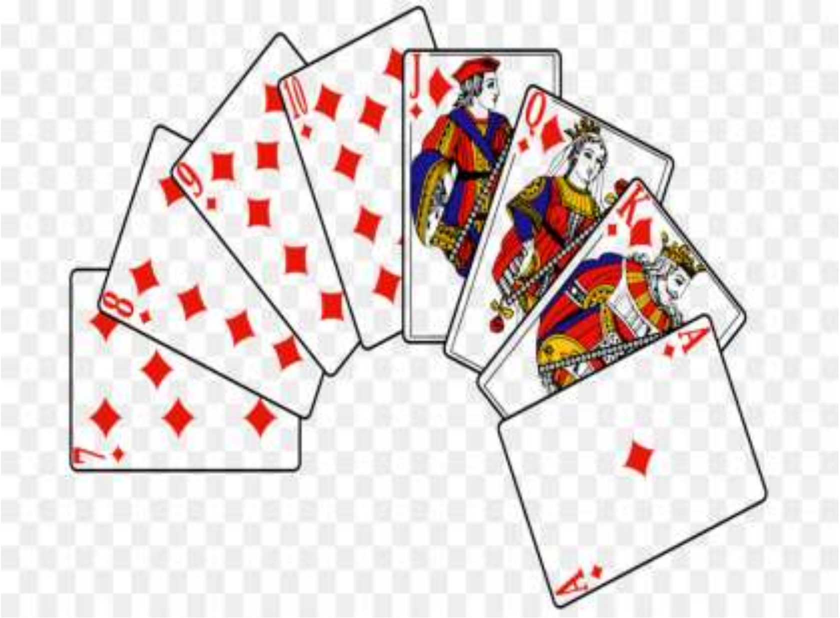
(4 queen, 52 total cards)

Probability Basics

Pick a random card and we know that it is a **diamond**. Now what is the probability of that card being a **queen**?

P(queen/diamond) = 1/

(1 queen, 13 diamond cards)



Probability Basics

P(queen/diamond) = 1/13

It is called Conditional probability

P(A/B) = probability of event A, knowing that event B has already occurred.

Naïve Bayes Theorem

Naïve Bayes Classifier

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

Thomas Bayes
1702 - 1761

Naïve Bayes Theorem

Now let context our queen and diamond problem…

$$P(queen/diamond) = \frac{P(diomond|queen) * p(queen)}{p(diomond)}$$

P(diamond/queen) = ¼

P(queen) = 1/13

P(diamond) = 1/4

$$= \frac{(\frac{1}{4})*(\frac{1}{13})}{(\frac{1}{4})}$$

$$= 1/13$$

Naïve Bayes Classification

Let we have D training data set having
n-dimensional attribute vector, $\boldsymbol{X = (x_1, x_2, x_3, \ldots, x_n)}$ and
m classes, $\boldsymbol{C_1, C_2, C_3, \ldots, C_m}$
Now, given a tuple, ***X*** the classifier will predict that ***X*** belongs to the class having the highest posterior probability, conditioned on ***X***. That is ***X*** belongs to the class $\boldsymbol{C_i}$ if and only if

$$\boldsymbol{P(C_i \mid X) > P(C_j \mid X)} \quad \text{for } 1 \le j \le m, j \ne i.$$

Now the formula is, $\boldsymbol{P(C_i \mid X)} = \dfrac{\boldsymbol{P(X \backslash Ci) * P(Ci)}}{\boldsymbol{P(X)}}$

As $\boldsymbol{P(X)}$ is constant for all classes, so only $\boldsymbol{P(X \backslash Ci) * P(Ci)}$ need to be calculated.

Example: Play Tennis

Here,

C_i = {Yes, No}
X ={outlook, Temperature, Humidity, Winc

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

Learning Phase:

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

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

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

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: Play Tennis

Learning Phase:

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

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

Temperature	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

Example: Play Tennis

Test Phase:

- Given a new instance,

  x′=(Outlook=Sunny, Temperature=Cool, Humidity=High, Wind=Strong)

- Look up tables:

  P(Play=*Yes*) = 9/14

  P(Outlook=*Sunny*|Play=*Yes*) = 2/9

  P(Temperature=*Cool*|Play=*Yes*) = 3/9

  P(Huminity=*High*|Play=*Yes*) = 3/9

  P(Wind=*Strong*|Play=*Yes*) = 3/9

  P(Play=*No*) = 5/14

  P(Outlook=*Sunny*|Play=*No*) = 3/5

  P(Temperature=*Cool*|Play==*No*) = 1/5

  P(Huminity=*High*|Play=*No*) = 4/5

  P(Wind=*Strong*|Play=*No*) = 3/5

- MAP rule:

  P(*Yes*|**x′**): [P(*Sunny*|Yes)P(*Cool*|Yes)P(*High*|Yes)P(*Strong*|Yes)]P(Play=*Yes*) = 0.0053

  P(*No*|**x′**): [P(*Sunny*|No) P(*Cool*|No)P(*High*|No)P(*Strong*|No)]P(Play=*No*) = 0.0206

Given the fact P(*Yes*|x′) < P(*No*|x′), we label x′ to be *“No”*.

Example: Play Tennis

Test Phase:

- MAP rule:

$$P(Yes|\mathbf{x}'): [P(Sunny|Yes)P(Cool|Yes)P(High|Yes)P(Strong|Yes)]P(\text{Play}=Yes) = 0.0053$$

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

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

- Normalization:

We know that sum of possible event is 1, then we normalize above calculation as bellow:

$$\left.\begin{aligned} P(Yes|\mathbf{x}') &:= \frac{0.0053}{0.0053+0.0206} = 0.2046 \\ P(No\;|\mathbf{x}') &:= \frac{0.0206}{0.0053+0.0206} = 0.7954 \end{aligned}\right\} \quad \text{Sum is 1 (one)}$$

Example: Play Tennis

Test Phase:

- Given a new special instance,
  x′=(Outlook=Overcast, Temperature=Hot, Humidity=High, Wind=Strong)
- MAP rule

P(*Yes*|**x′**): [P(*Overcast*|Yes)P(*Hot*|*Yes*)P(High|*Yes*)P(Strong|*Yes*)]P(Play=*Yes*)

= (4/9) * (3/9) * (2/9) * (3/9) * (9/14) = **0.0071**

P(*No*|**x′**): [P(*Overcast*|No) P(*Hot*|No)P(High|*No*)P(Strong |*No*)]P(Play=*No*)

= (0/5) * (2/5) * (4/5) * (3/5) * (5/14) = **0**

Given the fact P(*Yes*|x′) > P(*No*|x′), we label x′ to be *"Yes"*.

Laplace Estimation/Smoothing:

P(*No*|**x′**): ((0+3)/(5+3)) * (2/5) * (4/5) * (3/5) * (5/14) = **0.0257**

Now, given the fact P(*Yes*|x′) < P(*No*|x′), we label x′ to be *"No"*.

Naïve Bayes Classification

Applications:

- To mark an email as spam, or not spam ?
- Classify a news article about technology, politics, or sports ?
- Check a piece of text expressing positive emotions, or negative emotions?
- Also used for face recognition software's.
- Etc.

Thank you everyone.