Naïve Bayes Classifier

Jing-Doo Wang

jdwang@asia.edu.tw

Asia University

Bayes' Theorem

Bayes' Formula:

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

Where: `P(A)` is the prior probability of A occurring independently, for us this is `P(HIV)`.

`P(B)` is the prior probability of B occurring independently, for us this is `P(Positive)`.

`P(A|B)` is the posterior probability of A occurring given B occurs, for us this is `P(HIV | Positive)`. This is the probability that an individual has HIV given their test results are positive and what we're trying to calculate.

`P(B|A)` is the likelihood probability of B occurring, given A occurs. In our example this is `P(Positive | HIV)`. This value is given to us.

Stringing these together we get:

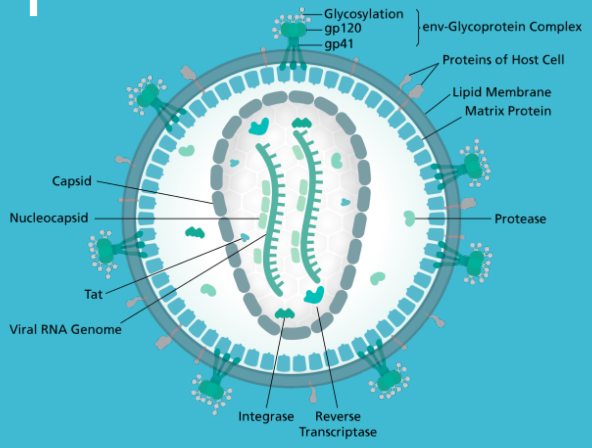
`P(HIV | Positive) = ((P(HIV) * P(Positive | HIV)) / P(Positive)`

Thus the probability of getting a positive HIV test result `P(HIV)` becomes:

`P(Positive) = [P(HIV) * Sensitivity] + [P(~HIV) * (1-Specificity)]`

From: https://etav.github.io/algorithms/bayes_theorem_scratch.html

a doctor testing patients for HIV



https://en.wikipedia.org/wiki/Structure_and_genome_of_HIV

Here we'll create a fictitious world in which we're a doctor testing patients for HIV, subject to the following assumptions:

`P(HIV)` = The odds of a person having HIV is .015 or 1.5%

`P(Positive)` = The probability the test results are positive

`P(Negative)` = The probability the test results are negative.

`P(Positive | HIV)` = The probability the test results are positive given someone has HIV. This is also called Sensitivity or True Positive Rate. We'll assume the test is correct .95 or 95% of the time.

`P(Negative | ~HIV)` = The probability the test results are negative given someone does not have HIV. This is also called Specificity or True Negative Rate. We'll assume this is also correct .95 or 95% of the time.

From: https://etav.github.io/algorithms/bayes_theorem_scratch.html

Priors

Calculations - Priors

```
#performing calculations:

p_hiv = .015 #P(HIV) assuming 1.5% of the population has HIV

p_no_hiv = .98 # P(~HIV)

p_positive_hiv = .95 #sensitivity

p_negative_hiv = .95#specificity

#P(Positive)
p_positive = (p_hiv * p_positive_hiv) + (p_no_hiv * (1-p_negative_hiv))
print "The probability of getting a positive test result is:", p_positive, "this is our prior"
```

```
The probability of getting a positive test result is: 0.06325 this is our prior
```

Using this prior we can calculate our posterior probabilities as follows:

The probability of an individual having HIV given their test result is positive.

`P(HIV|Positive) = (P(HIV) * Sensitivity)) / P(Positive)`

The probability of an individual not having HIV given their test result is positive.

`P(~HIV|Positive) = (P(~HIV) * (1-Sensitivity))) / P(Positive)`

Note: the sum of posteriors must equal one because combined they capture all possible states within our set of probabilities.

From: https://etav.github.io/algorithms/bayes_theorem_scratch.html

Posterios

Calculations - Posteriors

```
#P(HIV | Positive)
p_hiv_positive = (p_hiv * p_positive_hiv) / p_positive

print "The probability of a person having HIV, given a positive test result is:", p_hiv_positive
```

```
The probability of a person having HIV, given a positive test result is: 0.225296442688
```

```
#P(~HIV | Positive)
p_positive_no_hiv = 1 - p_positive_hiv
p_no_hiv_positive = (p_no_hiv * p_positive_no_hiv) / p_positive

print "The probability of an individual not having HIV a positive test result is:", p_no_hiv_positive
```

```
The probability of an individual not having HIV a positive test result is: 0.774703557312
```

From: https://etav.github.io/algorithms/bayes_theorem_scratch.html

What is the meaning of "Naïve" Bayes?

- P(D) : the probability of a given data sample.
- P(D|C): the probability of the data D point belonging to the class C.
- Assume that each feature is independent of each other.
  - $P(D|C) = P(D_1|C)*P(D_2|C)*\ldots*P(D_n|C)$
  - $D_1, D_2, \ldots, D_n$: feature

From: Learning Data Mining with Python - Second Edition

Example: Naive Bayes

Example of Naïve Bayes Classifier

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

A: attributes

M: mammals

N: non-mammals

$$P(A|M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A|N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A|M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A|N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

P(A|M)P(M) > P(A|N)P(N)

=> Mammals

From: http://slideplayer.com/slide/4732608/