GENERAL ASSEMBLY

INTRO TO DATA SCIENCE

NAÏVE BAYESIAN CLASSIFICATION

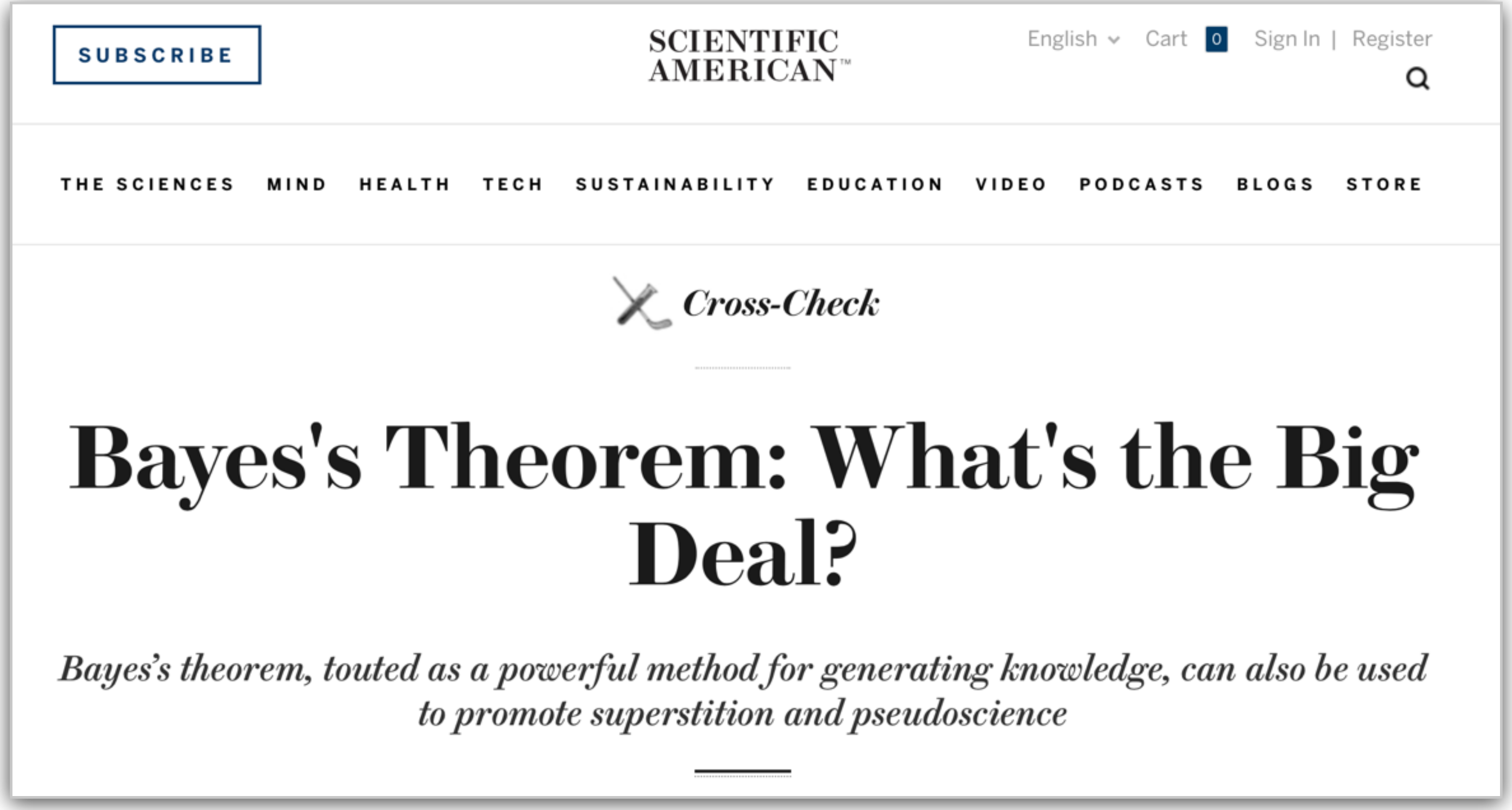
DATA SCIENCE IN THE NEWS

DataTau new | comments | leaders | submit login

1. ▲ Using Pipeline and GridSearchCV for More Compact and Comprehensive Code (civisanalytics.com)
   3 points by michelangelo 8 hours ago | discuss
2. ▲ Bayes's Theorem: What's the big deal? (scientificamerican.com)
   4 points by jpiburn 12 hours ago | discuss
3. ▲ Nando de Freitas Machine Learning course Slides and Videos (ox.ac.uk)
   6 points by Anon84 23 hours ago | discuss
4. ▲ Data manipulation with tidyr (datascienceplus.com)
   2 points by klo99 17 hours ago | discuss
5. ▲ Apache Spark 1.6 Release Notes with Examples (databricks.com)
   4 points by brakmic 1 day ago | discuss
6. ▲ Hastie - Statistical Learning with Sparsity: The Lasso and Generalizations (stanford.edu)
   3 points by dmagee 1 day ago | discuss
7. ▲ N.F.L. Playoff Picture: Every Team's Remaining Paths to the Postseason (nytimes.com)
   5 points by Veerle 2 days ago | discuss
8. ▲ How to make Tufte's discrete sparklines using d3.js (blogspot.com.es)
   4 points by rooviz 2 days ago | discuss
9. ▲ Ggplot2 quick reference by tasks (r-statistics.co)
   8 points by selva86 3 days ago | discuss
10. ▲ Generating Synthetic Data with Random Forests (scweiss.blogspot.com)
    5 points by assumednormal 2 days ago | discuss
11. ▲ Feature selection approaches with R (r-statistics.co)
    6 points by selva86 3 days ago | discuss
12. ▲ Doing Data Science at Twitter (medium.com)
    5 points by Veerle 4 days ago | discuss
13. ▲ Regression Models with R (collection of posts) (datascienceplus.com)
    9 points by klo99 5 days ago | discuss
14. ▲ Sentiment Analysis on Donald Trump using R and Tableau (datascienceplus.com)
    5 points by klo99 4 days ago | discuss
15. ▲ Attention and Memory in Deep Learning and NLP (wildml.com)
    3 points by dbritz 3 days ago | discuss
16. ▲ Rodeo 1.2: Python Paths, Interrupt, Stickers (yhathq.com)
    11 points by glamp 8 days ago | 1 comment
17. ▲ Google scholar scraping with rvest package (datascienceplus.com)
    4 points by klo99 5 days ago | discuss
18. ▲ Evaluation of Deep Learning Toolkits (github.com)
    6 points by rishy 6 days ago | discuss

http://www.datatau.com/

AGENDA



SUBSCRIBE

SCIENTIFIC AMERICAN™

English Cart 0 Sign In | Register

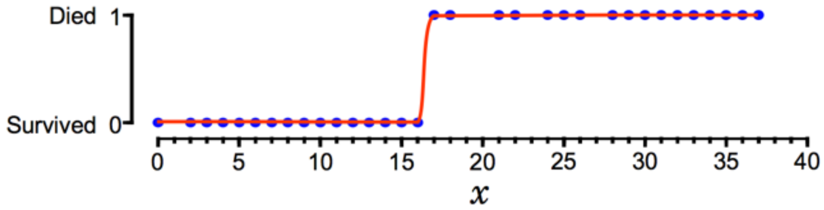
THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS STORE

Cross-Check

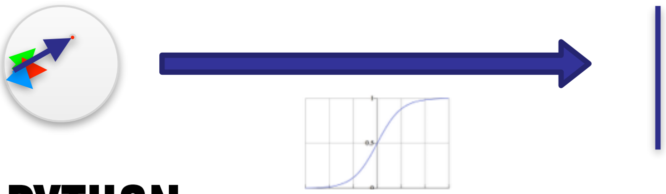
Bayes's Theorem: What's the Big Deal?

Bayes's theorem, touted as a powerful method for generating knowledge, can also be used to promote superstition and pseudoscience

LAST TIME:



I. LOGISTIC REGRESSION
II. OUTCOME VARIABLES
III. ERROR TERMS
IV. INTERPRETING RESULTS



LAB: IMPLEMENTING LOGISTIC REGRESSION IN PYTHON

QUESTIONS?

QUESTIONS?

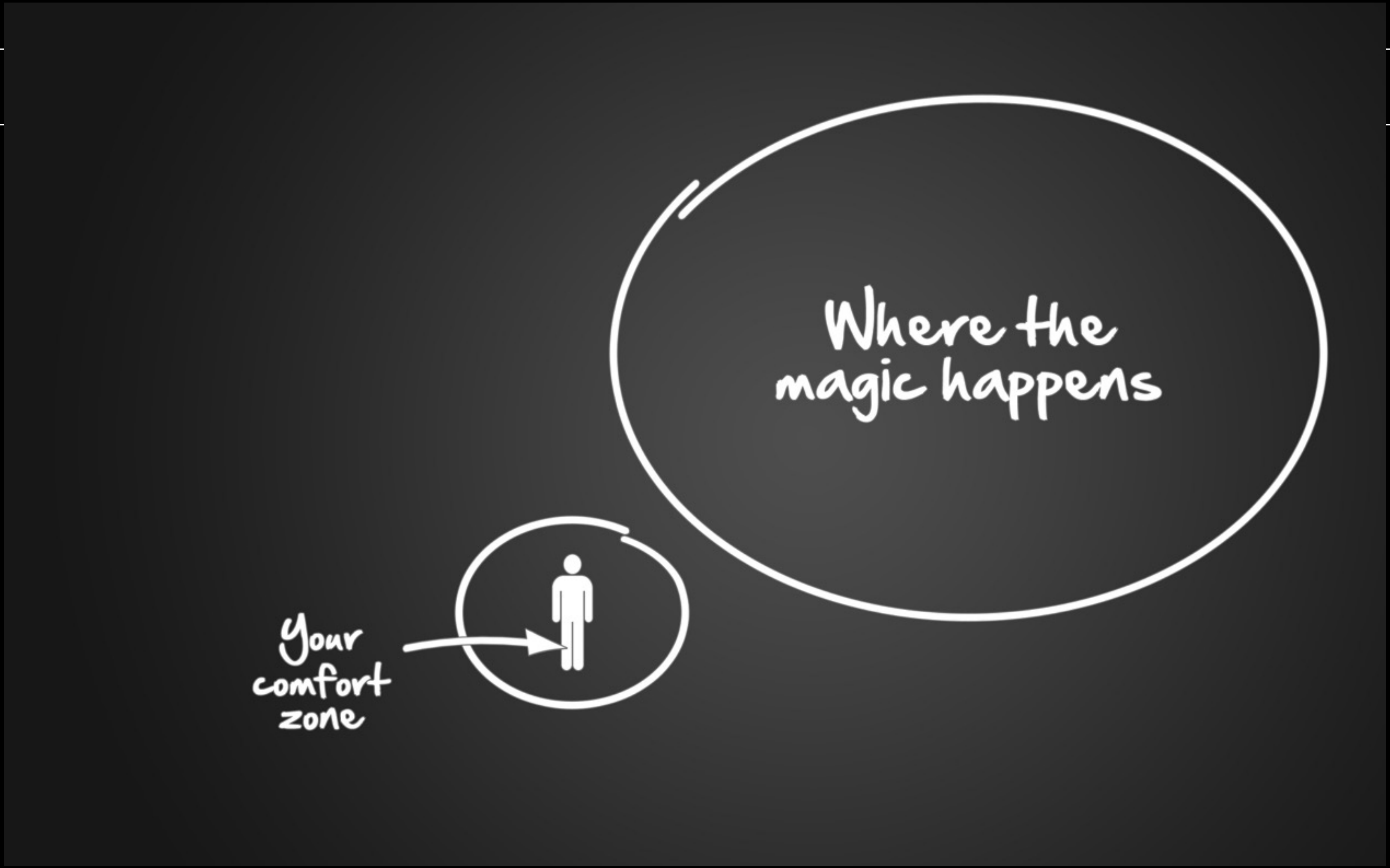
WHAT WAS THE MOST INTERESTING THING YOU LEARNT?

WHAT WAS THE HARDEST TO GRASP?

GENERAL ASSEMBLY

HOW'S THE HOMEWORK GOING?


TIGERS


Where the
magic happens
Your
comfort
zone

I. INTRO TO PROBABILITY
II. NAÏVE BAYESIAN CLASSIFICATION

LAB:
III. NAÏVE BAYES CLASSIFICATION IN PYTHON

- UNDERSTAND PROBABILITY AND CONDITIONAL PROBABILITY
- GET AN INTUITIVE SENSE FOR BAYES' THEOREM
- UNDERSTAND WHY "NAÏVE" BAYES
- BE ABLE TO PERFORM CLASSIFICATION WITH NAIVE BAYES IN PYTHON

INTRO TO PROBABILITY

BEWARE: EQUATIONS AHEAD !

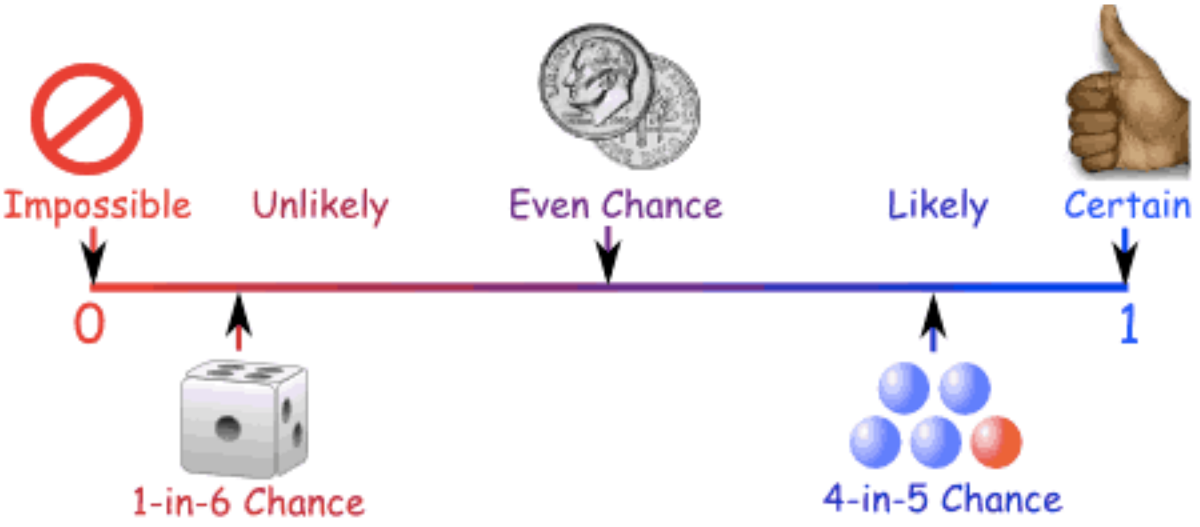
Q: What is a **probability***?*

YOU TELL ME

The probability p(A) for some event A *is number between 0 and 1 that characterizes the* ***likelihood*** *the event A will occur.*

Q: If the probability is zero, how often do we expect that event to occur?

***The probability p(A) for some event A** is number between 0 and 1 that characterizes the **likelihood** the event A will occur.*



https://www.mathsisfun.com/data/images/probability-line.gif

Ω

the **SAMPLE SPACE** is the set of all possible events

EXAMPLES?

Ω

the **SAMPLE SPACE** is the set of all possible events

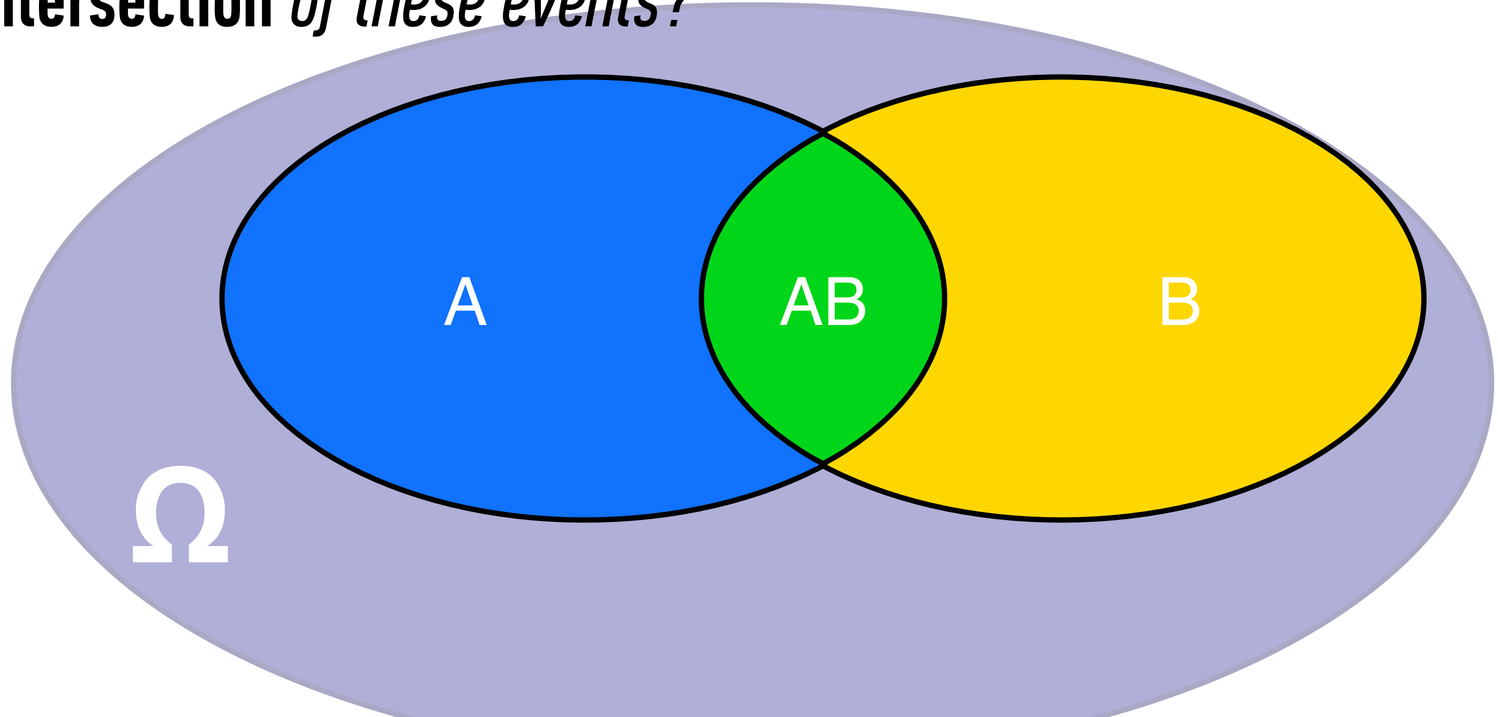
$p(\Omega) = ?$

Ω

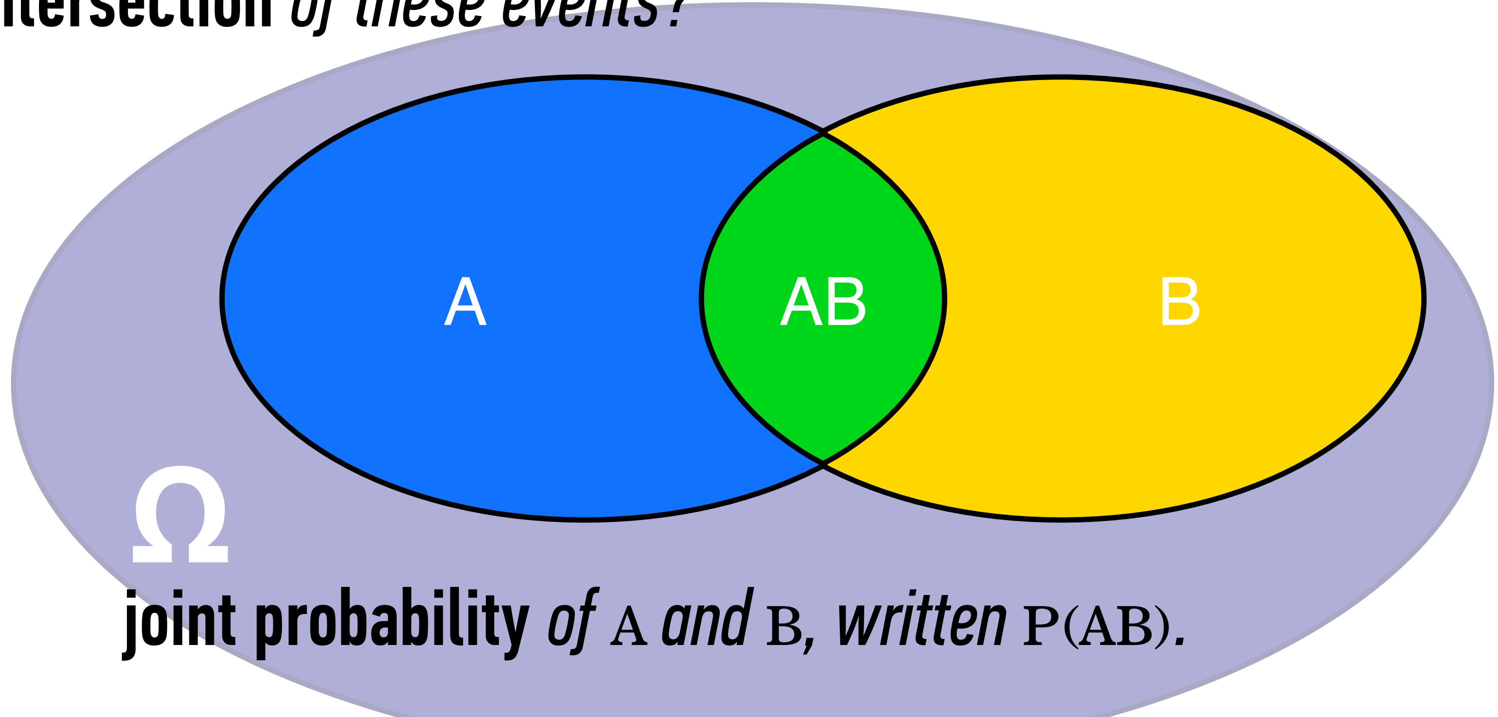
the **SAMPLE SPACE** is the set of all possible events

$p(\Omega) = 1$

Q: Consider two events A & B. *How can we characterize the* **intersection** *of these events?*



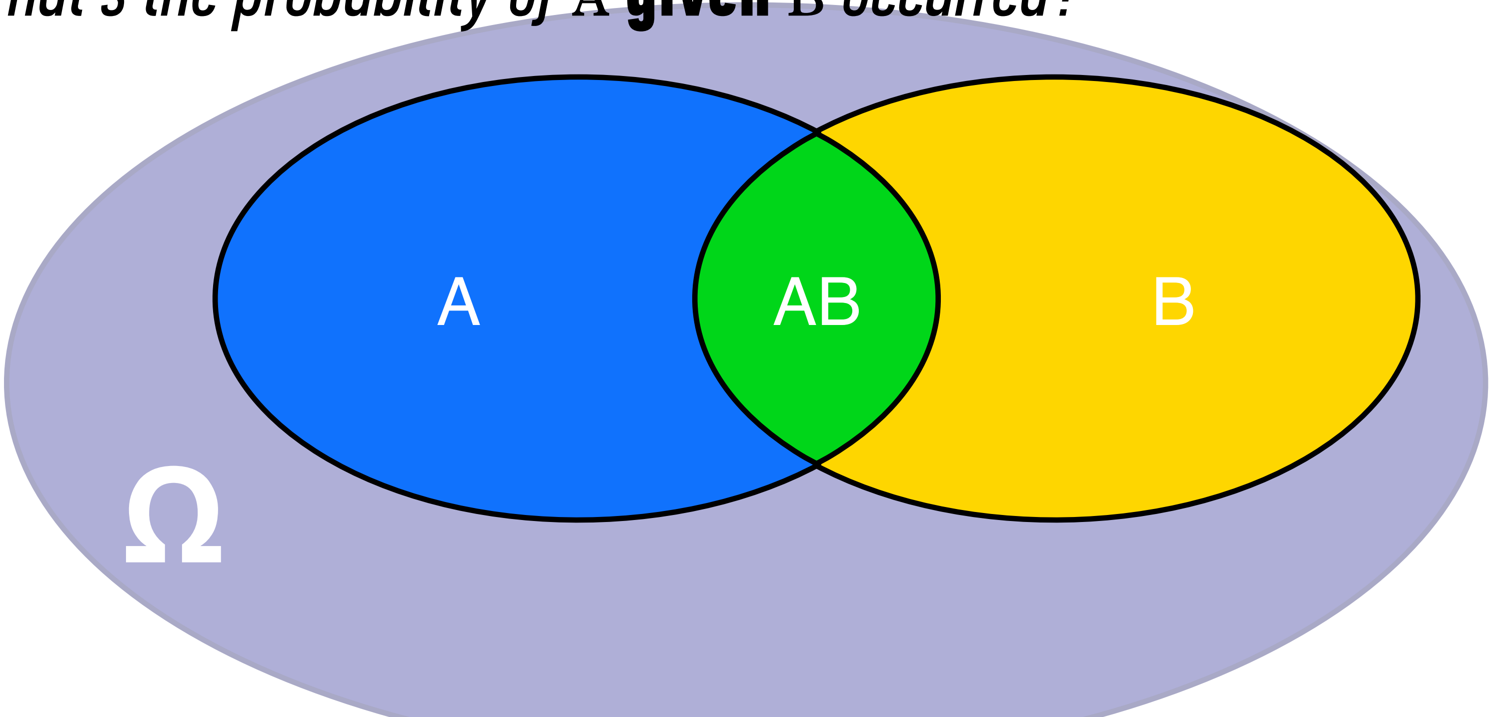
Q: Consider two events A & B. *How can we characterize the* **intersection** *of these events?*



joint probability *of* A *and* B, *written* P(AB).

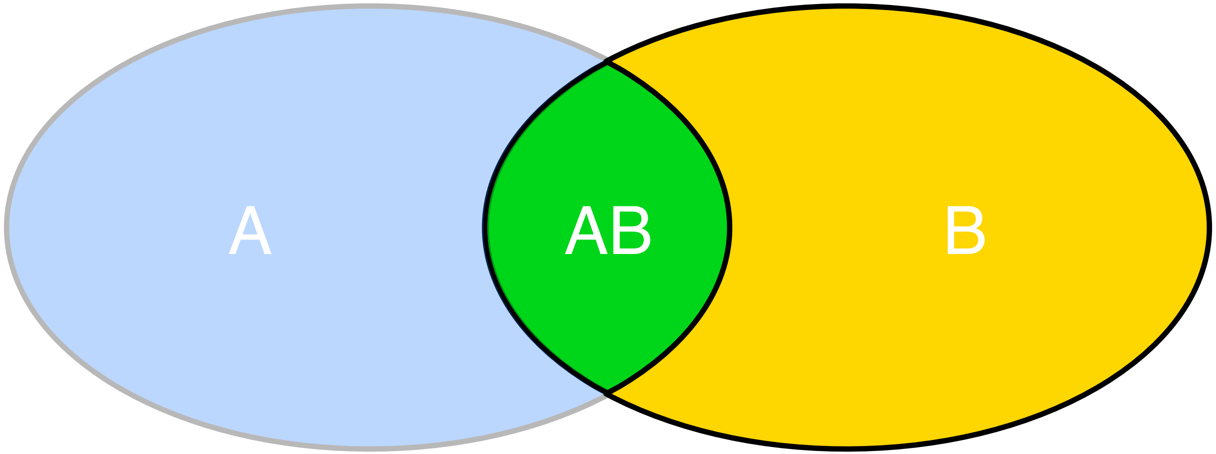
Suppose event B *has occurred*

What's the probability of A **given** B *occurred?*



Suppose event B *has occurred*

What's the probability of A **given** B *occurred?*



NOTE

This information about B transforms the sample space.

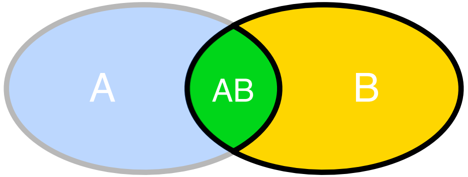
The intersection of A & B *divided by region* B.

Suppose event B *has occurred*

What's the probability of A **given** B *occurred?*

This is called the **conditional probability** *of* A *given* B

written $P(A|B) = P(AB) / P(B)$.



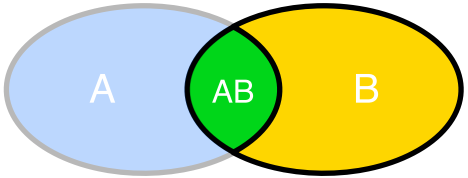
Suppose event B *has occurred*

What's the probability of A **given** B *occurred?*

This is called the **conditional probability**

of A *given* B

written $P(A|B) = P(AB) / P(B)$

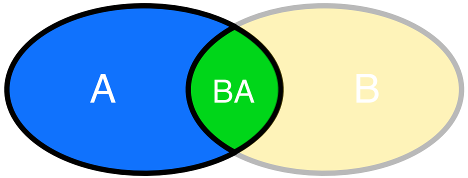


or $P(A|B)\,P(B) = P(AB)\ldots$

Now let's ask the converse question:
what is $P(B|A)$ *?*

This is called the **conditional probability**
of B *given* A

written $P(B|A) = P(BA) / P(A)$



or $P(B|A)\,P(A) = P(BA)\ldots$

Let's recap

Let's recap

$$P(AB) = P(A|B) * P(B)$$

conditional probability of A given B

Let's recap

$P(AB) = P(A|B) * P(B)$ *conditional probability of A given B*

$P(BA) = P(B|A) * P(A)$ *by substitution*

Let's recap

$P(AB) = P(A|B) * P(B)$ *conditional probability of A given B*

$P(BA) = P(B|A) * P(A)$ *by substitution*

But $P(AB) = P(BA)$ *since event* $AB =$ *event* BA

Let's recap

$P(AB) = P(A|B) * P(B)$ *conditional probability of A given B*

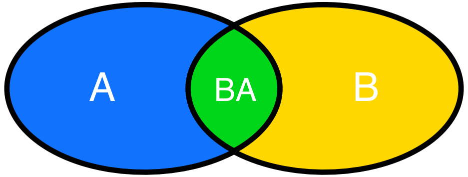
$P(BA) = P(B|A) * P(A)$ *by substitution*

But $P(AB) = P(BA)$ *since event* $AB =$ *event* BA

$$P(A|B) * P(B) = P(B|A) * P(A)$$

This result is called **Bayes' theorem**

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



BAYES' THEOREM: EXAMPLE

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

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

A = breast cancer
B = positive mammogram

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

A = breast cancer
B = positive mammogram

> 1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

$P(A) = ?$

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

A = breast cancer
B = positive mammogram

$$P(A) = 0.01$$

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

A = breast cancer
B = positive mammogram

$P(A) = 0.01$
$P(B|A) = ?$

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

> 1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

A = breast cancer
B = positive mammogram

$P(A) = 0.01$
$P(B|A) = 0.80$

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

A = breast cancer
B = positive mammogram

$P(A) = 0.01$
$P(B|A) = 0.80$
$P(B|\sim A) = ?$

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

A = breast cancer
B = positive mammogram

$P(A) = 0.01$
$P(B|A) = 0.80$
$P(B|{\sim}A) = 0.096$

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

A = breast cancer
B = positive mammogram

$$P(A) = 0.01$$
$$P(B|A) = 0.80$$
$$P(B|\sim A) = 0.096$$

$$P(A|B) = ?$$

What is the probability that she actually has breast cancer?

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

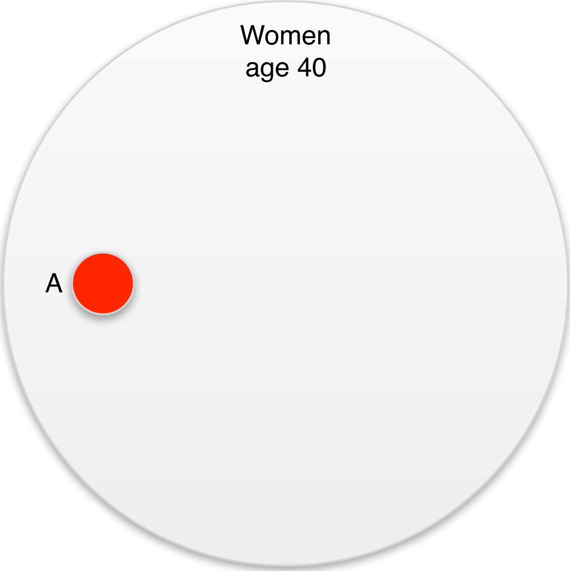
BAYES' THEOREM: EXAMPLE

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

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

A = breast cancer
B = positive mammogram

$P(A) = 0.01$
$P(B|A) = 0.80$
$P(B|\sim A) = 0.096$



$P(A|B) = ?$

What is the probability that she actually has breast cancer?

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

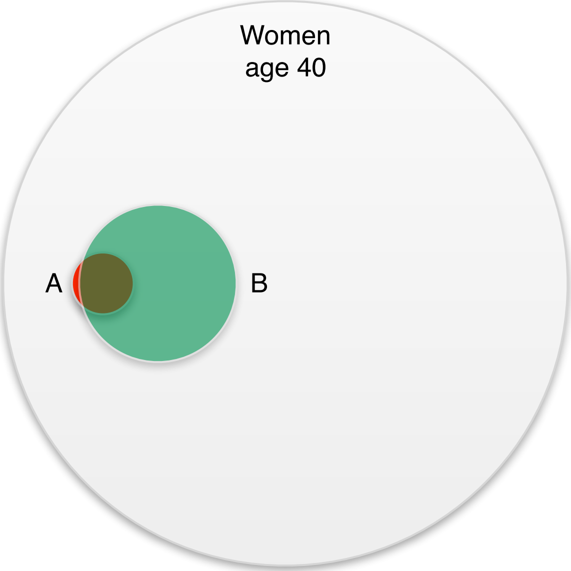
$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

A = breast cancer
B = positive mammogram

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

$P(A) = 0.01$
$P(B|A) = 0.80$
$P(B|\sim A) = 0.096$

$P(A|B) = ?$



What is the probability that she actually has breast cancer?

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

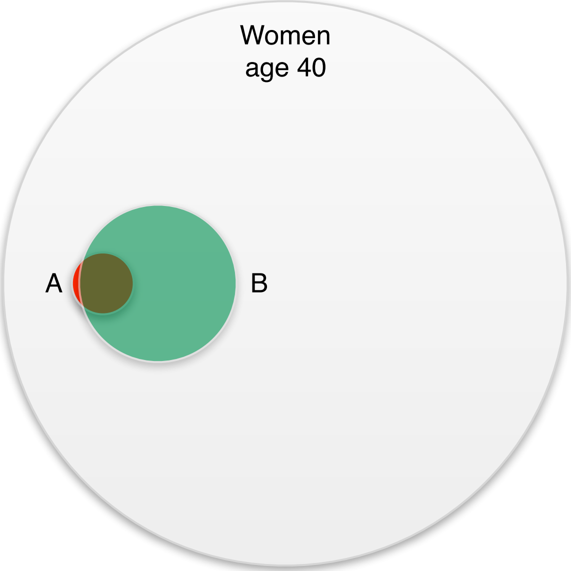
$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

A = breast cancer
B = positive mammogram

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

$P(A) = 0.01$
$P(B|A) = 0.80$
$P(B|{\sim}A) = 0.096$
$P(B) = ?$

$P(A|B) = ?$



What is the probability that she actually has breast cancer?

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

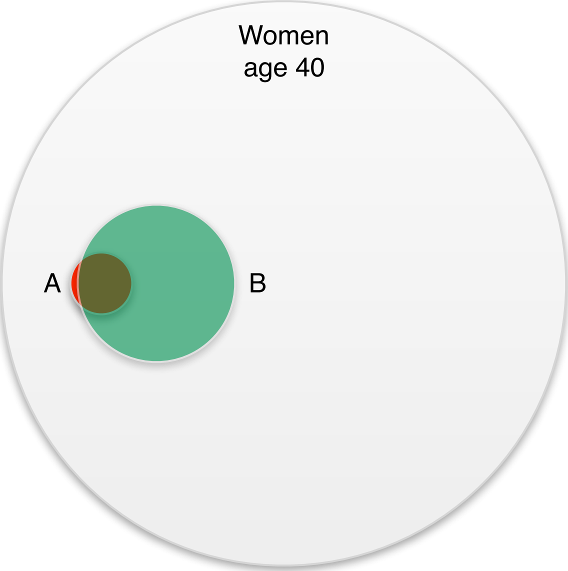
$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

A = breast cancer
B = positive mammogram

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

$$P(A) = 0.01$$
$$P(B|A) = 0.80$$
$$P(B|\sim A) = 0.096$$
$$P(B) = P(B|A)*P(A) + P(B|\sim A)*P(\sim A)$$
$$= 0.80 * 0.01 + 0.096 * 0.99 = 0.10304$$

$$P(A|B) = ?$$



What is the probability that she actually has breast cancer?

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

BAYES' THEOREM: EXAMPLE

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

A = breast cancer
B = positive mammogram

1% of women at age forty who participate in routine screening have breast cancer. 80% of women with breast cancer will get positive mammograms. 9.6% of women without breast cancer will also get positive mammograms. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

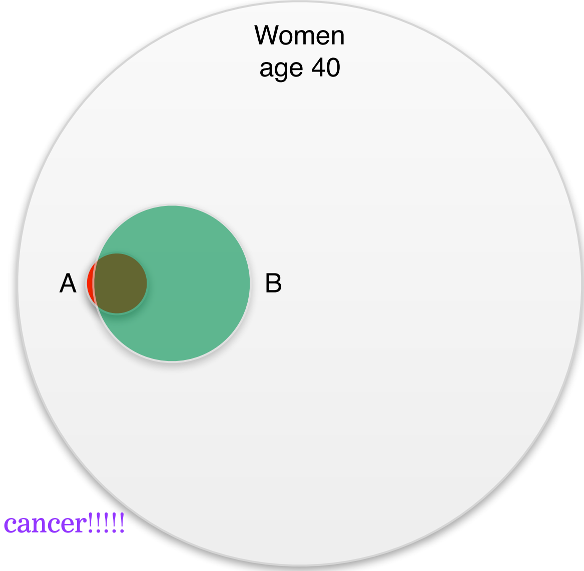
$P(A) = 0.01$

$P(B|A) = 0.80$

$P(B|{\sim}A) = 0.096$

$$P(B) = P(B|A)*P(A) + P(B|{\sim}A)*P({\sim}A)$$
$$= 0.80 * 0.01 + 0.096 * 0.99 = 0.10304$$

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{0.8 * 0.01}{0.10304} = 0.0776$$



What is the probability that she actually has breast cancer? About 7.8% chance of actually having cancer!!!!!

http://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

INTERPRETATIONS OF PROBABILITY

There are 2 interpretations of probability:

There are 2 interpretations of probability:

*The **frequentist** interpretation regards an event's probability as its limiting frequency across a very large number of trials.*

There are 2 interpretations of probability:

*The **frequentist** interpretation regards an event's probability as its limiting frequency across a very large number of trials.*

*The **Bayesian** interpretation regards an event's probability as a "degree of belief," which can apply even to events that have not yet occurred.*

INDEPENDENT EVENTS

Q: When are 2 events **independent***?*

Q: When are 2 events **independent***?*

A: Information about one does not affect the probability of the other.

Q: When are 2 events **independent***?*

A: Information about one does not affect the probability of the other.

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

Q: When are 2 events **independent***?*

A: Information about one does not affect the probability of the other.

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

using the definition of conditional probability:

$P(A|B) = P(AB) / P(B) = P(A)$ → $$P(AB) = P(A) * P(B)$$

ADDITIONAL RESOURCES

http://www.yudkowsky.net/rational/bayes

https://en.wikipedia.org/wiki/Bayes%27_theorem

http://betterexplained.com/articles/an-intuitive-and-short-explanation-of-bayes-theorem/

http://alexanderetz.com/2015/08/09/understanding-bayes-visualization-of-bf/

http://jakevdp.github.io/blog/2015/08/07/frequentism-and-bayesianism-5-model-selection/

What's the difference between frequentist and Bayesian interpretations of probability?

What does Bayes Theorem allow us to do?

NAÏVE BAYESIAN CLASSIFICATION

Suppose we have a dataset with features $x_1, \ldots, x_n$ *and a class label* C*. What can we say about classification using Bayes' theorem?*

Suppose we have a dataset with features $x_1, \ldots, x_n$ and a class label C. What can we say about classification using Bayes' theorem?

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

Bayes' theorem can help us to determine the probability of a record belonging to a class, given the data we observe.

source: Data Analysis with Open Source Tools, by Philipp K. Janert. O'Reilly Media, 2011.

Each term in this relationship has a name, and each plays a distinct role in any Bayesian calculation (including ours).

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

This term is the **likelihood function**. *It represents the joint probability of observing features* $\{x_i\}$ *given that that record belongs to class* C.

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

This term is the **likelihood function**. *It represents the joint probability of observing features $\{x_i\}$ given that that record belongs to class C.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

Thought experiment: How might we calculate the likelihood function?

This term is the **likelihood function**. *It represents the joint probability of observing features $\{x_i\}$ given that that record belongs to class C.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

We can observe the value of the likelihood function from the training data.

This term is the **prior probability** *of* C*. It represents the probability of a record belonging to class* C *before the data is taken into account.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

This term is the **prior probability** *of* C. *It represents the probability of a record belonging to class* C *before the data is taken into account.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

The value of the prior is also observed from the training data.

This term is the **normalization constant**. *It doesn't depend on* C, *and is generally ignored until the end of the computation.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

This term is the **normalization constant**. *It doesn't depend on* C, *and is generally ignored until the end of the computation.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

The normalization constant doesn't tell us much.

This term is the **posterior probability** *of* C*. It represents the probability of a record belonging to class* C *after the data is taken into account.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

This term is the **posterior probability** *of* C. *It represents the probability of a record belonging to class* C *after the data is taken into account.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

The goal of any Bayesian computation is to find ("learn") the posterior distribution of a particular variable.

The idea of Bayesian inference, then, is to **update** *our beliefs about the distribution of* C *using the data ("evidence") at our disposal.*

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

Then we can use the posterior for prediction.

EXAMPLE: SPAM DETECTION

EXAMPLE: SPAM DETECTION

C: IS SPAM ? {1,0}
xi: how many times is word i present in email {0, Inf}

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

P({xi}|C) = count of emails with words frequencies {xi} in the SPAM subset
P(C) = ratio of SPAM emails
P({xi}) = normalization constant

Q: What piece of the puzzle we've seen so far looks like it could intractably difficult in practice?

Remember the likelihood function?

$$P(\{x_i\}|C) = P(\{x_1, x_2, \ldots, x_n\})|C)$$

Remember the likelihood function?

$$P(\{x_i\}|C) = P(\{x_1, x_2, \ldots, x_n\})|C)$$

Observing this exactly would require us to have enough data for every possible combination of features to make a reasonable estimate.

Q: What piece of the puzzle we've seen so far looks like it could intractably difficult in practice?

A: Estimating the full likelihood function.

Q: So what can we do about it?

Q: So what can we do about it?

A: Make a simplifying assumption. In particular, we assume that the features $\mathrm{x_i}$ *are conditionally independent from each other:*

Q: So what can we do about it?

A: Make a simplifying assumption. In particular, we assume that the features $\mathrm{x_i}$ *are conditionally independent from each other:*

$$\mathrm{P(\{x_i\}\,|\,C) \;=\; P(x_1, x_2, \ldots, x_n\,|\,C) \;\approx\; P(x_1\,|\,C) * P(x_2\,|\,C) * \ldots * P(x_n\,|\,C)}$$

Q: So what can we do about it?

A: Make a simplifying assumption. In particular, we assume that the features $\mathrm{x_i}$ *are conditionally independent from each other:*

$$\mathrm{P(\{x_i\}\,|\,C) \;=\; P(x_1, x_2, \ldots, x_n\,|\,C) \;\approx\; P(x_1\,|\,C) * P(x_2\,|\,C) * \ldots * P(x_n\,|\,C)}$$

This "naïve" assumption simplifies the likelihood function to make it tractable.

$$P(\text{class } C \mid \{x_i\}) = \frac{P(\{x_i\} \mid \text{class } C) \cdot P(\text{class } C)}{P(\{x_i\})}$$

the training phase *of the model involves computing the* likelihood function*, which is the conditional probability of each feature given each class.*

the prediction phase *of the model involves computing the* posterior probability *of each class given the observed features, and choosing the class with the highest probability.*

Advantages:

– Fast to train (single scan). Fast to classify

– Not sensitive to irrelevant features

– Handles real and discrete data

– Handles streaming data well

Disadvantages:

– Assumes independence of features

LAB
IV. NAIVE BAYESIAN CLASSIFICATION