Modeling with Naïve Bayes:

Mathematics, Example with School Grades via Hand, & Example with Titanic via Python

SAIL ON ● OCTOBER 2016

PAMELA TOMAN ● PTOMAN@CS.STANFORD.EDU

Where are we going tonight?

Educational agenda:

- We’re going to model tonight! …what is that again?
- So what is Kaggle?
- Let’s review probabilities…
- How does Naïve Bayes get us a model?
- Let’s code this thing!

Machine learning has three parts

We want our AI to be able to learn (robots? monitoring & labeling danger? personalized help?)

In machine learning, we provide (1) a mathematical equation with lots of unknown variables (a model) and (2) data, and we ask the computer to learn what the best values for the variables are

We implement the mathematical equation as an *algorithm* in some programming language

Machine learning

Mathematics (in slides)

How does my model work?
* You need to understand the math behind the most popular models.

Programming (in Titanic Notebook)

How do I write my process as code?
* You need to learn to program & how to use libraries like sklearn and pandas.

The Art of Data (a bit in Titanic Notebook)

What variables can I get or derive?
How good is current performance?
* You need to think critically and be aware of known gotchas.

What’s “a mathematical equation with lots of unknown variables (a model)”?

We choose/we are given (or both) some input variables *x, y, z* and some target variable *t.*

Our model is an equation for how the input variables *x, y, z* can be used to produce the target *t.*

What's "a mathematical equation with lots of unknown variables (a model)"?

We choose/we are given (or both) some input variables *x, y, z* and some target variable *t.*

Our model is an equation for how the input variables *x, y, z* can be used to produce the target *t.*

Maybe *x* is hat size, *y* is weight, and *z* is foot length.
We want to predict *t*, which is height.

We might think they are "linearly related":

$$t = ax + by + cz + d$$

This equation is entirely variables!

- Any individual has their own *x, y, z, t*, so these are variables.
- We don't know *a*, *b, c*, or *d*, so these are also variables. (But we can learn estimates for these!)

What's "a mathematical equation with lots of unknown variables (a model)"?

We choose/we are given (or both) some input variables *x, y, z* and some target variable *t.*

Our model is an equation for how the input variables *x, y, z* can be used to produce the target *t.*

Maybe *x* is hat size, *y* is weight, and *z* is foot length. We want to predict *t*, which is height.

We might think they are "linearly related":

$$t = ax + by + cz + d$$

This equation is entirely variables!

- Any individual has their own *x, y, z, t*, so these are variables.
- We don't know *a, b, c*, or *d*, so these are also variables. (But we can learn estimates for these!)

We can answer the question with other models too. Tonight we'll talk about Naïve Bayes, which uses probabilities. This time we let *t* be "height > 5.5ft":

$$P(t|x,y,z) = \frac{P(x|t)P(y|t)P(z|t)P(t)}{P(x,y,z)}$$

This equation is still entirely variables!

- Any individual has their own *x, y, z, t*, so these are variables.
- We don't know *P(x|t), P(y|t), P(z|t), P(t), or P(x,y,z)*, so these are also variables. (But we can learn estimates!)

What are we going to do tonight?

What are we going to do tonight?

Training Data: x, y, z
(each row contains attributes)

hat size, weight, foot length

What are we going to do tonight?

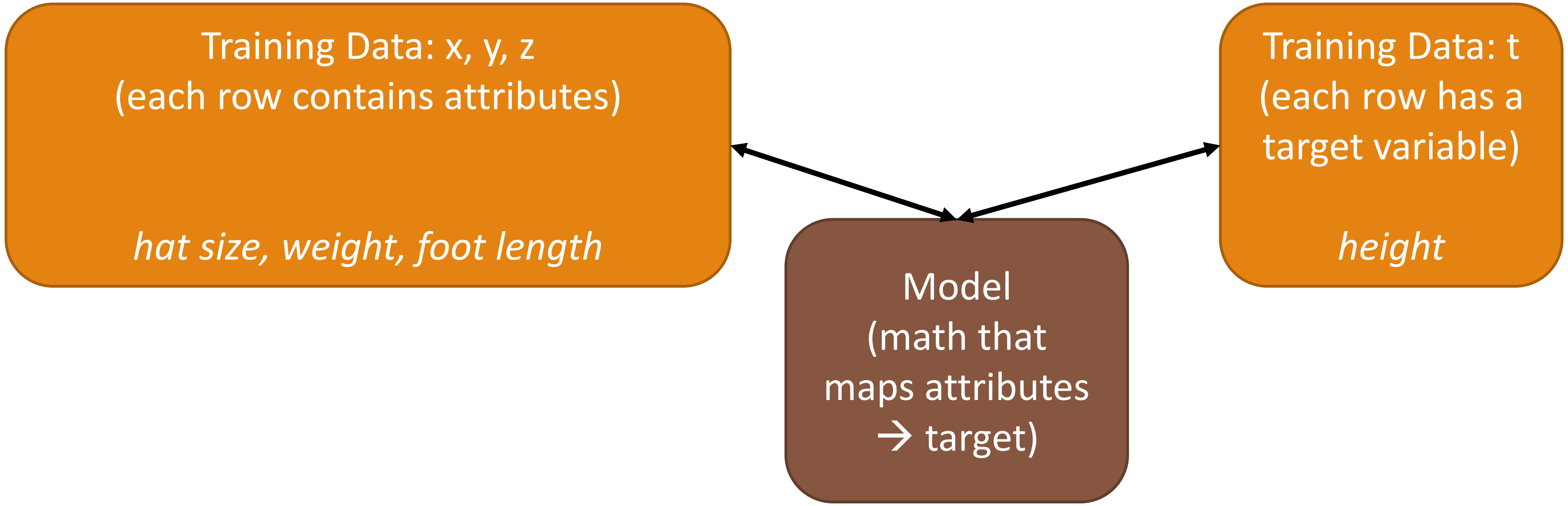
Training Data: x, y, z
(each row contains attributes)

hat size, weight, foot length

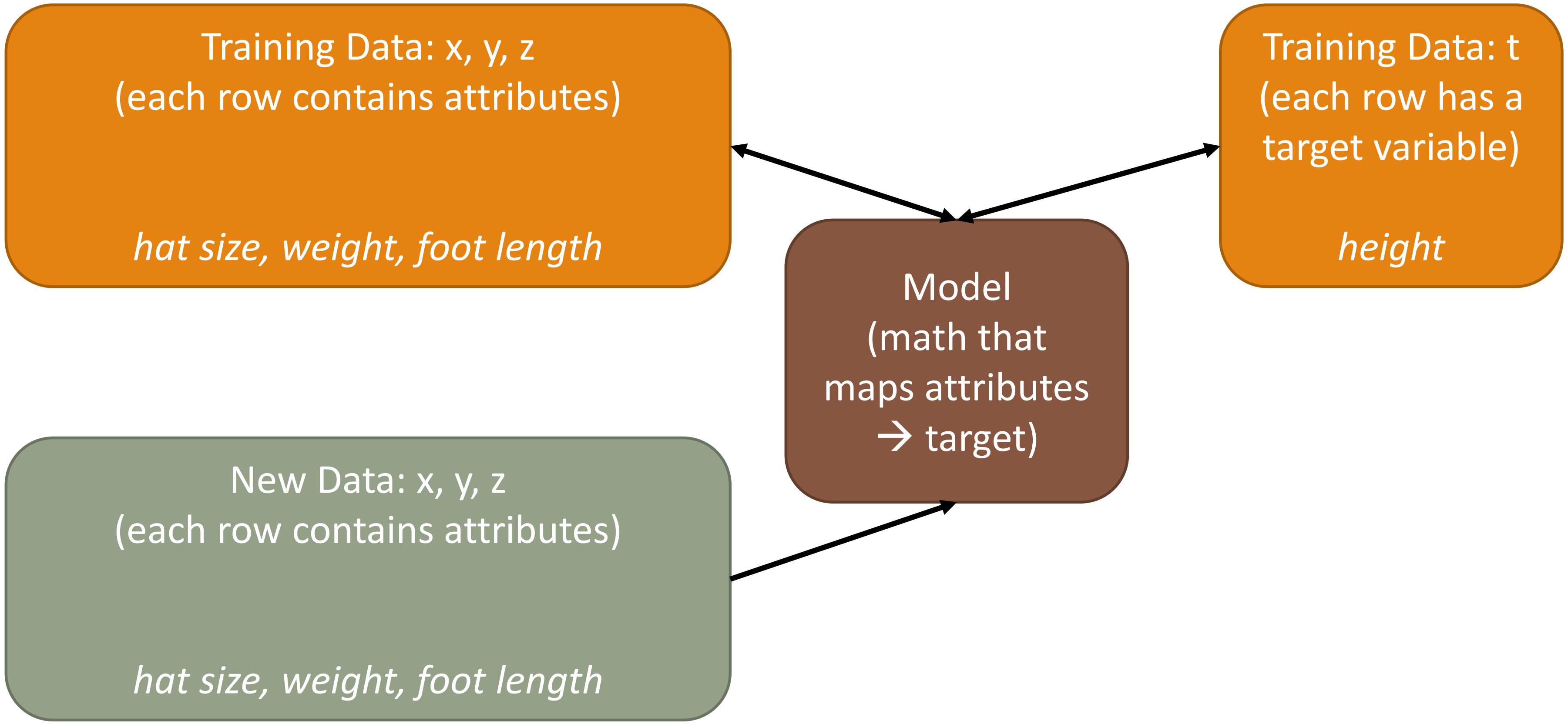
Training Data: t
(each row has a target variable)

height

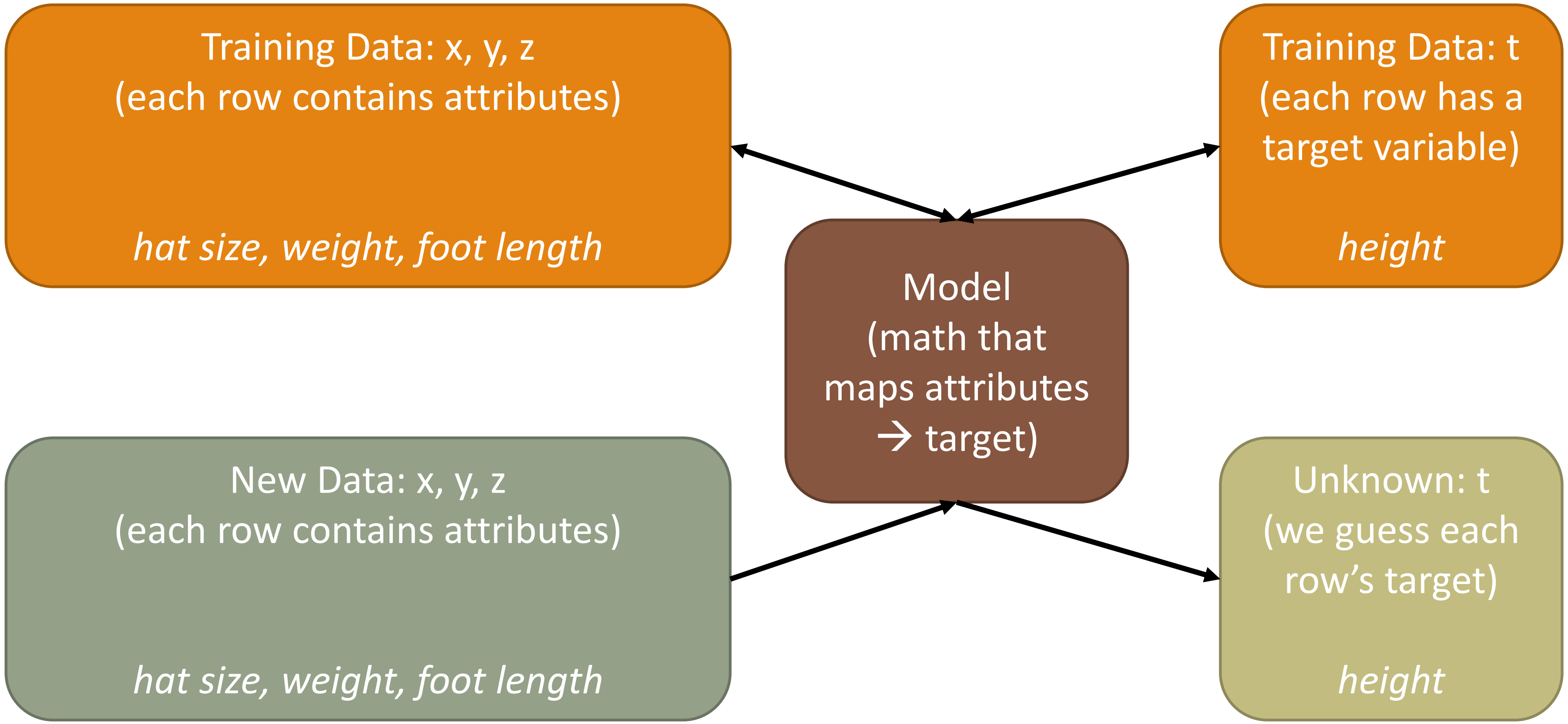
What are we going to do tonight?



What are we going to do tonight?



What are we going to do tonight?



What is Kaggle?

Kaggle is at https://www.kaggle.com/

Kaggle is a platform for “data science” – the application of machine learning to problems

- A company (usually) provides data & a target variable
- Participants build models to predict the target variable and they submit their predictions
- Participant solutions are ranked twice: (1) a public leaderboard visible to participants, (2) a private leaderboard visible only to the problem provider (… why twice?)

Participants get bragging rights & sometimes money or jobs

One of their “Getting Started” challenges is predicting who survives on the Titanic

We’re going to build a mathematical foundation for Naïve Bayes, and then work on the Titanic

Let's review probabilities…

Select 2 cards from a deck of 52 cards with replacement.
What's the probability of obtaining 2 kings?

Select 2 cards from a deck of 52 cards *without* replacement.
What's the probability of obtaining 2 kings?

Let’s review probabilities…

Select 2 cards from a deck of 52 cards with replacement.
What’s the probability of obtaining 2 kings?

“Independent” events
~0.0059

Select 2 cards from a deck of 52 cards *without* replacement.
What’s the probability of obtaining 2 kings?

Let’s review probabilities…

Select 2 cards from a deck of 52 cards with replacement.
What’s the probability of obtaining 2 kings?

“Independent” events
~0.0059

Select 2 cards from a deck of 52 cards *without* replacement.
What’s the probability of obtaining 2 kings?

“Dependent” events
~0.0045

Let’s review probabilities…

Select 2 cards from a deck of 52 cards with replacement.
What’s the probability of obtaining 2 kings?

“Independent” events
~0.0059

If two events *A* and *B* are **independent events**,

then the probability of event A *and* B is given by the following rule:

P(A, B) = P(A) * P(B)

We read P(X, Y) as “the probability of X *and* Y (both occurring)”.

Let’s review probabilities…

Select 2 cards from a deck of 52 cards *without* replacement.
What’s the probability of obtaining 2 kings?

“Dependent” events
~0.0045

If two events *A* and *B* are **dependent events**,

then the probability of event A and B is given by the following rule:

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

Here, P(X|Y) is a “conditional probability”: the probability that an event X will occur given that Y has already occurred.

We read P(X|Y) as “the probability of X, given Y”.

Let's tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

Let's tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

What's the probability that A = 2?

Let’s tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

What’s the probability that A = 2?

+		B					
		1	**2**	**3**	**4**	**5**	**6**
A	**1**	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Let's tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

What's the probability that A = 2?

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

P(A=2) = 6/36 = 0.167

Let's tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

What's the probability that A = 2?

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

P(A=2) = 6/36 = 0.167

It's revealed that A+B ≤ 5.

Let’s tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

What’s the probability that A = 2?

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

P(A=2) = 6/36 = 0.167

It’s revealed that A+B ≤ 5.

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Let's tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

What's the probability that A = 2?

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

P(A=2) = 6/36 = 0.167

It's revealed that A+B ≤ 5.

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Now what's the probability that A = 2?

Let's tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

What's the probability that A = 2?

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

P(A=2) = 6/36 = 0.167

It's revealed that A+B ≤ 5.

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Now what's the probability that A = 2?

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

source

P(A=2 | A+B ≤ 5) = 3/10 = 0.300

Let's tie P(X|Y) to a concrete example

We roll 2 fair dice, A and B. We write the possible A+B outcomes in a table.

What's the probability that A = 2?

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

P(A=2) = 6/36 = 0.167

It's revealed that A+B ≤ 5.

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Now what's the probability that A = 2?

+		B					
		1	2	3	4	5	6
A	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

source

P(A=2 | A+B ≤ 5) = 3/10 = 0.300

The probability that A=2 went up given the additional information!

Let's think more about conditional probabilities…

Here's the conditional probability rule again:

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

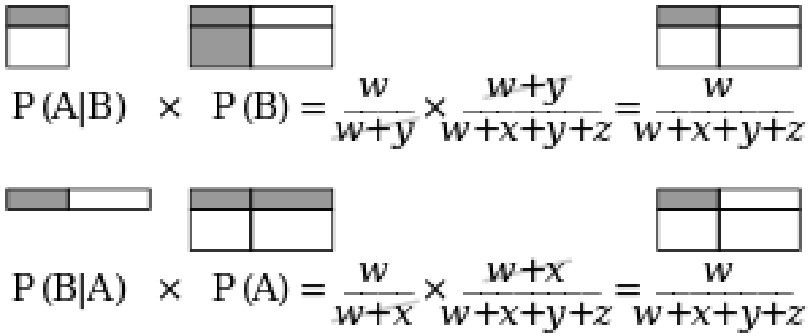
Let’s think more about conditional probabilities…

Here’s the conditional probability rule again:

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

Let’s look at why it works in a picture:
(Spend a bit of time reflecting on why this works.)

Relative size	Case B	Case $\bar{B}$	Total
Condition A	w	x	$w+x$
Condition $\bar{A}$	y	z	$y+z$
Total	$w+y$	$x+z$	$w+x+y+z$



source

Let’s think more about conditional probabilities…

Here’s the conditional probability rule again:

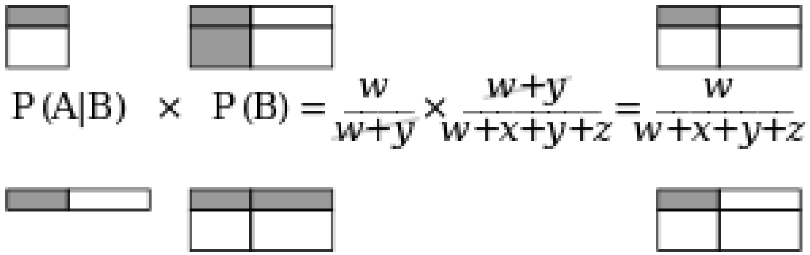
P(A, B) = P(A) * P(B|A) = P(B) * P(A|B)

Let’s look at why it works in a picture:
(Spend a bit of time reflecting on why this works.)

Let’s consider some more examples:

P(K1 and K2) = P(K1) * P(K2|K1) = P(K2) * P(K1|K2)

Relative size	Case B	Case $\bar{B}$	Total
Condition A	w	x	$w+x$
Condition $\bar{A}$	y	z	$y+z$
Total	$w+y$	$x+z$	$w+x+y+z$



$$P(A|B) \times P(B) = \frac{w}{w+y} \times \frac{w+y}{w+x+y+z} = \frac{w}{w+x+y+z}$$

$$P(B|A) \times P(A) = \frac{w}{w+x} \times \frac{w+x}{w+x+y+z} = \frac{w}{w+x+y+z}$$

source

Let’s think more about conditional probabilities…

Here’s the conditional probability rule again:

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

Relative size	Case B	Case $\bar{B}$	Total
Condition A	w	x	$w+x$
Condition $\bar{A}$	y	z	$y+z$
Total	$w+y$	$x+z$	$w+x+y+z$

Let’s look at why it works in a picture:
(Spend a bit of time reflecting on why this works.)

$$P(A|B) \times P(B) = \frac{w}{w+y} \times \frac{w+y}{w+x+y+z} = \frac{w}{w+x+y+z}$$

$$P(B|A) \times P(A) = \frac{w}{w+x} \times \frac{w+x}{w+x+y+z} = \frac{w}{w+x+y+z}$$

source

Let’s consider some more examples:

P(K1 and K2) = P(K1) * P(K2|K1) = P(K2) * P(K1|K2)

P(A and passed) = P(A) * P(passed|A) = P(passed) * P(A|passed)

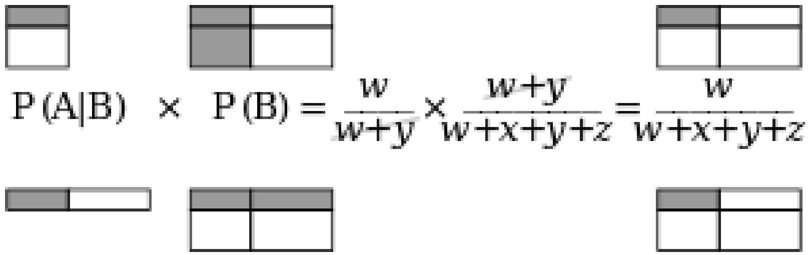
Let’s think more about conditional probabilities…

Here’s the conditional probability rule again:

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

Relative size	Case B	Case $\bar{B}$	Total
Condition A	w	x	$w+x$
Condition $\bar{A}$	y	z	$y+z$
Total	$w+y$	$x+z$	$w+x+y+z$

Let’s look at why it works in a picture:
(Spend a bit of time reflecting on why this works.)



$$P(A|B) \times P(B) = \frac{w}{w+y} \times \frac{w+y}{w+x+y+z} = \frac{w}{w+x+y+z}$$

$$P(B|A) \times P(A) = \frac{w}{w+x} \times \frac{w+x}{w+x+y+z} = \frac{w}{w+x+y+z}$$

source

Let’s consider some more examples:

P(K1 and K2) = P(K1) * P(K2|K1) = P(K2) * P(K1|K2)

P(A and passed) = P(A) * P(passed|A) = P(passed) * P(A|passed)

P(A+B ≤ 5 and A=2) = P(A+B ≤ 5) * P(A=2 | A+B ≤ 5) = P(A=2) * P(A+B ≤ 5 | A=2)

Bayes' Law (= Bayes' Theorem, Bayes' Rule)

Here's our conditional probability rule again:

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

Bayes' Law (= Bayes' Theorem, Bayes' Rule)

Here's our conditional probability rule again:

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

Let's rewrite it with P(B|A) on the left using algebra:

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

Bayes' Law (= Bayes' Theorem, Bayes' Rule)

Here's our conditional probability rule again:

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

Let's rewrite it with P(B|A) on the left using algebra:

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

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

Bayes' Law (= Bayes' Theorem, Bayes' Rule)

Here's our conditional probability rule again:

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

Let's rewrite it with P(B|A) on the left using algebra:

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

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

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

This equation is called "Bayes' Law"!

Let’s think more about Bayes’ Law…

Remember Bayes’ law:

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

Let’s think more about Bayes’ Law…

Remember Bayes’ law:

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

Let’s apply it.

Everyone who gets an A will pass. The average passing rate is 85%, and 20% of the class gets As. Of the people who pass, what percentage gets As?

Let’s think more about Bayes’ Law…

Remember Bayes’ law:

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

Let’s apply it.

Everyone who gets an A will pass. The average passing rate is 85%, and 20% of the class gets As. Of the people who pass, what percentage gets As?

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

Let’s think more about Bayes’ Law…

Remember Bayes’ law:

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

Let’s apply it.

Everyone who gets an A will pass. The average passing rate is 85%, and 20% of the class gets As. Of the people who pass, what percentage gets As?

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

$$P(A|\text{passed}) = \frac{1.00 * 0.20}{0.85}$$

Let’s think more about Bayes’ Law…

Remember Bayes’ law:

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

Let’s apply it.

Everyone who gets an A will pass. The average passing rate is 85%, and 20% of the class gets As. Of the people who pass, what percentage gets As?

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

$$P(A|passed) = \frac{1.00 * 0.20}{0.85}$$

$$P(A|passed) \approx 0.23$$

Let's think more about Bayes' Law…

Remember Bayes' law:

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

Let's apply it.

Everyone who gets an A will pass. The average passing rate is 85%, and 20% of the class gets As. Of the people who pass, what percentage gets As?

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

$$P(A|passed) = \frac{1.00 * 0.20}{0.85}$$

$$P(A|passed) \approx 0.23$$

We just used math to *infer* implicit & hidden information from given information.

Okay, back to modeling….

Okay, back to modeling….

Let’s say we want to infer whether *you* got an A.

Okay, back to modeling....

Let's say we want to infer whether *you* got an A.

You are very shy and say nothing, but some classmates are more forthcoming.

Okay, back to modeling....

Let’s say we want to infer whether *you* got an A.

You are very shy and say nothing, but some classmates are more forthcoming.

We do know a few things about you:

- Did you pass the class?
- Did you look happy when grades came out last week?
- What’s your GPA (or our guess at it)?
- What’s the average grade of your three best friends in the class?

Okay, back to modeling....

Let's say we want to infer whether *you* got an A.

You are very shy and say nothing, but some classmates are more forthcoming.

We do know a few things about you:

- Did you pass the class?
- Did you look happy when grades came out last week?
- What's your GPA (or our guess at it)?
- What's the average grade of your three best friends in the class?

We create a table for our information:

Okay, back to modeling….

Let's say we want to infer whether *you* got an A.

You are very shy and say nothing, but some classmates are more forthcoming.

We do know a few things about you:

- Did you pass the class?
- Did you look happy when grades came out last week?
- What's your GPA (or our guess at it)?
- What's the average grade of your three best friends in the class?

We create a table for our information:

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
99 (you!)	Yes	Yes	3.9	B	????

… + more rows!

If we reframe as a conditional probability problem, we'll need lots of data

If we reframe as a conditional probability problem, we'll need lots of data

We want to know:

If we reframe as a conditional probability problem, we'll need lots of data

We want to know:

P(A | Yes, Yes, 3.9, B)

If we reframe as a conditional probability problem, we'll need lots of data

We want to know:

P(A | Yes, Yes, 3.9, B)

Bayes' rule means we could plug values into this equation & get the answer:

If we reframe as a conditional probability problem, we'll need lots of data

We want to know:

$$P(A \mid \text{Yes}, \text{Yes}, 3.9, B)$$

Bayes' rule means we could plug values into this equation & get the answer:

$$P(A \mid \text{Yes}, \text{Yes}, 3.9, B) = \frac{P(\text{Yes}, \text{Yes}, 3.9, B \mid A) * P(A)}{P(\text{Yes}, \text{Yes}, 3.9, B)}$$

If we reframe as a conditional probability problem, we'll need lots of data

We want to know:

P(A | Yes, Yes, 3.9, B)

Bayes' rule means we could plug values into this equation & get the answer:

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

But this is *hard*!! We don't have nearly enough data.

If we reframe as a conditional probability problem, we'll need lots of data

We want to know:

$$P(A \mid Yes, Yes, 3.9, B)$$

Bayes' rule means we could plug values into this equation & get the answer:

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

But this is *hard*!! We don't have nearly enough data.

→ In particular:

Looking only at the people in the class who got As, how many people match your answers?

Probably 0. But that means we have small data, not that you didn't get an A…

Our solution out of the small data problem: a "naïve" independence assumption

Our solution out of the small data problem: a “naïve” independence assumption

We can resolve the small data problem through making a “naïve” assumption:

Our solution out of the small data problem: a “naïve” independence assumption

We can resolve the small data problem through making a “naïve” assumption:

Within each target class (A or not-A), the probability of each variable is *independent of others*:

- Probability of passing is independent from happiness, GPA, and friends, among those who (didn’t) get an A
- Happiness is independent from passing, GPA, and friends, among those who (didn’t) get an A
- …

So instead of:

Our solution out of the small data problem: a “naïve” independence assumption

We can resolve the small data problem through making a “naïve” assumption:

Within each target class (A or not-A), the probability of each variable is *independent of others*:

- Probability of passing is independent from happiness, GPA, and friends, among those who (didn’t) get an A
- Happiness is independent from passing, GPA, and friends, among those who (didn’t) get an A
- …

So instead of:

$$P(A \mid Yes, Yes, A, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

We will make a “Naïve Bayes” assumption and instead use:

Our solution out of the small data problem: a "naïve" independence assumption

We can resolve the small data problem through making a "naïve" assumption:

Within each target class (A or not-A), the probability of each variable is *independent of others*:

- Probability of passing is independent from happiness, GPA, and friends, among those who (didn't) get an A
- Happiness is independent from passing, GPA, and friends, among those who (didn't) get an A
- …

So instead of:

$$P(A \mid Yes, Yes, A, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

We will make a "Naïve Bayes" assumption and instead use:

$$P(A \mid Yes, Yes, A, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Our solution out of the small data problem: a "naïve" independence assumption

We can resolve the small data problem through making a "naïve" assumption:

Within each target class (A or not-A), the probability of each variable is *independent of others*:

- Probability of passing is independent from happiness, GPA, and friends, among those who (didn't) get an A
- Happiness is independent from passing, GPA, and friends, among those who (didn't) get an A
- …

So instead of:

$$P(A \mid Yes, Yes, A, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

We will make a "Naïve Bayes" assumption and instead use

$$P(A \mid Yes, Yes, A, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

If we just want *whether* you got an A, we can simplify even more....

Let's keep simplifying. So currently our formula will tell us the *probability* you got an A.

$$P(A \mid Yes, Yes, A, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

If we just want to know *whether* you got an A, we can simplify even more.

If we just want *whether* you got an A, we can simplify even more….

Let’s keep simplifying. So currently our formula will tell us the *probability* you got an A.

$$P(A \mid Yes, Yes, A, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

If we just want to know *whether* you got an A, we can simplify even more.

Whether you got an A is just:

$$P(A \mid whatever) \quad >? \quad P(not\ A \mid whatever)$$

$$\frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)} \quad >? \quad \frac{P(Yes \mid not\ A) * P(Yes \mid not\ A) * P(3.9 \mid not\ A) * P(B \mid not\ A) * P(not\ A)}{P(Yes, Yes, 3.9, B)}$$

Since the denominator is the same, we can drop the denominator….
The key here is that whether you got an A or you didn’t, we still have the same observations.

So finally we have a tractable (usable) formula for whether you got an A…

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)

Rearranging the equations into a format known as “Bayes’ law”:

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)

Rearranging the equations into a format known as "Bayes' law":

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

$$P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)$$

Rearranging the equations into a format known as "Bayes' law":

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Deciding we can deal with small data through the "Naïve Bayes" assumption of independence:

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

$$P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)$$

Rearranging the equations into a format known as "Bayes' law":

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Deciding we can deal with small data through the "Naïve Bayes" assumption of independence:

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)

Rearranging the equations into a format known as "Bayes' law":

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Deciding we can deal with small data through the "Naïve Bayes" assumption of independence:

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Identifying we just want *whether*, not a probability:

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

$$P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)$$

Rearranging the equations into a format known as "Bayes' law":

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Deciding we can deal with small data through the "Naïve Bayes" assumption of independence:

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Identifying we just want *whether*, not a probability:

$$\frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)} >? \frac{P(Yes \mid not A) * P(Yes \mid not A) * P(3.9 \mid not A) * P(B \mid not A) * P(not A)}{P(Yes, Yes, 3.9, B)}$$

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

$$P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)$$

Rearranging the equations into a format known as "Bayes' law":

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Deciding we can deal with small data through the "Naïve Bayes" assumption of independence:

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Identifying we just want *whether*, not a probability:

$$\frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)} >? \frac{P(Yes \mid not\ A) * P(Yes \mid not\ A) * P(3.9 \mid not\ A) * P(B \mid not\ A) * P(not\ A)}{P(Yes, Yes, 3.9, B)}$$

Recognizing the denominator doesn't matter:

So finally we have a tractable (usable) formula for whether you got an A…

Starting with the definition of conditional probability:

$$P(A, Yes, Yes, 3.9, B) = P(A) * P(Yes, Yes, 3.9, B|A) = P(Yes, Yes, 3.9, B) * P(A|Yes, Yes, 3.9, B)$$

Rearranging the equations into a format known as "Bayes' law":

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes, Yes, 3.9, B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Deciding we can deal with small data through the "Naïve Bayes" assumption of independence:

$$P(A \mid Yes, Yes, 3.9, B) = \frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)}$$

Identifying we just want *whether*, not a probability:

$$\frac{P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A)}{P(Yes, Yes, 3.9, B)} >? \frac{P(Yes \mid not\ A) * P(Yes \mid not\ A) * P(3.9 \mid not\ A) * P(B \mid not\ A) * P(not\ A)}{P(Yes, Yes, 3.9, B)}$$

Recognizing the denominator doesn't matter:

$$P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A) >? P(Yes \mid not\ A) * P(Yes \mid not\ A) * P(3.9 \mid not\ A) * P(B \mid not\ A) * P(not\ A)$$

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A		
Not A		

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A					
Not A					

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	
Not A		

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A					
Not A					

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A)   >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A		

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A					
Not A					

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A	2	

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A					
Not A					

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A	2	2

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A					
Not A					

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A	2	2

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A	1				
Not A					

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A	2	2

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A	1	0	0	0	0
Not A					

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A	2	2

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A	1	0	0	0	0
Not A	0				

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A	2	2

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A	1	0	0	0	0
Not A	0	3			

We can solve the problem now!

So here is our formula:

$$P(Yes \mid A) * P(Yes \mid A) * P(3.9 \mid A) * P(B \mid A) * P(A) \quad >? \quad P(Yes \mid not\ A) * P(Yes \mid not\ A) * P(3.9 \mid not\ A) * P(B \mid not\ A) * P(not\ A)$$

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A	2	2

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A	1	0	0	0	0
Not A	0	3	1		

We can solve the problem now!

So here is our formula:

P(Yes| A) * P(Yes | A) * P(3.9 | A) * P(B | A) * P(A) >?
P(Yes| not A) * P(Yes | not A) * P(3.9 | not A) * P(B | not A) * P(not A)

Using the data we collected from your forthcoming classmates, we can estimate *all* these probs.

Person	Passed?	Happy?	GPA?	Friends'?	Grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	No	No	3.3	A	F
5	Yes	Yes	3.2	B	B
…	…	…	…	…	…
99 (you!)	Yes	Yes	3.9	B	????

	Passed = Yes	Passed = No
A	1	0
Not A	2	2

	GPA > 3.5	GPA 3.0-3.5	GPA 2.5-3.0	GPA 1.5-2.5	GPA <1.5
A	1	0	0	0	0
Not A	0	3	1	0	0

Your turn!

That's a lot of potentially new math.

Check your understanding at the lowest level by filling out the worksheet by hand.

If that's straightforward, try the deeper thought questions….

Naïve Bayes Modeling Example:

Estimating *Your* Hidden Grade From More Forthcoming Classmates' Statements

You're shy and don't want to say whether you got an A. But many of your classmates are pretty forthcoming. Can we figure your grade out from what they say about themselves plus a handful of facts about you – with mathematical rigor?

Person	Passed the class?	Looked happy seeing grades?	GPA estimate?	Average grade of friends?	True grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	Yes	No	3.3	A	D
5	Yes	Yes	3.6	B	B
6	Yes	No	4.0	A	A
7	Yes	Yes	3.7	A	A
8	Yes	Yes	3.2	A	B
9	Yes	Yes	3.1	B	B
10	No	No	1.8	D	F
11	Yes	No	3.3	B	A
12	Yes	No	3.6	A	B
13	Yes	Yes	2.7	B	C
14	No	No	3.1	C	F
15	Yes	Yes	1.9	C	D
99 (YOU)	Yes	Yes	3.9	B	????

To get a mathematically justifiable prediction of your grade, we tally up the counts:

Passed	*Yes*	*No*
A	4	0
Not A	8	3

Happy	*Yes*	*No*
A		
Not A		

GPA	*>3.5*	*3.0-3.5*	*2.5-3.0*	*1.5-2.5*	*<1.5*
A					
Not A					

Friends'	*A*	*B*	*C*	*D*	*F*
A					
Not A					

Then we calculate two Naïve Bayes scores for person #99 (you!), whose grade we want to predict:

*Passed * Happy * GPA * Friends' * P(A)*

Getting an A: P(Yes| A) * P(Yes | A) * P(>3.5 | A) * P(B | A) * P(A)

4/4 = 1.0 * ________ * ________ * ________ * _____ = _________

Not getting an A:

____ = _________

Given the observed data and our Naïve Bayes model, which target class (A or not A) is most likely? __________

Pamela Toman (ptoman@cs.stanford.edu)
SAIL ON, October 2016

Deeper thought questions for you…

source

Why did we "discretize" (put into buckets) the GPA variable? How do we avoid counting a border case like 3.5 in more than one bucket?

If we need to bucket variables (like GPA), what are some good ways of choosing the bucket size?

Does the Naïve Bayes method get better with a bigger sample size (more data)? Why?

What do we need to estimate the actual probability that you got an A?

Can we use this framework to figure out your most likely grade (A-F)? How?

What happens if we want to make a prediction about someone who has a 1.4 GPA but we don't have anyone in the dataset with that characteristic? How could we fix that problem?

Suppose the teacher tells us P(A) is really 45% even though our data estimate says 27% (our sample was skewed). Which value should we use in our equation for P(A)? Why?

Your turn! (answers)

Our prediction, based on the 15 rows of data, is that you **don’t** get an A. *sad trombones*

The score for getting an A: 0.0127

The score for not getting an A: 0.0186

Naïve Bayes Modeling Example:
Estimating *Your* Hidden Grade From More Forthcoming Classmates’ Statements

You’re shy and don’t want to say whether you got an A. But many of your classmates are pretty forthcoming. Can we figure your grade out from what they say about themselves plus a handful of facts about you – with mathematical rigor?

Person	Passed the class?	Looked happy seeing grades?	GPA estimate?	Average grade of friends?	True grade
1	Yes	No	3.8	A	A
2	No	No	2.1	D	F
3	Yes	Yes	3.1	B	B
4	Yes	No	3.3	A	D
5	Yes	Yes	3.6	B	B
6	Yes	No	4.0	A	A
7	Yes	Yes	3.7	A	A
8	Yes	Yes	3.2	A	B
9	Yes	Yes	3.1	B	B
10	No	No	1.8	D	F
11	Yes	No	3.3	B	A
12	Yes	No	3.6	A	B
13	Yes	Yes	2.7	B	C
14	No	No	3.1	C	F
15	Yes	Yes	1.9	C	D
99 (YOU)	Yes	Yes	3.9	B	????

To get a mathematically justifiable prediction of your grade, we tally up the counts:

Passed	*Yes*	*No*
A	4	0
Not A	8	3

Happy	*Yes*	*No*
A		
Not A		

GPA	*>3.5*	*3.0-3.5*	*2.5-3.0*	*1.5-2.5*	*<1.5*
A					
Not A					

Friends’	*A*	*B*	*C*	*D*	*F*
A					
Not A					

Then we calculate two Naïve Bayes scores for person #99 (you!), whose grade we want to predict:

*Passed * Happy * GPA * Friends’ * P(A)*

Getting an A: P(Yes| A) * P(Yes | A) * P(>3.5 | A) * P(B | A) * P(A)

4/4 = 1.0 * ________ * ________ * ________ * _____ = _________

Not getting an A:

____ = _________

Given the observed data and our Naïve Bayes model, which target class (A or not A) is most likely? __________

Pamela Toman (ptoman@cs.stanford.edu)
SAIL ON, October 2016

Naïve Bayes is a machine learning model

Naïve Bayes is a machine learning model

All models connect input variables to a target output variable through math/algorithms

- The Naïve Bayes model goes through each target class individually, and then it estimates the probability that the data we observed came from that target class. It predicts the most likely target class.
- Other models follow other assumptions & use other approaches

Naïve Bayes is a machine learning model

All models connect input variables to a target output variable through math/algorithms

- The Naïve Bayes model goes through each target class individually, and then it estimates the probability that the data we observed came from that target class. It predicts the most likely target class.
- Other models follow other assumptions & use other approaches

The best & most commonly used models are *mathematically justifiable*

- We like rigor
- We like having a guarantee that the solution is correct (or optimal) under some assumptions
- No one takes you seriously if you don't have a defensible reason to do what you did

Naïve Bayes is a machine learning model

All models connect input variables to a target output variable through math/algorithms

- The Naïve Bayes model goes through each target class individually, and then it estimates the probability that the data we observed came from that target class. It predicts the most likely target class.
- Other models follow other assumptions & use other approaches

The best & most commonly used models are *mathematically justifiable*

- We like rigor
- We like having a guarantee that the solution is correct (or optimal) under some assumptions
- No one takes you seriously if you don't have a defensible reason to do what you did

What makes this *machine learning*?

We defined a **general** approach for using conditional probabilities to get the most likely target class. We only used & programmed generic logic. The same approach works anywhere – with predicting cancer, with predicting whether an email is something you'll care about, etc. …!

Let’s turn to the Titanic

The Titanic data is much more than 15 rows (yay!)

But hand counting was painful on only 15 rows.

Now that we thoroughly understand the math & algorithm… Let’s use a computer!

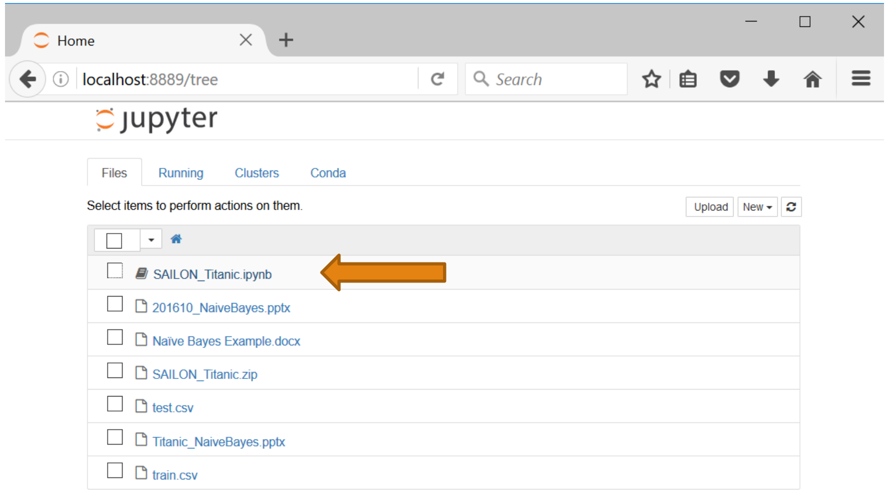
Opening up the Jupyter Notebook

Quickstart from scratch:

1. Install the Anaconda distribution of Python (other distributions won't necessarily come with all that we need)
2. Download the .ipynb file to /some/path/for/my/ipynb
   - Jupyter notebooks (*.ipynb) are for "literate programming" – intermingling code and detailed text discussions (even pictures!)
3. From Kaggle's website, download the Titanic *train.csv* and *test.csv* files to /some/path/for/my/ipynb
4. Start Jupyter Notebook in a parent folder to where the .ipynb file is
   - Navigate to /some/path/for/my/ipnb in the command line & run `jupyter notebook` in the command line

   -OR-

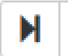
   - Open up Jupyter Notebook & move the .ipynb & .csv files so they show up in the file list
5. Click on the *SAILON_Titanic.ipynb* file within your web browser

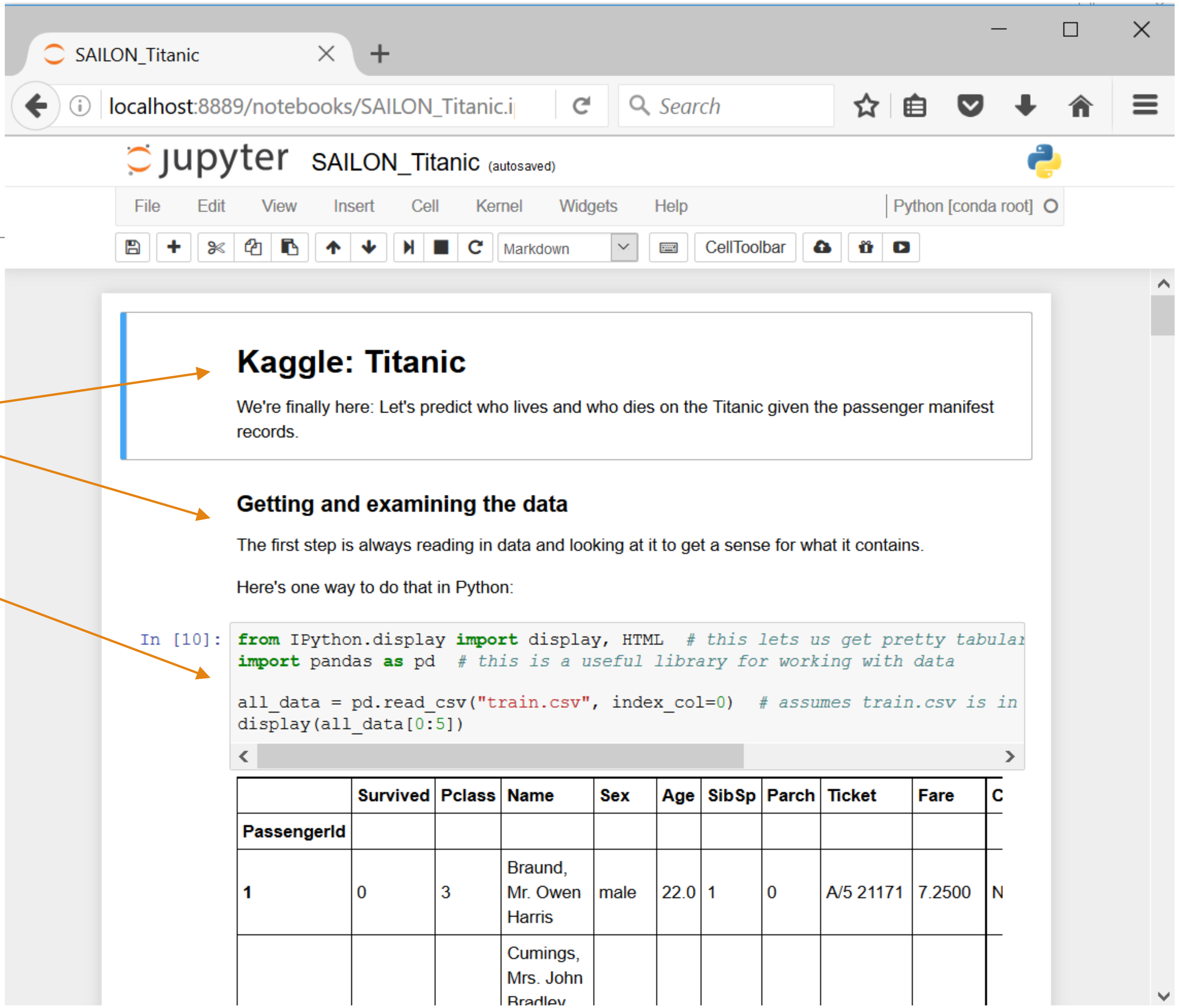


Jupyter Notebook

Text cells

Code cell

- Execute it with:
  - SHIFT+Enter
  - 
  - Cell→Run Cells
- Edit it and rerun!



For even more info, there is a Youtube [tutorial on Jupyter Notebooks](#).

When coding, we almost always use libraries for the math (which we need to understand to not make silly mistakes & get stuck). Starting to code raises new challenges: thinking on & usefully preparing lots of data.

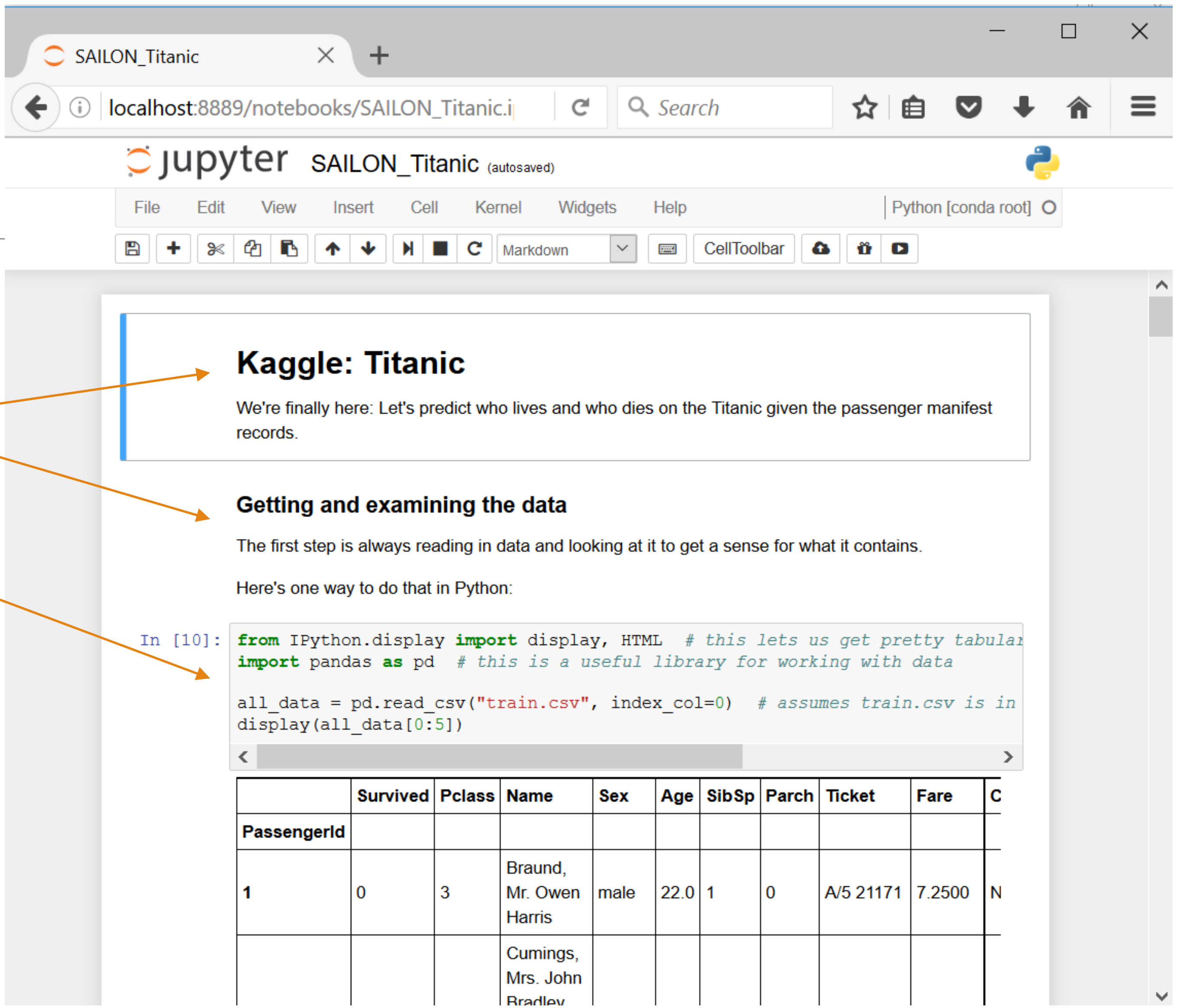
Jup er ote book

lls

cell

with:

- S -Enter
-
- Cell→Run Cells
- Edit it and rerun!



For even more info, there is a Youtube [tutorial on Jupyter Notebooks](#).

To get an even better handle on this…

1) Try to answer the thought questions posed earlier

2) Tinker with & fully understand the Jupyter Notebook we worked on tonight

3) Build an amazing Naïve Bayes model for the Titanic

- Create a Kaggle account (https://www.kaggle.com)
- Review the Titanic data and tutorials (look for DIY Tutorials and Kaggle Kernels – look especially for ones that name their method)
- Build a Naïve Bayes model that predicts who will survive (starting from our notebook or elsewhere)
- Explore excluding, including, and deriving new columns to try to improve performance
- Keep your best models & predictions – we'll share insights later on!

4) Write your own implementation of Naïve Bayes in Python (from scratch!) and/or write out the derivation of the Naïve Bayes model (without peaking!) – this is the way to verify you have a deep & thorough understanding