

# Assignment 1: Bayes' Rule

## Machine Learning

Fall 2019

### 🔗 Learning Objectives

- Gain familiarity with key ideas in ML (with a focus on probabilistic methods).
- Review of mathematical concepts we will be using in the beginning part of this course.
- Familiarize yourself with computational tools for machine learning.
- Learn linear regression using a “top-down” approach.

## 1 Motivation

We've circled around the idea of uncertainty a lot in this class. For instance, when we studied logistic regression, we learned that logistic regression outputs a probability that the class output is 1, which lets us capture the model's certainty in its prediction.

TODO: fix the rest of this.

We've also thought about models that don't fit the data perfectly (e.g., a line of best fit will usually not perfectly fit out data). One way we can think of this lack of fit is due to inherent uncertainty in the thing we are trying to predict. In this next part of the class we're going to formalize our notion of uncertainty using various concepts from probability theory. The major outcomes of this move will be to:

- make more explicit our assumptions about where data comes from and how uncertainty impacts that data.
- allow us to quantify our confidence in our model - instead of giving back the best fitting model we will actually be able to say how confident we are that this model is the right model.
- give us a way to incorporate prior knowledge into our machine learning models, giving us more interpretable models.
- give us powerful tools to reason about fairness and bias in machine learning models.

## 2 Six Big Ideas in (Probabilistic) Machine Learning

We're going to reprise the Six Big ideas in ML from our very first assignment with a focus on connections to probabilistic machine learning. Our intent is to provide a larger framework for you to interpret the content you are learning. Could potentially point to [this article in Nature Reviews](#).

The [pymc examples](#) are also super cool (e.g., [GP Mauna Loa](#))

*Idea 1: Probabilistic Methods Help Us Understand Learning in Artificial Systems and Biological Systems*

*Idea 2: Probabilistic Methods Can Illuminate Hidden Structures*

- Phylogenetic trees
- [Indus Script](#)

*Idea 3: Probabilistic Methods Give Us a Language for Reasoning about Algorithmic Fairness*

*Idea 4: Probabilistic Methods Allow Us to Attach Confidences to Our Predictions*

[Mauna Loa CO2 example](#)

*Idea 5: Probabilistic Methods Allow for Learning More Interpretable Models*

<https://towardsdatascience.com/what-is-bayesian-statistics-used-for-37b91c2c257c>

*Idea 6: Probabilistic Methods Let you Learn Hierarchies of Models*

TODO (clustering or topic modeling?)

*notes: other ideas*

- Decision-making under uncertainty
- [Probabilistic Programming?](#) (Note: there is an example of inverse computer graphics that is cool)
- Compression?

### Exercise 1 🧠 (60 minutes)

Now, we want to hear from you! Choose one of the big ideas above and write a short response to it. Your response could incorporate something surprising you read, a thought-provoking question, your personal experience, an additional resource that builds upon or shifts the discussion. We hope that this reflection will help scaffold class discussions and get you thinking about your interests in the big space that is ML. Also, you have license from us to customize the structure of your response as you see fit. As a rough guide, you should aim for a response of a 1-2 paragraphs.

## 3 Probability

First we'll build some intuition around probability, and then we'll define it formally.

### 3.1 Intuition

Most of us are used to thinking that events can be probabilistic, that is we can attach some probability to whether or not they occur. Take for example flipping a coin. We could think of the event that the coin comes up heads as having probability 0.5. That is, there is an even chance that it happens versus doesn't happen. Further, we can talk about events as being observable if we are able to ultimately know whether or not they occurred. For instance, whether a coin comes up heads is an observable event since you can ultimately observe the outcome of the flip. In contrast, some events would be considered unobservable if they are unable to be directly ascertained by human senses. A classic example of this would be whether a scientific theory is true or not. It is impossible to directly observe whether the theory is true, but you might be able to observe events that are consistent or inconsistent with the theory.

## Exercise 2

Come up with some examples of observable events that are probabilistic in nature. For each event, either provide the probability or explain what factors would determine the probability.

### 3.2 Formal Definition

TODO link to external resource.

#### External Resource(s)

- Goals for external resources:
- Define a probability space (sample space, event space, pmf)
- Give some basic properties
- Complement
- Probability of a union of disjoint events

## Exercise 3

Here are some diagnostic questions to make sure that you go the basic ideas.

- What are the components needed to define a probability space?
- If  $\mathcal{E}$  refers to a particular event that occurs with probability  $p(\mathcal{E})$  what is  $p(\overline{\mathcal{E}})$  (the line over  $\mathcal{E}$  means the complement, or the event that  $\mathcal{E}$  does not occur).
- For any event  $\mathcal{E}$ , what must be true of  $p(\mathcal{E})$ ?
- if  $\mathcal{A}$  and  $\mathcal{B}$  are disjoint events (they cannot occur simultaneously, what is the probability of at least one of  $\mathcal{A}$  or  $\mathcal{B}$  occurring in terms of  $p(\mathcal{A})$  and  $p(\mathcal{B})$ .

### 4 From Basic Definitions to Bayes' Rule

#### External Resource(s)

Allen Downey's (ever heard of him?) book Think Bayes is an excellent introduction to Bayesian analysis. The [first chapter](#) (which you should read) starts with a less formal definition of probability than we gave earlier. The chapter then gives intuitions around conjoint probability (the probability that multiple events occur simultaneously), conditional probability (the probability that some event occurs conditioned on another event having occurred), and finally to Bayes' rule (a surprisingly easy theorem to derive that allows you to write one conditional probability distribution in terms of another).

Allen's treatment of the material is, of course, not the only one out there (we like it for its focus on building intuition and focusing on the key ideas). Here are some other resources you might consider checking out.

- [Khan Academy Video on Bayes' Theorem](#): this shows some simple applications of Bayes' rule and explains why it is a convenient way to reason about the probability of hypothesis given data).
- [Veritasium Episode on Bayes' Theorem](#): this has a bit more history and philosophy of Bayes' Theorem along with some nice visualizations. It also includes the presenter randomly walking on a very scenic mountain, so there's that if nothing else.
- I (Paul) ran across [this example of applying Bayes' rule to a real world problem](#) that was created by a grad school friend of mine. It's hilarious (lots of Cat Memes).

#### Exercise 4

Here are some exercises to test your understanding of this material.

(a) TODO

### 5 *Marginalization of Probability Distributions*

TODO: This was used in some of the previous readings, make it more formal. Show tree creation as a way to do visualize this.

#### Exercise 5

TODO

### 6 *Applying Bayes' Rule*

Have a few options of where Bayes could be applied. Have them pick one of them and figure out an estimate.

### 7 *Companion Notebook*