FLATIRON SCHOOL

Mapping Distributions with Bayesian Algorithms.

DS

hosted by KASH

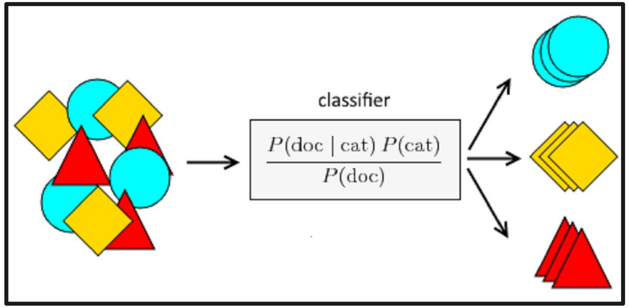
CRITICAL QUESTION

How can we construct and apply algorithms that leverage Bayesian statistics and conditional probability distributions in order to learn patterns in data?

CRITICAL QUESTION

How can we utilize Gaussian Naïve Bayes and Multinomial Naïve Bayes algorithms to classify new data?

BAYESIAN MODELS AS PREDICTIVE CLASSIFIERS



Before diving deep into **Bayesian statistics**, what's most important to know is that Bayesian classifiers are a powerful predictive algorithm due to their ability to analyze **how features in a dataset *influence the outcome* of a target class**.

While the precise mathematical process may be arcane at the moment, the technical evocation of this is quite straightforward and **similar to other utilizations of machine learning classifiers in Python**.

//

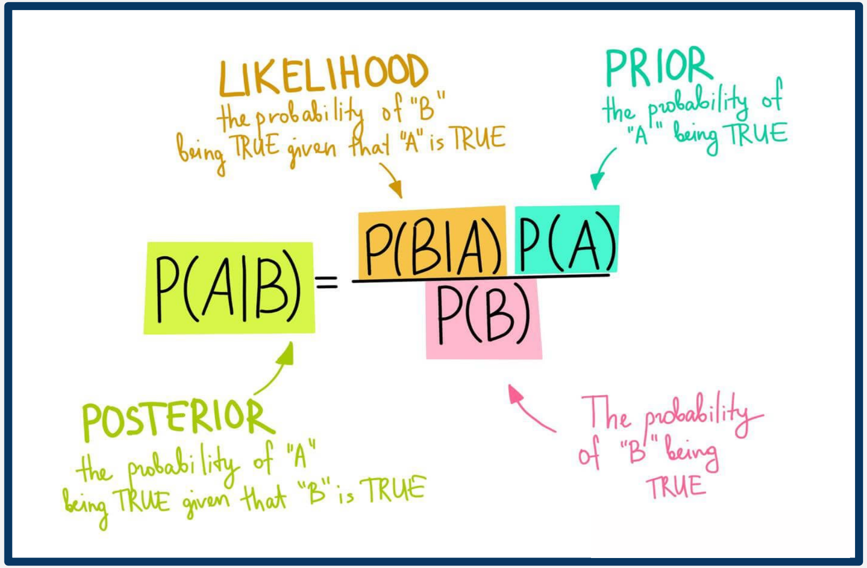
WRITTEN IDEATION

HOW DOES **PRIOR KNOWLEDGE** ABOUT A DATASET **IMPACT PREDICTIONS** WE MAKE ACROSS THE DATA?

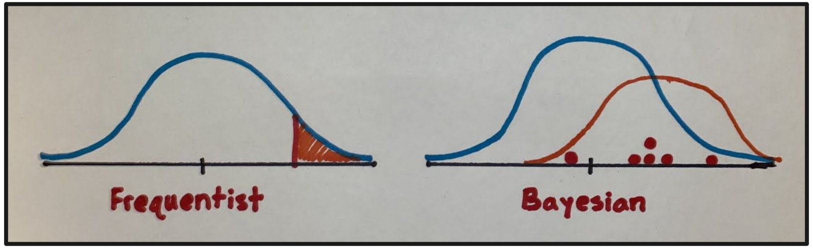
CODING WALKTHROUGH

Applying Bayesian Classifiers for Basic Text Classification

AN ILLUSTRATIVE INTRODUCTION TO THE BAYES THEOREM



FROM FREQUENTISTS TO BAYESIANS



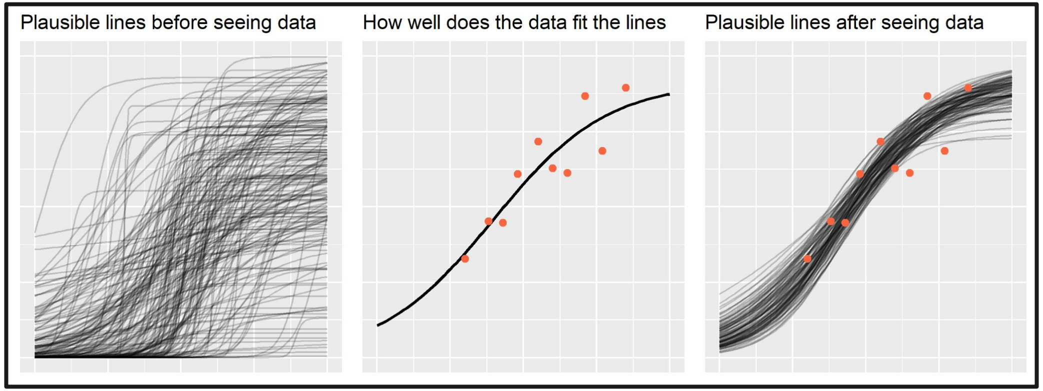
So far throughout our journey in data science, we've examined data and its classes and heuristics through a **frequentist lens**: by assuming non-randomness in our parameter and viewing events as samples rather than probability distributions.

However, many opt to use a **Bayesian lens** for predictive modeling in order to understand the relationship between classes and data more intimately.

A CONCEPTUAL INTRODUCTION TO THE BAYES THEOREM

The **Bayes Theorem** is at the heart of Bayesian modeling: it is a mathematical formula that asserts that events across data maintain **conditional probabilities**, in that the occurrences of some events and data influence the occurrences of other events and data.

Through this perspective on conditional probability, Bayesian inference can be applied to data science to understand **how some features can affect the outcome of a target class**.



PRIORS, POSTERIORS, AND CONDITIONAL PROBABILITIES

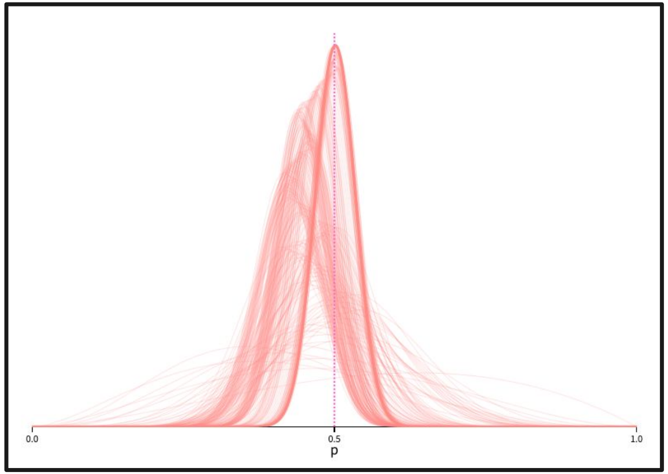
When applying Bayesian algorithms, it's important to understand how data is updated – in these cases, we always start with some **"prior belief(s)"** on the heuristics we're evaluating.

By analyzing *newly **observed data** from a conditional probability perspective*, we can convert our priors into **"posterior beliefs"** that are marginally more expressive of the true relationship that we're attempting to capture.

TURNING PRIORS TO POSTERIORS WITH M.A.P. OPTIMIZATION

The process for updating our prior beliefs to posterior beliefs (which, by the way, are both represented as probability distributions) is through a tuning step called **"maximum a posteriori" (MAP)** optimization.

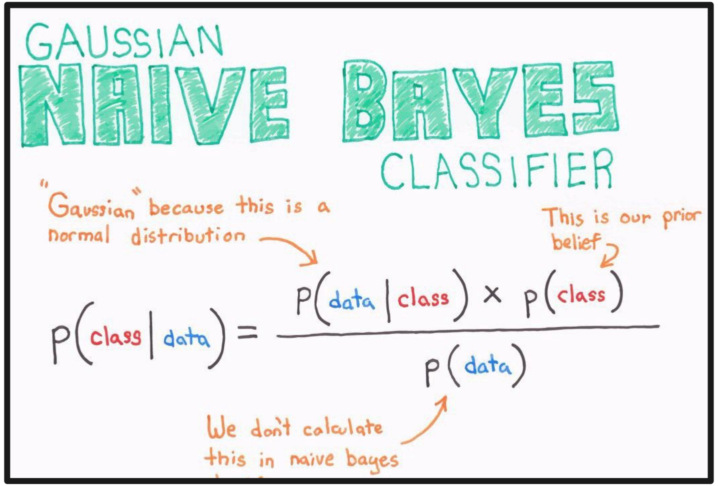
This process attempts to find the optimal probability for our target heuristic by **generating likelihood functions from new observations that *shift* our prior distributions in the direction of our expected heuristic**.



WRITTEN IDEATION

WHAT **TYPES OF DISTRIBUTIONS** CAN **BAYESIAN CLASSIFICATION** BE USED FOR?

ILLUSTRATING GAUSSIAN NAÏVE BAYESIAN CLASSIFIERS



The simplest of Bayesian classifier models is that of a **Naïve Bayes** model, which performs Bayesian classification with an extremely powerful naive assumption that **feature occurrences are independent of one another** (which dramatically simplifies the process).

One type of naive Bayesian model is the **Gaussian classifier**, which assumes that the class distribution is Gaussian or normal.

ILLUSTRATING MULTINOMIAL NAÏVE BAYESIAN CLASSIFIERS

Another type of naïve Bayesian classifier is that of the **multinomial model**, which can be a little abstract to conceptualize – plainly put, multinomial models assume that the **probability distribution is discretized across separate categories** rather than a fully continuous range.

While these distributions can still be treated as probability distributions, it's important to distinguish them as they relate to **domains where our data has explicitly discrete outcomes**, such as a coin toss (evokes a binomial distribution).

RECOMMENDED RESOURCE

Grant Sanderson's Explanatory Deconstruction of Bayes' Theorem

RECOMMENDED RESOURCE

Cassie Kozyrkov's Article on Bayesianism vs. Frequentism

THANK YOU
FOR YOUR TIME!