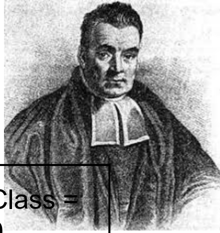
Jake Dineen

Understanding Multinomial Naive Bayes



Bayesian Rule:

P(class|X) = P(X|class) * P(class) / P(X)

Or

Posterior = Conditional * Prior / P(X)*

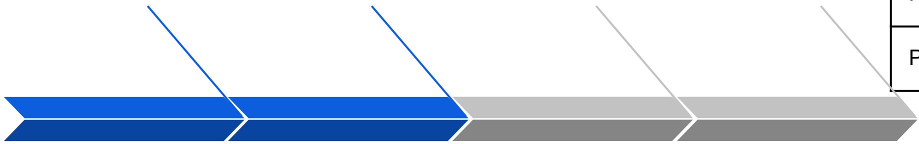
*P(X) can be ignored as we are trying to derive the highest posteriors

For Text Classification:

- Multinomial model for word frequency
- Benoulli model for boolean

Why the Naivete?

A naive bayes classifier is aptly named due to the assumption of independence by all features whose conditional probabilities are being calculated. The model refuses to reinforce the idea that the occurrence of one feature may be predicated on the presence of another - Correlation in linguistics is highly likely to occur, but fitting on joint probabilities may lead to memorization and overfitting come inference time. The ability to rank features based on their posteriors makes NB algorithms interpretable, and lower dimensional compute makes induction quicker than alternatives like Decision Trees, which can quickly grow too large and lose explainability.



Tokenize

Extract meaningful words.

Vectorize

Create term freq. matrix

Compute Prior and Conditional

Ref. Bayes theorem.

Compute Posterior

Solve Bayes for class mapping.

How it works

Doc1	This **place** is **terrible**.	Class = 0
Doc2	I **love** the **food** at this **place**. **Great customer service**.	Class = 1
Test	**Love**, **love**, **love** this **place**. It's **great**	Class = ?

*Words in bold would likely remain after removing stopwords.They contain some level of semantic meaning when looking at sentiment.

Token	Conditional Formula	Prior	Cond. Prob
Sample	(Count of word in class + 1 (smoothing)) / (Length of *total* vocab in class + Length of total unique tokens in both classes)	Class distribution	Equation solved.
P(Love|1)	(1 + 1) / (5 + 8)	1/2	.15
P(Love|0)	(0 + 1) / (2 + 8)	1/2	.10
P(Place|1)	(1 + 1) / (5 + 8)	1/2	.15
P(Place|0)	(1 + 1) / (2 + 8)	1/2	.20
P(great|1)	(1 + 1) / (5 + 8)	1/2	.15
P(great|0)	(0 + 1) / (2 + 8)	1/2	.10

	Formula	Posterior
P(1|test)	½ * (.15)^3 * .15 * .15	**0.000038**
P(0|test)	½ * (0.10)^3 *.2 * .10	0.000010