PGE 383 Lecture xx
Naïve Bayes

- Recall of Bayesian Approach
- Naïve Bayes Prediction
- Naïve Bayes Hands-on

Michael Pyrcz, The University of Texas at Austin

PGE 383 Lecture xx
Naïve Bayes

- Recall of Bayesian Approach

Michael Pyrcz, The University of Texas at Austin

Probability Definitions
Bayesian Statistics

Product Rule:

$$P(B \cap A) = P(A|B)\, P(B)$$

$$P(A \cap B) = P(B|A)\, P(A)$$

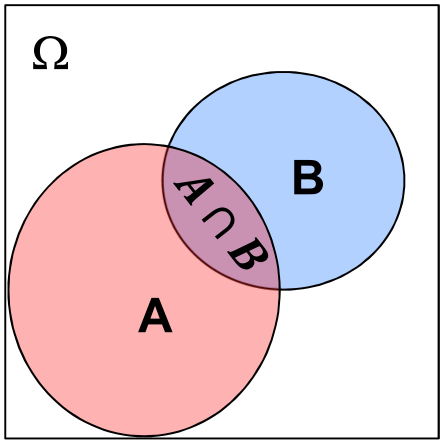
It follows that:

$$P(B \cap A) = P(A \cap B)$$

Therefore we combine two product rules, substitute:

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

We get Bayes' Theorem!



Venn Diagram – illustrating intersection.

Probability Definitions
Bayesian Statistics

Bayesian Statistical Approaches:

- probabilities based on a degree of belief in an event
- updated as new information is available
- solve probability problems that we cannot use simple frequencies

Probability Definitions
Bayesian Statistics

Bayes' Theorem:

Make a easy adjustment and we get the popular form.

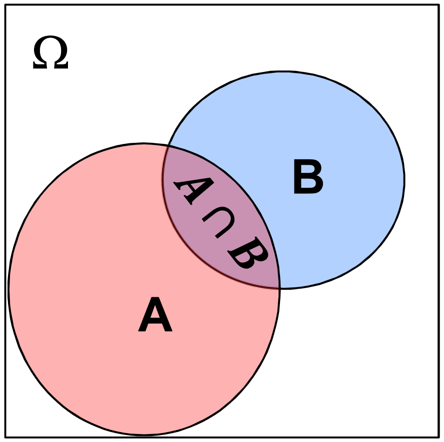
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Observations:

1. We are able to get P(A | B) from P(B | A) as you will see this often comes in handy.
2. Each term is known as:

$$\text{Posterior} = \frac{\text{Likelihood x Prior}}{\text{Evidence}}$$

3. Prior should have no information from likelihood.
4. Evidence term is usually just a standardization to ensure closure.



Venn Diagram – illustrating intersection.

Note: we got to Bayes' Theorem by fundamental frequentist approaches.

Probability Definitions
Bayesian Example

Bayes Theorem:

A common example of Bayes' Theorem for modeling the subsurface.

Model Updating with a New Data Source:

$$\underbrace{P(\text{Model} \mid \text{New Data})}_{\textbf{Posterior}} = \frac{\overbrace{P(\text{New Data} \mid \text{Model})}^{\textbf{Likelihood}}\ \overbrace{P(\text{Model})}^{\textbf{Prior}}}{\underbrace{P(\text{New Data})}_{\textbf{Evidence}}}$$

PGE 383 Lecture xx
Naïve Bayes

- **Naïve Bayes Prediction**

Michael Pyrcz, The University of Texas at Austin

Naïve Bayes Classifier

The Prediction Problem:

Given predictor features $x_1, \ldots, x_m$, predict the probability of response category C_k, with $\mathrm{k} = 1, \ldots, K$ possible categories.

This is our prediction problem, predict $C_k = f(x_1, \ldots, x_m)$.

With the naïve Bayes approach we will utilize the conditional probability

$$\mathrm{P}(C_k | x_1, \ldots, x_m) \ , \ \forall \ k = 1, \ldots, K$$

This would be a difficult inference problem, for which we would likely not have enough data, so we will make a simplifi

Naïve Bayes Classifier

The Naïve Bayes Method

$$\mathrm{P}(C_k | x_1, \dots, x_m)\ ,\ \forall\ k = 1, \dots, K$$

Let’s pose this as a Bayesian problem.

$$P(C_k | x_1, \dots, x_m) = \frac{P(x_1, \dots, x_m | C_k) P(C_k)}{P(x_1, \dots, x_m)}$$

Notice that we have prior, likelihood, evidence and posterior terms.

$$\text{Posterior} = \frac{\text{Likelihood x Prior}}{\text{Evidence}}$$

Naïve Bayes Classifier

The Naïve Bayes Method

$$P(C_k|x_1, \ldots, x_m) = \frac{P(x_1, \ldots, x_m|C_k)\boxed{P(C_k)}}{P(x_1, \ldots, x_m)}$$

The Prior:

We need the prior probability of cateogories $C_k, k = 1, \ldots, K$ independent of the predictor features.

This could be the global proportions seen in the training data or set naïve as a uniform distribution.

For example, if we are predicting low and high production wells from porosity and brittleness in an unconventional reservoir we could:

1. use the global proportion of low and high production wells observed
2. use 50% for low and 50% for high production

Naïve Bayes Classifier

The Naïve Bayes Method

$$P(C_k|x_1,\dots,x_m) = \frac{P(x_1,\dots,x_m|C_k)P(C_k)}{\boxed{P(x_1,\dots,x_m)}}$$

The Evidence:

Note that the evidence term does not consider the response category, C_k

The evidence term is constant over the categories, $C_k, k = 1,\dots,K$

It's only role is to standardize the resulting probabilities to sum to 1.0

This is the closure constraint – the sum of probabilities of all exhaustive, mutually exclusive outcomes must be 1.0.

$$P(x_1,\dots,x_m) = \sum_{k=1}^{K} P(x_1,\dots,x_m|C_k)P(C_k)$$

Naïve Bayes Classifier

The Naïve Bayes Method

$$P(C_k|x_1, \dots, x_m) = \frac{\boxed{P(x_1, \dots, x_m|C_k)}P(C_k)}{P(x_1, \dots, x_m)}$$

The Likelihood:

This is the difficult part of naïve Bayes. That's a potentially high dimensional joint conditional! Let's try working with it using basic Bayesian concepts.

Combine the likelihood and the prior to get one joint:

$$P(x_1, \dots, x_m|C_k)P(C_k) = P(x_1, \dots, x_m, C_k)$$

Naïve Bayes Classifier

The Naïve Bayes Method

$$P(C_k|x_1,\ldots,x_m) = \frac{\boxed{P(x_1,\ldots,x_m|C_k)}P(C_k)}{P(x_1,\ldots,x_m)}$$

The Likelihood:

Recursively expand the joint

$$P(x_1,\ldots,x_m,C_k) = P(x_1|x_2,\ldots,x_m,C_k)P(x_2,\ldots,x_m,C_k)$$

$$P(x_1,\ldots,x_m,C_k) = P(x_1|x_2,\ldots,x_m,C_k)P(x_2|x_3,\ldots,x_m,C_k)P(x_3,\ldots,x_m,C_k)$$

Let's generalize:

$$= P(x_1|x_2,\ldots,x_m,C_k)P(x_2|x_3,\ldots,x_m,C_k)\ldots P(x_{m-1}|x_m,C_k)P(x_m|C_k)P(C_k)$$

Naïve Bayes Classifier

The Naïve Bayes Method

$$P(C_k|x_1, \dots, x_m) = \frac{\boxed{P(x_1, \dots, x_m|C_k)}P(C_k)}{P(x_1, \dots, x_m)}$$

This is what we have now:

$$P(C_k|x_1, \dots, x_m) = \frac{\boxed{P(x_1|x_2, \dots, x_m, C_k)P(x_2|x_3, \dots, x_m, C_k) \dots P(x_{m-1}|x_m, C_k)P(x_m|C_k)}P(C_k)}{P(x_1, \dots, x_m)}$$

This is quite interesting. We can simplify this form greatly with an assumption of:

Conditional independence

Naïve Bayes Classifier

Conditional Independence

The predictor features are independent with eachother conditional the prediction of reponce feature.

For example:

$$P(x_1|x_2, \dots, x_m, C_k) = P(x_1|C_k)$$

We can exclude all the other predictor features from these terms!

- This greatly simplifies our inference problem.
- We now omit any interactions between features, $x_1, \dots, x_m$ with respect to predicting C_k

Naïve Bayes Classifier

The Naïve Bayes Method

$$P(C_k|x_1,\dots,x_m) = \frac{\boxed{P(x_1,\dots,x_m|C_k)}P(C_k)}{P(x_1,\dots,x_m)}$$

The Likelihood:

This is what we have now:

$$P(C_k|x_1,\dots,x_m) = \frac{\boxed{P(x_1|C_k)P(x_2|C_k)\dots P(x_{m-1}|C_k)P(x_m|C_k)}P(C_k)}{P(x_1,\dots,x_m)}$$

Now we need to estimate this set of conditional probabilities for each combination of predictor feature, $x_1,\dots,x_m$, and category, $C_k, k = 1,\dots,K$.

Naïve Bayes Classifier

Estimating the Likelihood Terms

$$P(x_1|C_k), P(x_2|C_k) \ldots P(x_{m-1}|C_k), P(x_m|C_k), k = 1, \ldots, K$$

We can estimate the conditional distribution by simply calculating the conditional probability density function.

For example:

- pool all predictor feature 1 values, $x_{1,j}$, over $j = 1, \ldots, n$ data samples
- calculate the associated continuous probability density function

We can further simplify our work by assuming a parametric conditional distribution

- for the complete Gaussian conditional we only need to estimate the conditioinal mean and variance.

Naïve Bayes Classifier

Let's work through an example by hand:

Let's say we want to estimate high or low production from average porosity and permeability over the well. We will use Gaussian naïve Bayes classification.

1. Pool all available data, separate training and testing data sets

Porosity	Brittleness	Production
26%	33%	High
28%	75%	High
7%	52%	High
29%	46%	High
14%	61%	High
28%	46%	High
22%	30%	High
30%	18%	Low
21%	82%	Low
29%	14%	Low
6%	78%	Low
24%	82%	Low
1%	87%	Low
3%	74%	Low
23%	80%	Low
17%	73%	Low
13%	98%	Low
8%	62%	Low

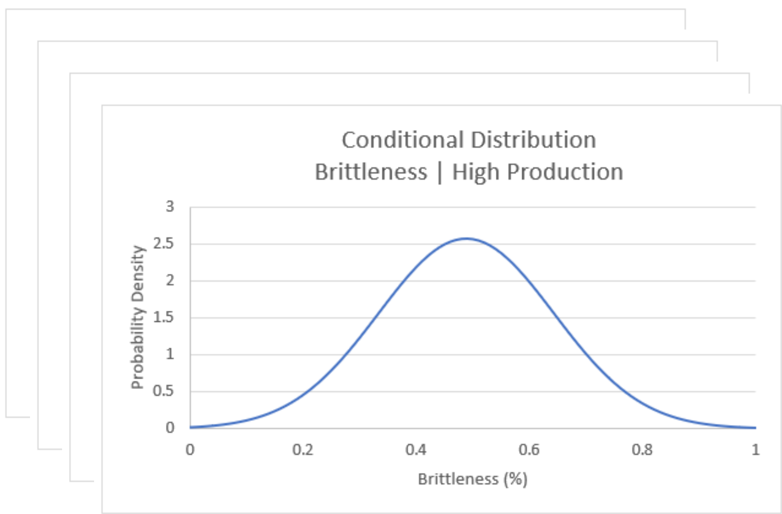
Naïve Bayes Classifier

Let's work through an example by hand:

2. Calculate the mean and variance for each coniditonal distribution.

	Porosity		Brittleness	
	Low	High	Low	High
Mean	16%	22%	68%	49%
StDev	10%	8%	27%	15%

3. Fit a Gaussian distribution to each conditional distribution.



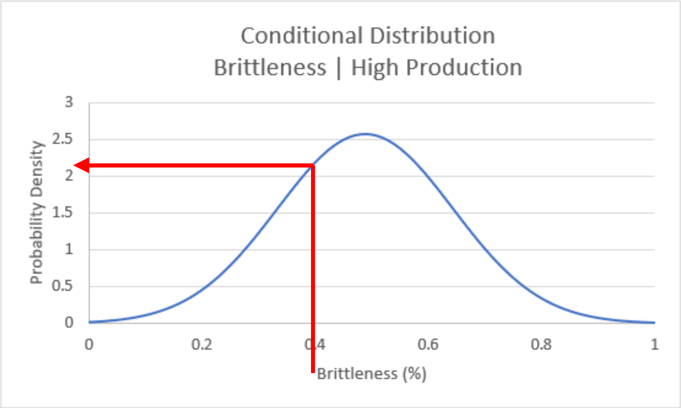
Naïve Bayes Classifier

Let's work through an example by hand:

4. Assign a prior probability

Prior Probability	
Low	39%
High	61%

5. We are ready to make predictions!



$$P(Brittleness = 40\%|High\ Production) = 2.2$$

We can use the density values, the evidence term will take care of closure.

Naïve Bayes Classifier

The Naïve Bayes Method

We estimate the posterior probability of production rate for any combination of porosity, φ, and brittleness, b.

$$P(High|por = \varphi, brittle = b)' \propto P(Porosity = \varphi|High)P(Brittle = b|High)P(High)$$

$$P(Low|por = \varphi, brittle = b)' \propto P(Porosity = \varphi|Low)P(Brittle = b|Low)P(Low)$$

We indicate proportional, $\propto$, as we will standardize to sum one, instead of calculating the evidence term directly.

$$P(High|por = \varphi, brittle = b) = \frac{(High|por = \varphi, brittle = b)'}{(High|por = \varphi, brittle = b)' + (Low|por = \varphi, brittle = b)'}$$

$$P(Low|por = \varphi, brittle = b) = \frac{(Low|por = \varphi, brittle = b)'}{(Low|por = \varphi, brittle = b)' + (High|por = \varphi, brittle = b)'}$$

PGE 383 Lecture xx
Naïve Bayes

- Naïve Bayes Hands-on

Michael Pyrcz, The University of Texas at Austin

Neural Net Demonstration

Demonstration workflow with naïve Bayes for supervised learning from training data.

Subsurface Data Analytics

Naive Bayes Classification for Subsurface Data Analytics in Python

Michael Pyrcz, Associate Professor, University of Texas at Austin

Twitter | *GitHub* | *Website* | *GoogleScholar* | *Book* | *YouTube* | *LinkedIn*

PGE 383 Exercise: Naive Bayes Classification for Subsurface Data Analytics in Python

Here's a simple workflow, demonstration of naive Bayes classification for subsurface modeling workflows. This should help you get started with building subsurface models that with predictions based on multiple sources of information.

This method is great as it builds directly on our knowledge Bayesian statistics to provide a simple, but flexible classification method.

Bayesian Updating

The naive Bayes classifier is based on the conditional probability of a category, k, given n features, $x_1, \dots, x_n$.

$$p(C_k | x_1, \dots, x_n)$$

we can solve this with Bayesian updating:

$$p(C_k | x_1, \dots, x_n) = \frac{p(x_1, \dots, x_n | C_k) p(C_k)}{p(x_1, \dots, x_n)}$$

we can exand the full joint distribution recursively as follows:

$$p(C_k, x_1, \dots, x_n)$$

with a simple reordering

$$p(x_1, \dots, x_n, C_k)$$

expansion of the joint with the conditional and prior

$$p(x_1 | x_2, \dots, x_n, C_k) p(x_2, \dots, x_n, C_k)$$

continue recursively expanding

$$p(x_1 | x_2, \dots, x_n, C_k) p(x_2 | x_3, \dots, x_n, C_k) p(x_3, \dots, x_n, C_k)$$

we can generalize as

$$p(x_1 | x_2, \dots, x_n, C_k) p(x_2 | x_3, \dots, x_n, C_k) p(x_3 | x_4, \dots, x_n, C_k) \dots (x_{n-1} | x_n, C_k)(x_n | C_k) p(C_k)$$

Naive Bayes Approach

The likelihood, conditional probability with the joint conditional is difficult to calculate. It requires information about the joint relationship between $x_1, \dots, x_n$ features. As n increases this requires a lot of data to inform the joint distribution.

File SubsurfaceDataAnalytics_NaiveBayes.ipynb at https://git.io/fj6ax.

PGE 383 Lecture xx
Naïve Bayes

- Recall of Bayesian Approach
- Naïve Bayes Prediction
- Naïve Bayes Hands-on

Michael Pyrcz, The University of Texas at Austin