PGE 383 Subsurface Machine Learning

Lecture 12: Naïve Bayes Classifier

Lecture outline:

- **Recall of Bayesian Approach**
- **Naïve Bayes Classifier**
- **Naïve Bayes Classifier Example**
- **Naïve Bayes Hands-on**

© 2019 Michael Pyrcz

Motivation

Another predictive machine learning with Bayesian methods

- classification for categorical response feature
- an intuitive classification method based on fundamental probability operators

© 2019 Michael Pyrcz

PGE 383 Subsurface Machine Learning

Lecture 12: Naïve Bayes Classifier

Lecture outline:

- **Recall of Bayesian Approach**

© 2019 Michael Pyrcz

Probability Definitions Bayesian Statistics

Recall the Multiplication Rule,

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

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

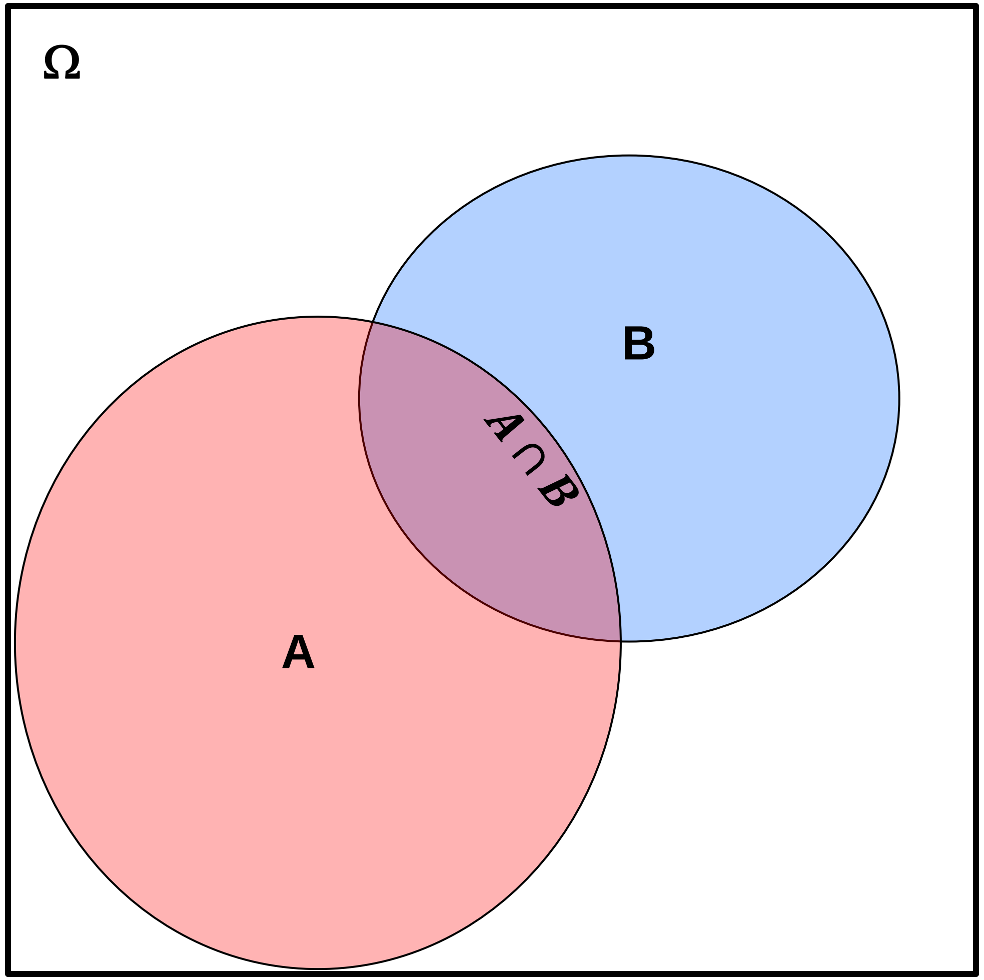
It follows that:

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

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.

© 2019 Michael Pyrcz

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
- credibility intervals that are more rational than confidence intervals
- integrate uncertainty in the model parameters, we will sample an uncertainty model!

© 2019 Michael Pyrcz

Bayes' Theorem Details

Bayes' Theorem Observations:

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

Observations:

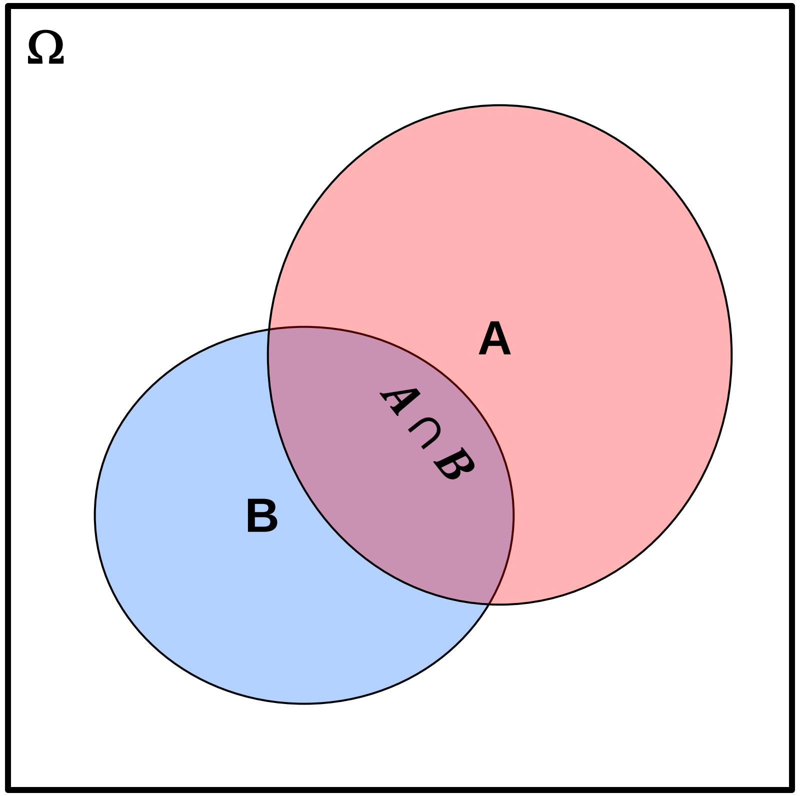
1. We can get $P(A \mid B)$ from $P(B \mid A)$, as you will see this often comes in handy.

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 – illustration of events and relations to each other.

© 2019 Michael Pyrcz

Bayes' Theorem
Common Approach

Bayes' Theorem:

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

Let's change the labels to communicate model updating with a new data source:

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

© 2019 Michael Pyrcz

PGE 383 Subsurface Machine Learning

Lecture 12: Naïve Bayes Classifier

Lecture outline:

- **Naïve Bayes Classifer**

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Classification 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.

$$\mathrm{P}(C_k | x_1, \ldots, x_m)$$

For example, we can conveniently state our prediction model as estimating conditional probabilities,

Mudstone: $\mathrm{P}(C_k = M | x_1, \ldots, x_m)$

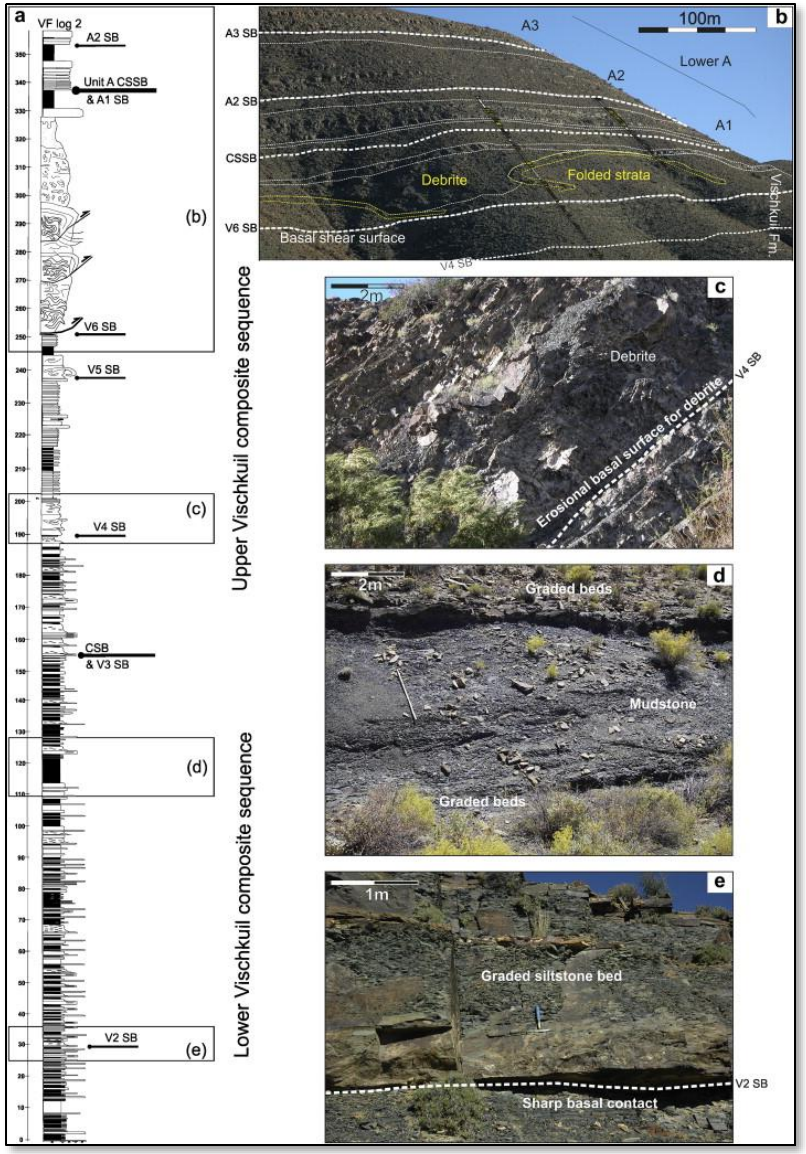
Siltstone: $\mathrm{P}(C_k = Si | x_1, \ldots, x_m)$

Sandstone: $\mathrm{P}(C_k = Ss | x_1, \ldots, x_m)$

Note, **cardinality** is the number of unique categories in a categorical feature. 3 in this case.

With closure,

$$\mathrm{P}(C_k = M | x_1, \ldots, x_m) + \mathrm{P}(C_k = Si | x_1, \ldots, x_m) + \mathrm{P}(C_k = Ss | x_1, \ldots, x_m) = 1.0$$



Summary characteristics of the Vischkuil Formation, which marks the initiation of the Laingsburg deepwater stratigraphy (Flint et al., 2011).

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Prediction Problem

Once again, here is our prediction model,

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

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

- $x_1, \dots, x_m$ is a potentially high dimensionality joint

The Bayesian approach can help us tackle this problem,

$$\mathrm{P}(C_k | x_1, \dots, x_m) \approx \lim_{n \to \infty} \frac{N(C_k, x_1, \dots, x_m)}{N(x_1, \dots, x_m)}$$

- let's not tackle this head on with the frequentist method.
- instead, we cast it as a posterior!

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Naïve Bayes Method

Given our classification problem,

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

Let's pose this as a Bayesian updating problem.

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

Recall,

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

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

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

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Naïve Bayes Method

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

Prior:

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

- **We cannot use the training data to estimate this – information leakage from likelihood to prior!**

This could be the global proportions seen in analog / other datasets 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 in the entire basin, outside your study

2. use 50% for low and 50% for high production for a naïve prior

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Naïve Bayes Method

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

Evidence:

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

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

Its 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.

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

Divide each prediction by the sum over all the categories.

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Naïve Bayes Method

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

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:

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

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Naïve Bayes Method

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

Likelihood:

Recursively expand the joint,

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

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

Let's generalize:

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

Prior is back!

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Naïve Bayes Method

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

Likelihood:

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

© 2019 Michael Pyrcz

Naïve Bayes Classifier

Conditional Independence

The predictor features are independent with each other conditional the prediction of response feature.

For example:

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

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

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

© 2019 Michael Pyrcz

Naïve Bayes Classifier

The Naïve Bayes Method

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

Likelihood:

Under the assumption of conditional independence, we have:

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

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

© 2019 Michael Pyrcz

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 $j = 1, \ldots, n$ training data

- calculate the associated continuous conditional probability density function (PDF)

$$f_{X_i|C_k}(X_i)$$

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

- for the complete Gaussian conditional we only need to estimate the conditional mean and variance to get the entire conditional PDF.

© 2019 Michael Pyrcz

PGE 383 Subsurface Machine Learning

Lecture 12: Naïve Bayes Classifier

Lecture outline:

- **Naïve Bayes Classifier Example**

© 2019 Michael Pyrcz

Naïve Bayes Classifier Example

Let's Calculate the Likelihoods By-hand

Let's say we want to estimate high or low production from average porosity and permeability over the well.

- We 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

© 2019 Michael Pyrcz

Naïve Bayes Classifier Example

Let's Calculate the Likelihoods By-hand

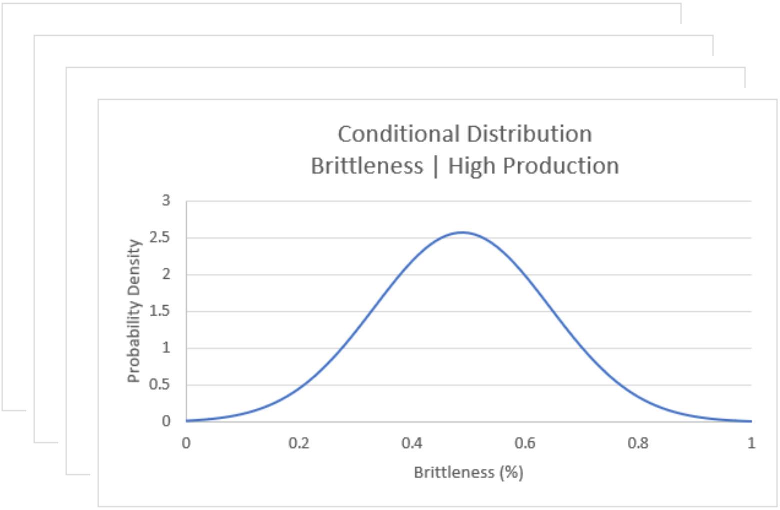
Let's say we want to estimate high or low production from average porosity and permeability over the well.

- We use Gaussian naïve Bayes classification.

2. Calculate the mean and variance for each conditional 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.



© 2019 Michael Pyrcz

Naïve Bayes Classifier Example

Let's Calculate the Likelihoods By-hand

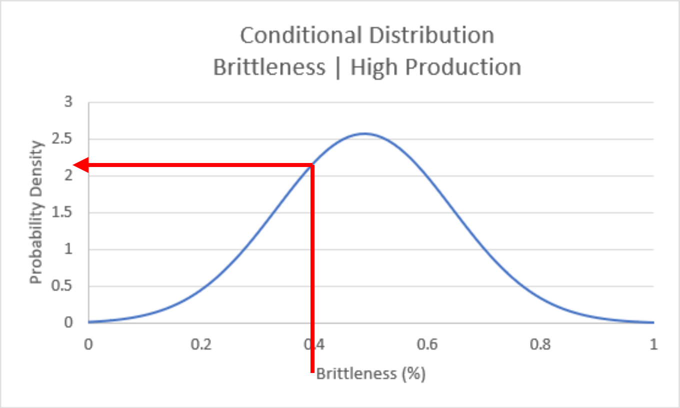
Let's say we want to estimate high or low production from average porosity and permeability over the well.

- We use Gaussian naïve Bayes classification.

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.

© 2019 Michael Pyrcz

Naïve Bayes Classifier Example

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.

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

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

© 2019 Michael Pyrcz

PGE 383 Subsurface Machine Learning

Lecture 12: Naïve Bayes Classifier

Lecture outline:

- **Naïve Bayes Hands-on**

© 2019 Michael Pyrcz

Naïve Bayes Classifier Demonstration in Python

Demonstration of Naïve Bayes Classifier with a well-documented workflow.

Applied Machine Learning in Python: a Hands-on Guide with Code

Applied Machine Learning in Python: a Hands-on Guide with Code

Machine Learning Concepts

Workflow Construction and Coding

Probability Concepts

Loading and Plotting Data and Models

Univariate Analysis

Multivariate Analysis

Feature Transformations

Feature Ranking

Cluster Analysis

Density-based Clustering

Spectral Clustering

Principal Components Analysis

Naive Bayes Classifier

Michael J. Pyrcz, Professor, The University of Texas at Austin

Twitter | GitHub | Website | GoogleScholar | Book | YouTube | Applied Geostats in Python e-book | LinkedIn

Chapter of e-book "Applied Machine Learning in Python: a Hands-on Guide with Code".

Cite this e-Book as:

Pyrcz, M.J., 2024, Applied Machine Learning in Python: a Hands-on Guide with Code, https://geostatsguy.github.io/MachineLearningDemos_Book.

The workflows in this book and more are available here:

Cite the MachineLearningDemos GitHub Repository as:

Pyrcz, M.J., 2024, MachineLearningDemos: Python Machine Learning Demonstration Workflows Repository (0.0.1). Zenodo. DOI 10.5281/zenodo.13835318

By Michael J. Pyrcz

© Copyright 2024.

This chapter is a tutorial for / demonstration of **Naive Bayes Classifier**.

YouTube Lecture: check out my lectures on:

Naïve Bayes chapter of Applied Machine Learning in Python e-book.

© 2019 Michael Pyrcz

PGE 383 Subsurface Machine Learning

Lecture 12: Naïve Bayes Classifier

Lecture outline:

- **Recall of Bayesian Approach**
- **Naïve Bayes Classifier**
- **Naïve Bayes Classifier Example**
- **Naïve Bayes Hands-on**

© 2019 Michael Pyrcz