Welcome to Online Engineering at George Washington University

Class will begin shortly

Audio: To eliminate background noise, please be sure your audio is muted. To speak, please click the hand icon at the bottom of your screen (**Raise Hand**). When instructor calls on you, click microphone icon to unmute. When you've finished speaking, ***be sure to mute yourself again***.

Chat: Please type your questions in Chat.

Recordings: Please note the recording of this class meeting will be available to download later today. The class recordings are to be used exclusively by registered students in this particular class. Releasing these recordings is strictly prohibited.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

SEAS 8510
Analytical Methods for Machine Learning

Lecture 6

Dr. Zachary Dennis

Adapted from Rodriguez, Etemadi, Fall 2023, George Washington University

Agenda

9:00 – 9:15		Discussion Group
9:15 – 10:00		Test Review
10:00 – 10:30		Homework Review
10:30 – 10:40		*BREAK (10 min)*
10:40 – 11:45		Naïve Bayes, Random Forests, Discrete Random Variables
11:45 – 12:00		Homework and Discussion Look Ahead

Assignments

Last week: Homework 4 and Discussion 4 due on 4/27 at 9 AM Eastern

This week: Homework 5 and Discussion 5 due on 5/4 at 9 AM Eastern

Midterm Statistics

High – 200

Low – 60

Mean – 151

Median – 160

Std Dev – 34

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Naïve Bayes

Naïve Bayes

Family of classification algorithms

- Set of features (X1, X2, X3, …. Xn)
- Predicting Class Y

Naïve Bayes “naively” assumes all features are independent.

Algorithm Steps:

1. Create frequency tables
2. Create likelihood tables
3. Calculate posterior probability using Bayes’ theorem
4. Predicts higher probability

Naïve Bayes Example

Data #	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Y
2	Red	Sports	Domestic	N
3	Red	Sports	Domestic	Y
4	Yellow	Sports	Domestic	N
5	Yellow	Sports	Imported	Y
6	Yellow	SUV	Imported	N
7	Yellow	SUV	Imported	Y
8	Yellow	SUV	Domestic	N
9	Red	SUV	Imported	N
10	Red	Sports	Imported	Y

1. Create frequency tables
2. Create likelihood tables

Let's start with **Color**….

Frequency Table

		Stolen	
		Y	N
Color	Red		
	Yellow		

Likelihood Table

		Stolen	
		P(Y)	P(N)
Color	Red		
	Yellow		

Naïve Bayes Example

Data #	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Y
2	Red	Sports	Domestic	N
3	Red	Sports	Domestic	Y
4	Yellow	Sports	Domestic	N
5	Yellow	Sports	Imported	Y
6	Yellow	SUV	Imported	N
7	Yellow	SUV	Imported	Y
8	Yellow	SUV	Domestic	N
9	Red	SUV	Imported	N
10	Red	Sports	Imported	Y

Type….

Frequency Table

		Stolen	
		Y	N
Type	Sports		
	Suv		

Likelihood Table

		Stolen	
		P(Y)	P(N)
Type	Sports		
	SUV		

Origin….

Frequency Table

		Stolen	
		Y	N
Origin	Domestic		
	Imported		

Likelihood Table

		Stolen	
		P(Y)	P(N)
Type	Domestic		
	Imported		

Naïve Bayes Example

3. Calculate posterior probability using Bayes' theorem
4. Predicts class with higher probability

Data #	Color	Type	Origin	Stolen?
New	Red	SUV	Domestic	?

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Random Forests

Ensemble Method

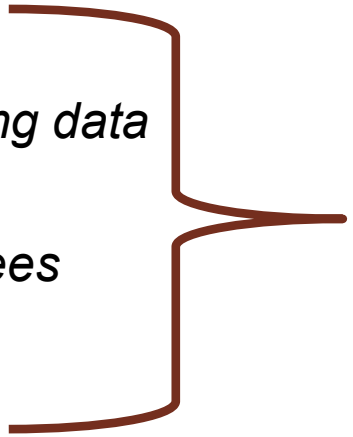
Example of an ensemble method:

Train a group of Decision Tree Classifiers

Each on a random subset of the training data

Make predictions using each of the individual trees

Then predict the class the gets the most votes

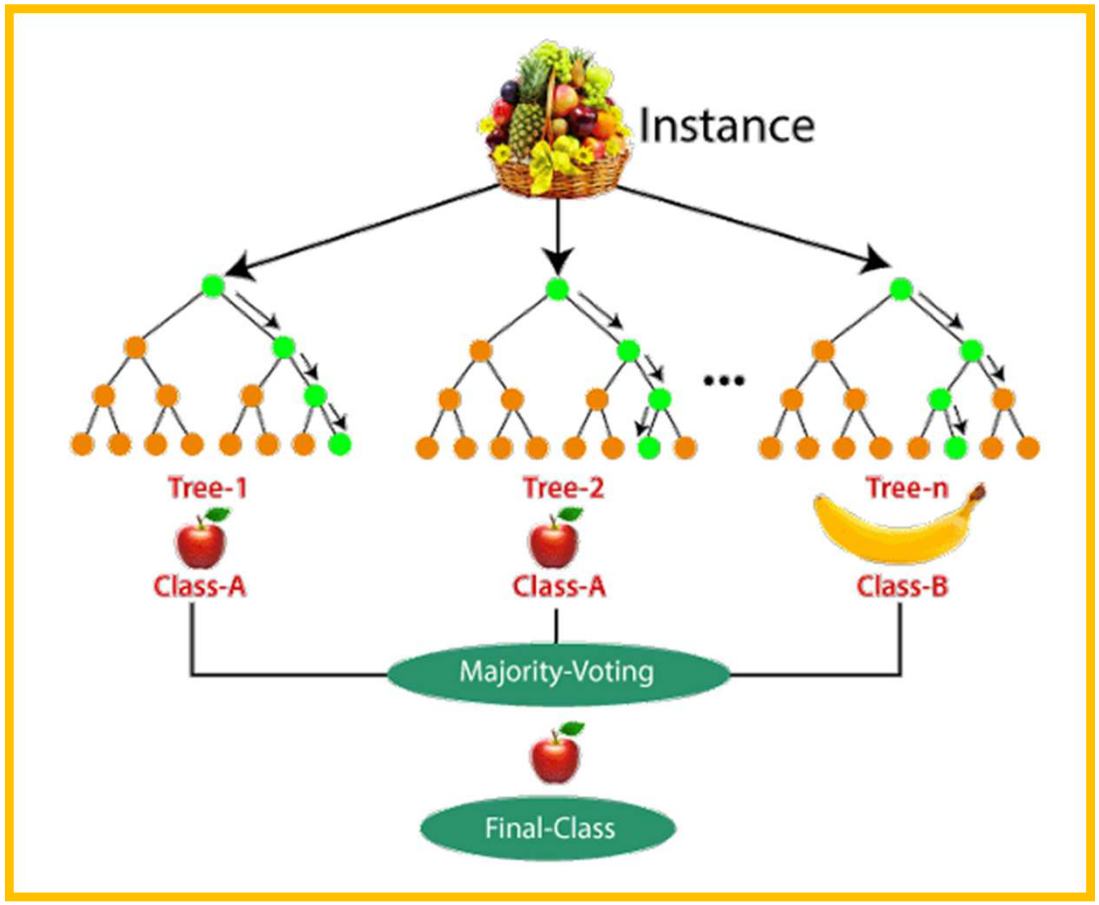


Random Forest

Often use ensemble method near end of a project, after already building a few good predictors. Combine into even better predictor.

(Geron, 2019, p. 189)

Random Forests



- Random Forest is an ***ensemble*** of Decision Trees, generally trained via the bagging method (or sometimes pasting)
- Random Forest introduces extra randomness when growing trees; instead of searching for very best feature when splitting a node, it searches for the best feature among a random subset of features
- Results in greater tree diversity, which trades a higher bias for a lower variance, generally yielding an overall better model

(Geron, 2019, p. 197-8)
Image: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

Random Forest Motivation Example

- Let's say we have this data set:

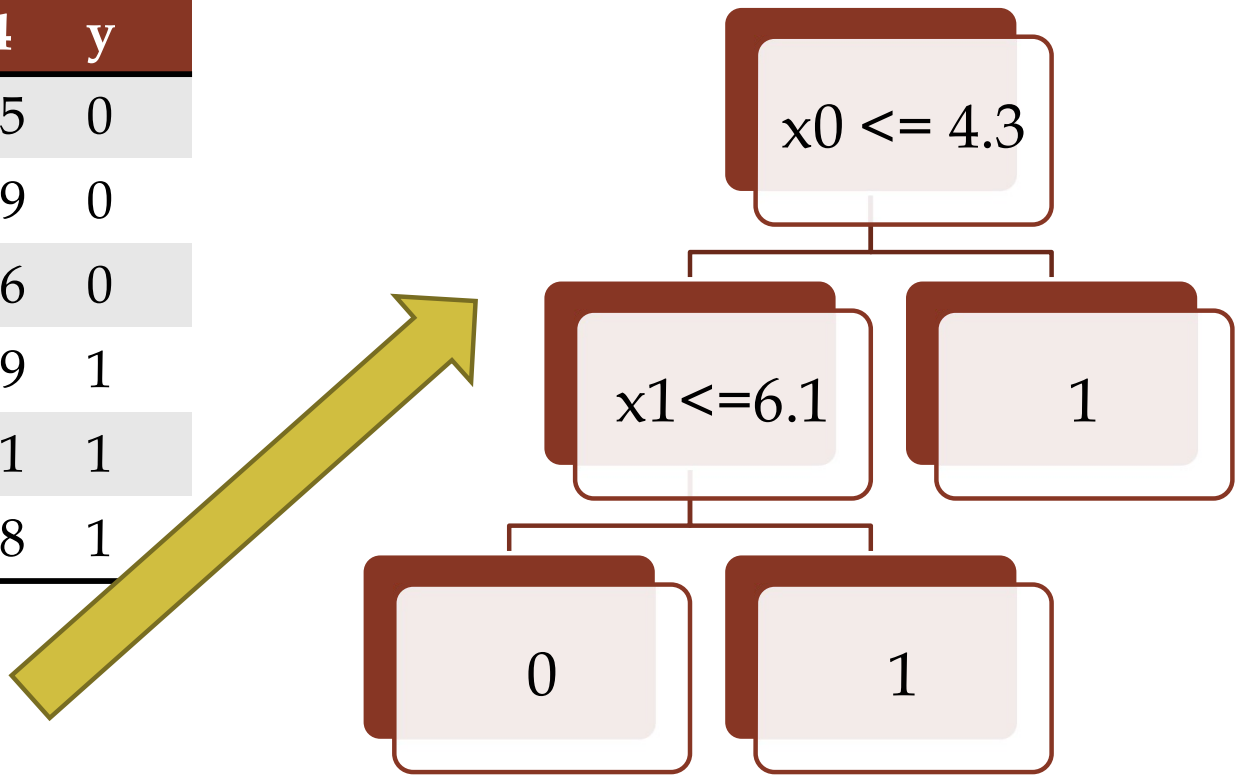
id	x0	x1	x2	x3	x4	y
0	4.3	4.9	4.4	4.7	5.5	0
1	3.9	6.1	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1

Random Forest Motivation Example

- Let's say we have this data set:

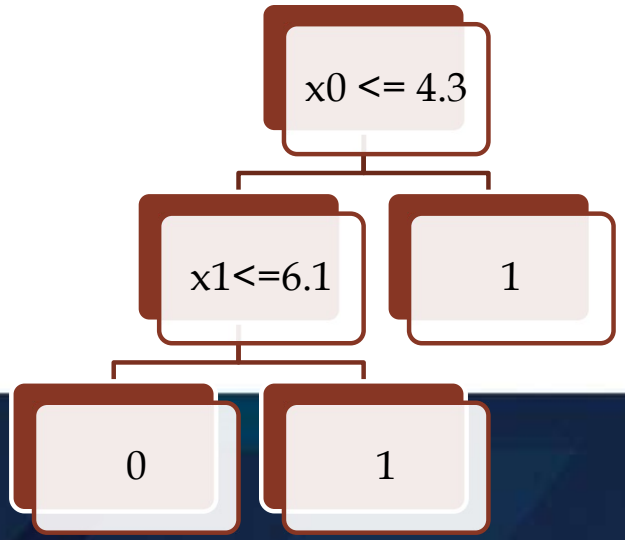
id	x0	x1	x2	x3	x4	y
0	4.3	4.9	4.4	4.7	5.5	0
1	3.9	6.1	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1

- And we train a decision tree



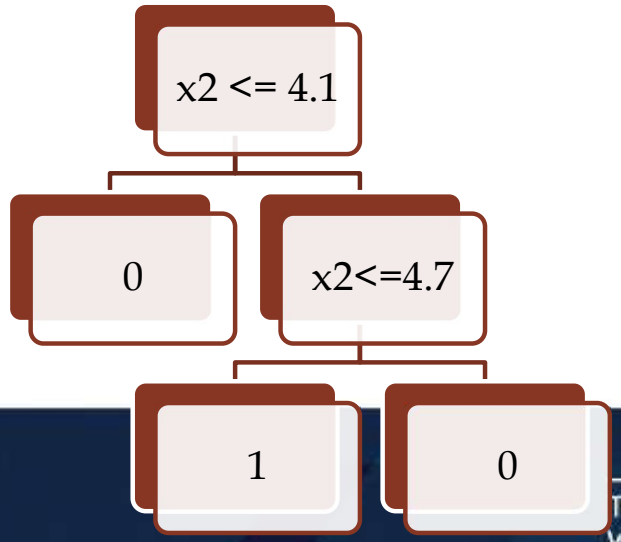
Random Forest Motivation Example

id	x0	x1	x2	x3	x4	y
0	4.3	4.9	4.4	4.7	5.5	0
1	3.9	6.1	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1



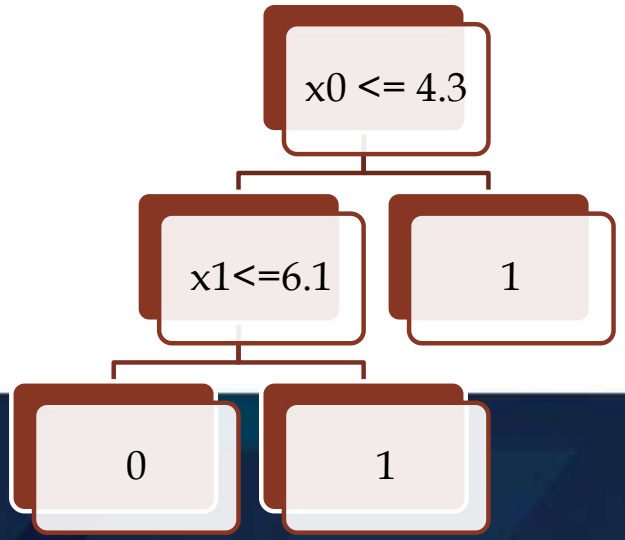
- If the data set changes:

id	x0	x1	x2	x3	x4	y
0	4.3	4.9	4.4	4.7	5.5	0
1	6.5	4.4	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1



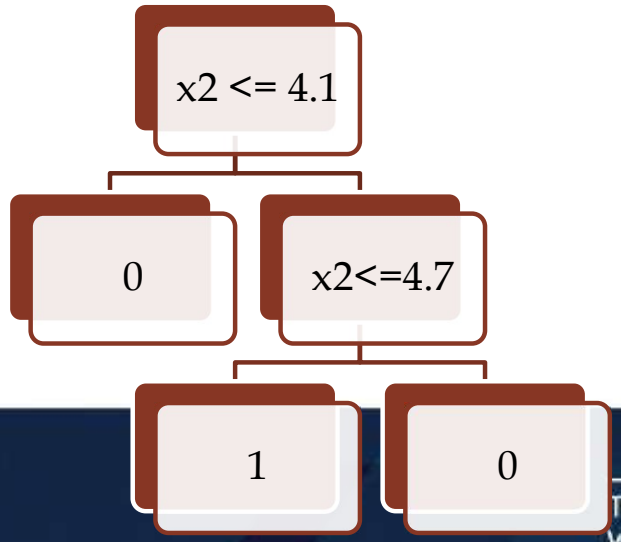
Random Forest Motivation Example

id	x0	x1	x2	x3	x4	y
0	4.3	4.9	4.4	4.7	5.5	0
1	3.9	6.1	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1



- If the data set changes:

id	x0	x1	x2	x3	x4	y
0	4.3	4.9	4.4	4.7	5.5	0
1	6.5	4.4	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1



Random Forest Example

id	x0	x1	x2	x3	x4	y
0	4.3	4.9	4.4	4.7	5.5	0
1	3.9	6.1	5.9	5.5	5.9	0
2	2.7	4.8	4.1	5.0	5.6	0
3	6.6	4.4	4.5	3.9	5.9	1
4	6.5	2.9	4.7	4.6	6.1	1
5	2.7	6.7	4.2	5.3	4.8	1

- Randomly select rows from original data set to build data sets for models to train with.
- Bootstrapping = with replacement

id
2
0
2
4
5
5

id
2
1
3
1
4
4

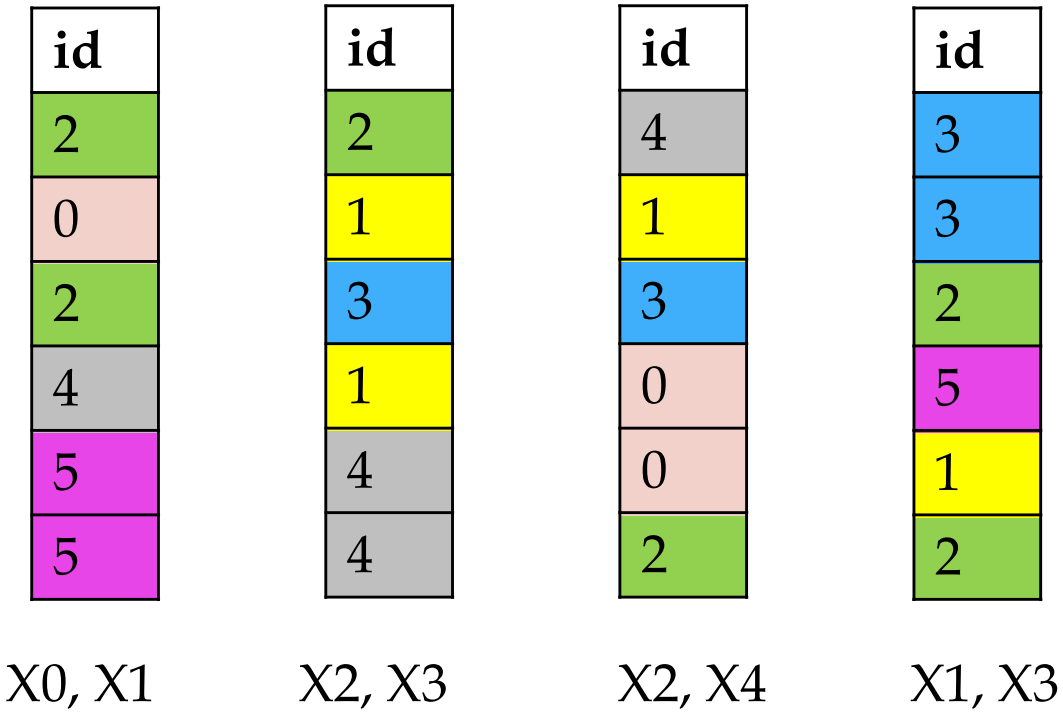
id
4
1
3
0
0
2

id
3
3
2
5
1
2

Random Forest Example Slides
Adapted from: https://youtu.be/v6VJ2RO66Ag

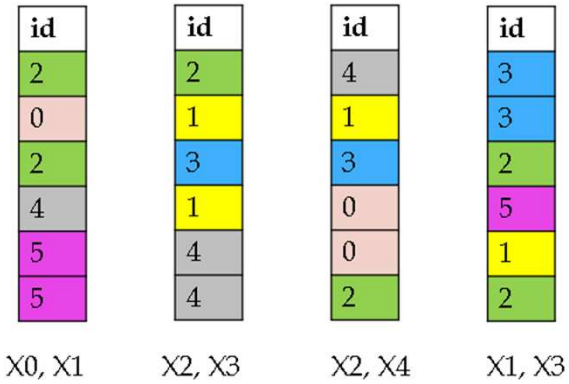
Random Forest Example

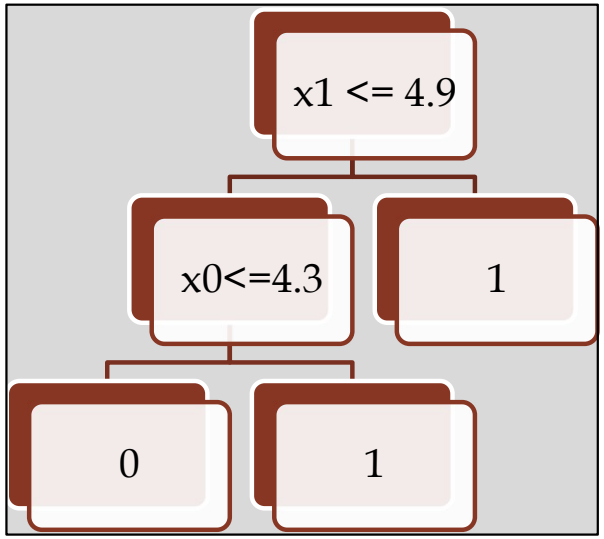
Train a decision tree for each of the random datasets, but only use a random subset of the features for each tree

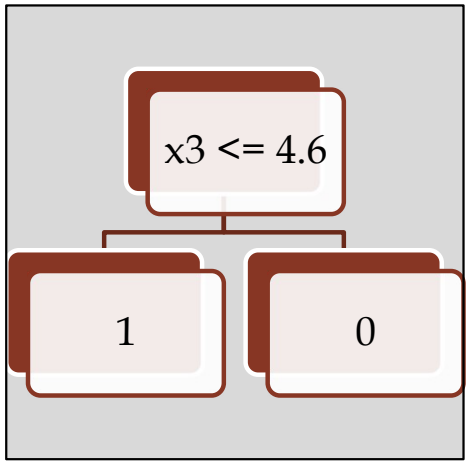


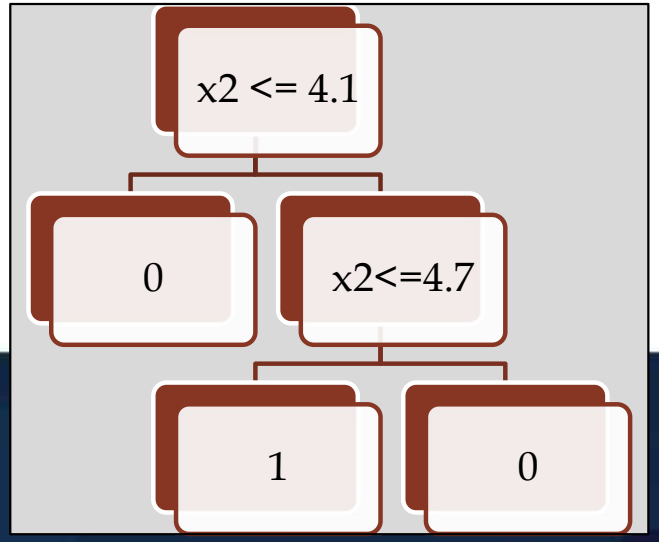
Random Forest Example

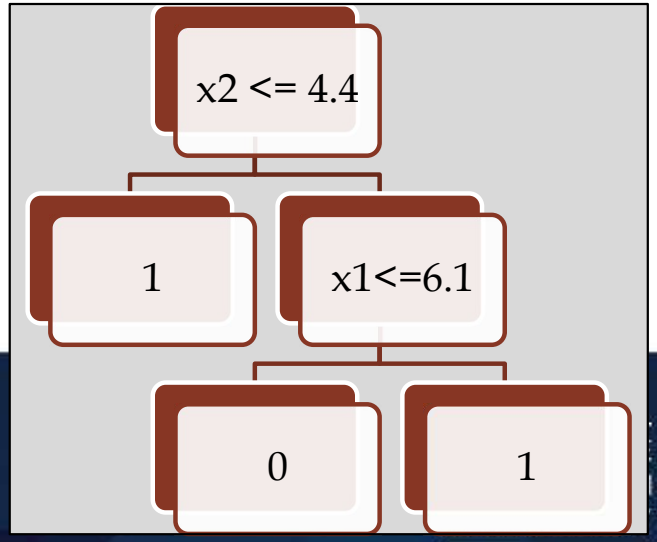
Train a decision tree for each of the random datasets, but only use a random subset of the features for each tree







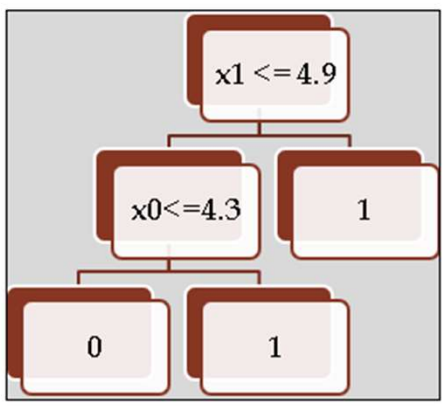


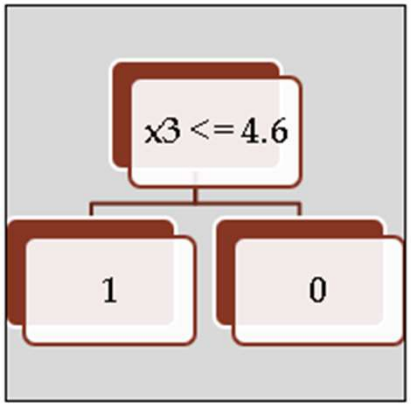


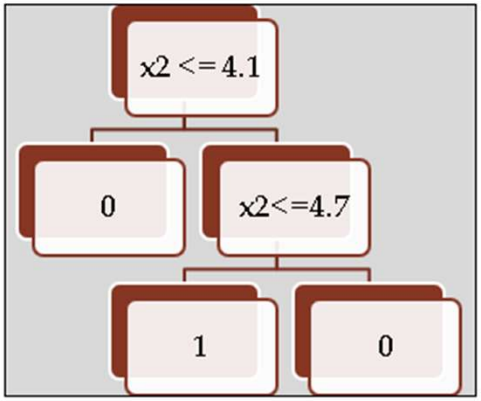
Random Forest Example

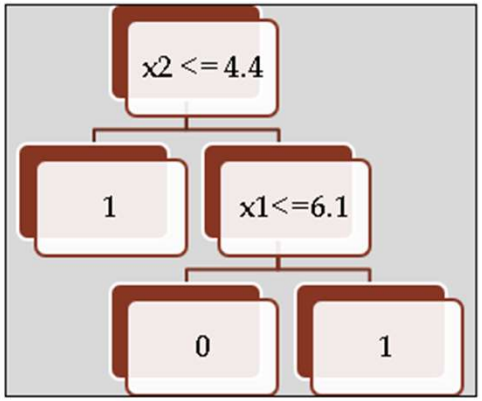
New instance:

x0	x1	x2	x3	x4
2.8	6.2	4.3	5.3	5.5









1 0 1 1

Random Forest Prediction = Class 1

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Discrete Distributions

Random Variables

Definition

For a given sample space S of some experiment, a **random variable** (rv) is any rule that associates a number with each outcome in S . In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real numbers.

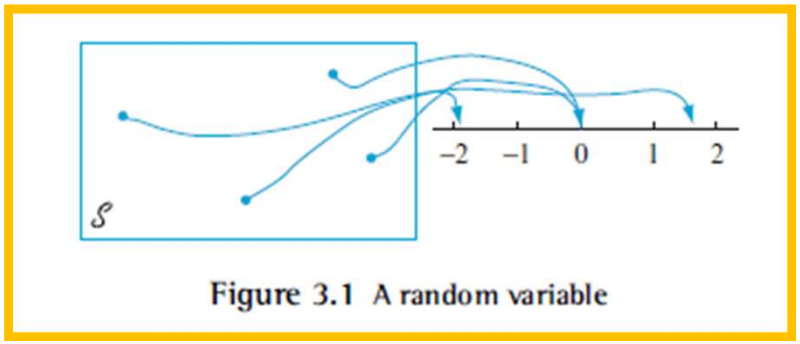


Figure 3.1 A random variable

(Devore & Berk, 2018, p.98)

Two Types of Random Variables

Definition

A **discrete** random variable is a rv whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on ("countably infinite").

A random variable is **continuous** if *both* of the following apply:

1. Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g., from $-\infty$ to ∞) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$)
2. No possible value of the variable has positive probability, that is, $P(X=c)=0$ for any possible value c

(Devore & Berk, 2018, p.98)

Two Types of Random Variables

Definition

A **discrete** random variable is a rv whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on ("countably infinite").

A random variable is **continuous** if *both* of the following apply:

1. Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g., from $-\infty$ to ∞) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$)
2. No possible value of the variable has positive probability, that is, $P(X=c)=0$ for any possible value c

(Devore & Berk, 2018, p.98)

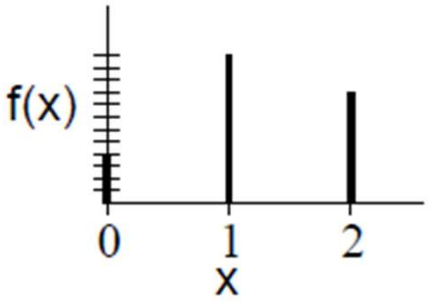
Representing Random Variables

Probability Mass Function (pmf)

When variables take on values that are countable or listable from smallest to largest they are called Discrete Random Variables

A pmf, f(x), of a random variable X is a function representing the values of a random variable with their associated probabilities – can be specified in tabular, graphic or equation form

X	f(x)
0	0.16
1	0.48
2	0.36



$$f(x) = \binom{2}{x}(0.6)^x(0.4)^{2-x}$$

$$x = 0,1,2$$

(Devore & Berk, 2018, p.101)

Representing Random Variables

Cumulative Distribution Function (cdf)

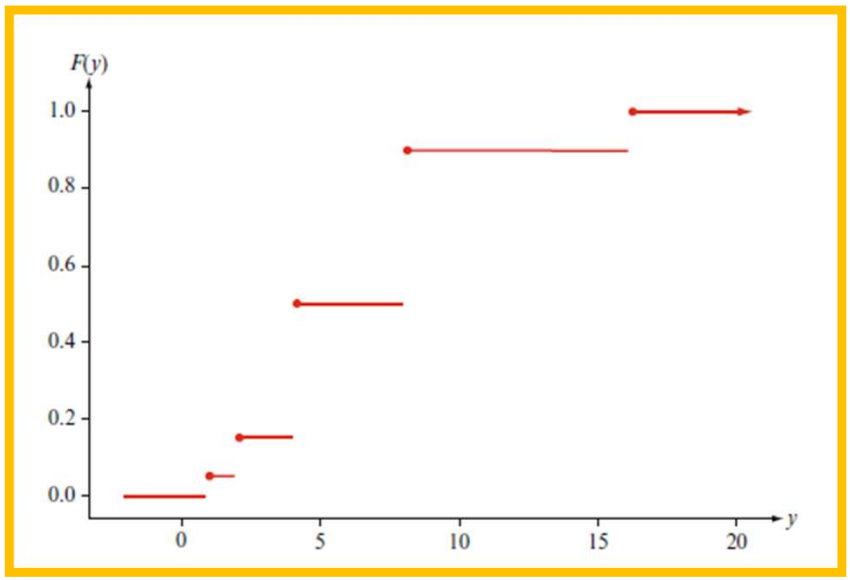
A CDF, F(x), of a random variable X is a function representing P(X<=x)

For discrete random variables this is a step function

pdf

y	1	2	4	8	16
$p(y)$	.05	.10	.35	.40	.10

cdf

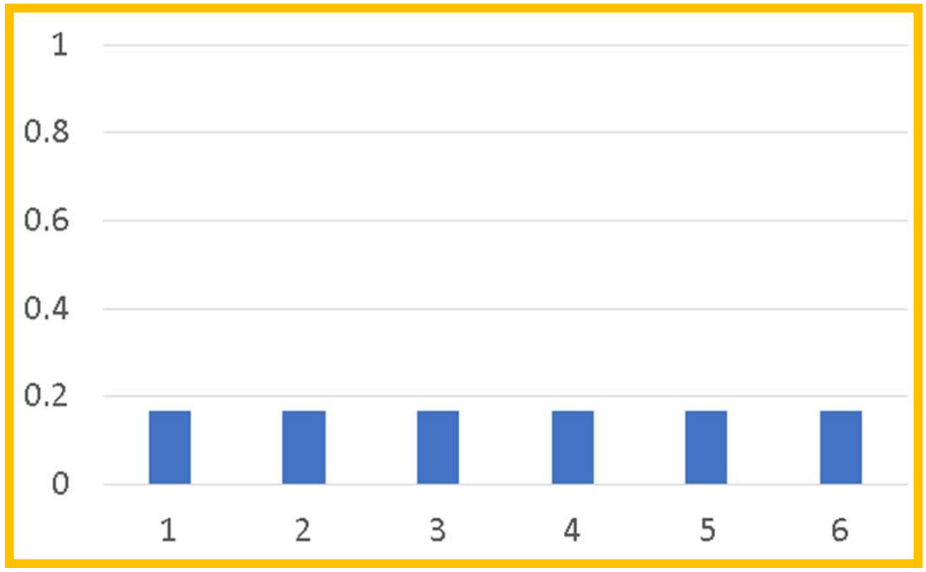


(Devore & Berk, 2018, p.105-106)

Uniform Distribution

The probabilities of each outcome are **evenly distributed** across the sample space

Ex: Rolling a fair die has 6 discrete, equally probable outcomes



Notation: $X \sim U(n)$

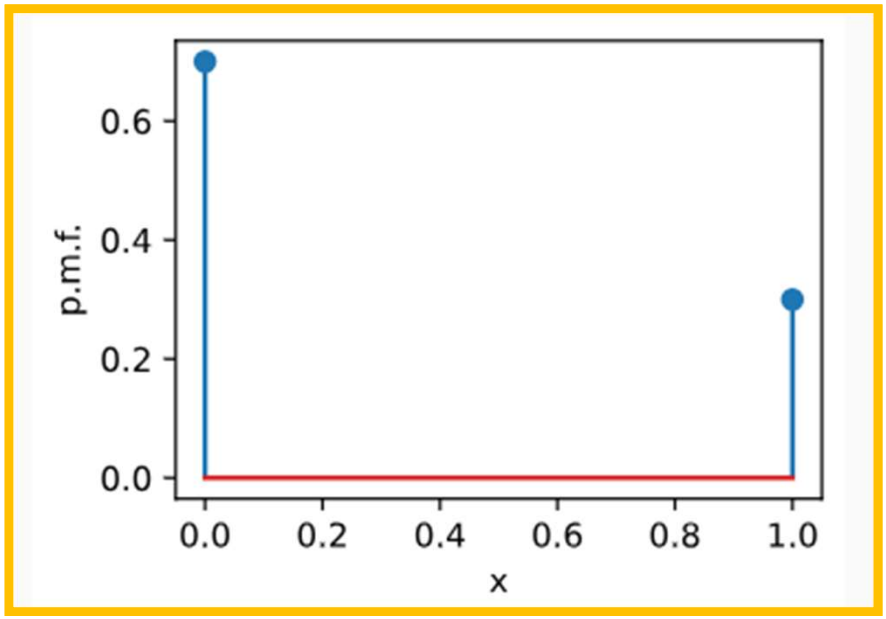
The probability of each value i $\in \{1, 2, \ldots . n\}$ is

$$p_i = \frac{1}{n}$$

Bernoulli Distribution

Any random variable whose only possible value are 0 and 1 is called a **Bernoulli random variable**

Notation: $X \sim Bernoulli(p)$



p = 0.3
1 – p = 0.7

https://d2l.ai/chapter_appendix-mathematics-for-deep-learning/distributions.html

Binomial Distribution

Binomial means there are two discrete, mutually exclusive outcomes

- Heads **or** tails
- On **or** off
- Defective **or** non defective
- Success **or** failure

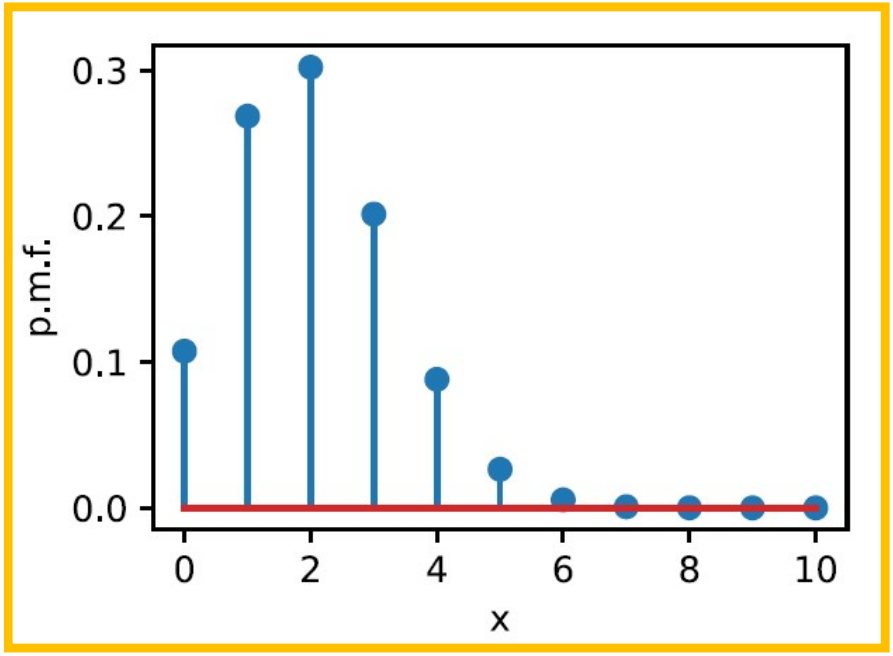
4 Criteria for an experiment to follow the Binomial Distribution:

- The experiment consist of a sequence of n smaller experiments called trials, where n is fixed in advance.
- Each trial can result in one of the same two possible outcomes, we generically denote by success (S) or failure (F).
- Trials are independent, so that the outcome of any particular trial does not influence the outcome on any other trial,
- The probability of success P(S) is constant from trial to trial; we denote this as probability p.

(Devore & Berk, 2018, p.128)

Binomial Distribution

- Performing a sequence of *n* independent experiments, each of which has probability *p* of succeeding, where $p \in \{0, 1\}$
- The probability of getting *k* successes in *n* trials is $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- Notation: $X \sim Binomial(n, p)$



https://d2l.ai/chapter_appendix-mathematics-for-deep-learning/distributions.html

Binomial Distribution

Example:

If you roll a die 16 times,

- What is the probability that a five comes up three times?

$p(x = 3)$

```
from scipy.stats import binom
binom.pmf(3,16,1/6)
```

```
0.24231376033713137
```

- What is the probability that a five comes up at least three times?

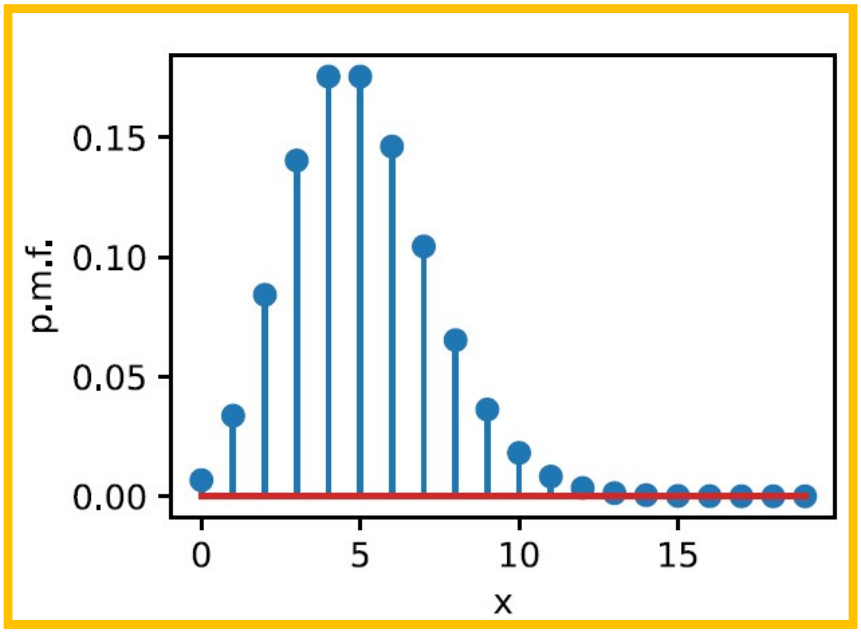
$p(x \geq 3) = 1 - p(x \leq 2)$

```
1 - binom.cdf(2, 16, 1/16)
```

```
0.07420726082533868
```

Poisson Distribution

- A number of events occurring independently in a fixed interval of time with a known rate λ
- A discrete random variable X with states $\{0, 1, 2, \dots\}$ has probability $P(X = k) = \frac{\lambda^X e^{-\lambda}}{X!}$
- The rate λ is the average number of occurrences of the event
- Notation: $X \sim Poisson(\lambda)$



https://d2l.ai/chapter_appendix-mathematics-for-deep-learning/distributions.html

Poisson Distribution

Example:

The number of arrivals at a store can be modeled as Poisson process with an average arrival rate of 10 customers per hours. What is the probability in the next 30 minutes that:

a) 3 customers arrive

```
from scipy.stats import poisson

poisson.pmf(3, 5)
```

```
0.1403738958142805
```

b) More than 4 customers arrive

```
1- poisson.cdf(4, 5)
```

```
0.5595067149347874
```

c) Less than 6 customers arrive

```
poisson.cdf(5, 5)
```

```
0.615960654833063
```

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Back-Up

Resources

- Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for Machine Learning.
- Modern Mathematical Statistics with Applications Second Edition by Jay L. Devore and Kenneth N. Berk
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, Aurelien Geron, 2019
- Introduction to Machine Learning (2014) by Ethem Alpayadin

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC