Week 6 (Tue):

e-Business

Semester 2, 2025

- **Today's Goal**
  - **Basics for Machine Learning**
  - **Google Colab Practice II**

A Case of Business Analytics

- The Case of Google
  - How do you think? Do you feel any difference in your life?

Our approach to AI is grounded in these three principles:

1. Bold innovation

We develop AI that assists, empowers, and inspires people in almost every field of human endeavor; drives economic progress; and improves lives, enables scientific breakthroughs, and helps address humanity's biggest challenges.

2. Responsible development and deployment

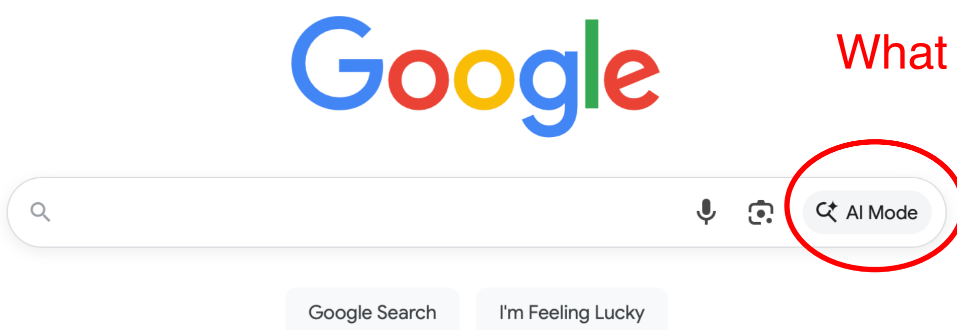
Because we understand that AI, as a still-emerging transformative technology, poses evolving complexities and risks, we pursue AI responsibly throughout the AI development and deployment lifecycle, from design to testing to deployment to iteration, learning as AI advances and uses evolve.

3. Collaborative progress, together

We make tools that empower others to harness AI for individual and collective benefit.

A Case of Business Analytics

- At the core of Google's business model lies its precise and globally scalable Value Proposition—highly targeted, text-based advertising.
- Through AdWords (now known as Google Ads), Google enables advertisers to display sponsored ads on search results pages and across a broad content network.
- These ads are contextually matched to user search queries, enhancing relevance and maximizing the click-through rate (CTR).



What is the role of AI?

?

A Case of Business Analytics

- Google's business model is highly attractive to advertisers, as it allows them to run customized ad campaigns based on keywords, interests, and demographics.
- Mechanism
  - The success of this model depends on a single critical factor, which is user traffic. The more users Google attracts to its search engine, the greater the number of ad impressions, and the more valuable the platform becomes for advertisers.
- Consumer relationship
  - To attract and retain users, Google has continuously strengthened its ecosystem through free, high-value services such as Gmail, Google Maps, and Picasa. This strategy not only boosts user engagement but also increases the volume of searchable queries, thereby expanding advertising opportunities.
- Provider relationship
  - To extend its advertising reach beyond its own platform, Google introduced the AdSense program. This initiative allows third-party websites to host Google ads and earn a share of the revenue. By using advanced algorithms to analyze site content and display relevant ads, AdSense created a new customer segment, content owners.

What are your thoughts on the role of business analytics using AI in this slide?

A Case of Business Analytics

- Google operates as a classic multi-sided platform, generating revenue from advertisers while offering free, value-rich services to web users and content publishers.
- This structure creates a reinforcing feedback loop in network effects:

  i) As the number of users increases → ad views rise.
  ii) As ad views increase → revenue from advertisers grows.
  iii) As advertising revenue grows → more content partners join AdSense.
  iv) As content expands → relevance and reach improve, which in turn attracts even more users.

- Google's revenue model operates on an auction-based system, where advertisers bid on keywords through the AdWords platform. The more competitive a keyword is, the higher the cost per click (CPC) becomes.

Then, what will be the role of business analytics using AI in this mechanism?

A Case of Business Analytics

- Google's Key Resource is its advanced search platform, which powers three major services:

  Web Search — Google.com
  AdWords — Advertising services
  AdSense — Content monetization services

- These services are built upon proprietary algorithms and a sophisticated IT infrastructure.

- Google's Key Activities include:
  Building and maintaining its search and advertising infrastructure
  Managing and continuously improving its core services
  Promoting the platform to attract and retain users, content creators, and advertisers

- Through these activities, Google achieves scalable growth, monetizes user attention, and delivers value across all segments of its platform.

Application of Business Analytics

- What key information would be important for applying a business analytic approach?
  - Put simply, what would present the Y (dependent variable) and X (independent variable)?

  i) As the number of users increases → ad views rise.
  ii) As ad views increase → revenue from advertisers grows.
  iii) As advertising revenue grows → more content partners join AdSense.
  iv) As content expands → relevance and reach improve, which in turn attracts even more users.

$$Y = \boldsymbol{f}(\ X\ )$$

→ AI (or machine learning models)

Apply to the Titanic Case
- What should you know?

```
df.head()
```

	PassengerId	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Application of Business Analytics

- What will be Y?
- What will be X?
- What are the characteristics of Y and X?
- Eventually, what is the role of machine learning in the context of Titanic case?
- How can the Titanic case be extended to the case of Google or others?

$$Y = \boldsymbol{f}(X)$$

AI (or machine learning models)

```
df.head()
```

	PassengerId	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

The Meaning of Machine Learning (ML)

- Machine Learning (ML) refers to algorithms that computers use to learn from data, allowing it to make predictions on future events.
- Early AI efforts did not pay much attention to ML, as computers were neither powerful enough nor had the capability to store the huge amount of data.
- The frame problem: Machine learning is the technique of pattern classification and prediction based on a data. A machine is not able to sort out the type of data.
- The symbol grounding problem: A machine recognize knowledge as a mark. Thus, a machine cannot recognize the difference of objects.
- The problem of feature engineering: An algorithm can only work well on training data that clearly features the value of identities.
- Accuracy rate is not high enough for the prediction model but helps to understand the trend of fitness.

The Meaning of Machine Learning (ML)

- Machine Learning & Deep Learning

	MACHINE LEARNING	DEEP LEARNING
Approach	Requires structure data	Does not require structure data
Human Intervention	Requires human intervention for mistakes	Does not require human intervention for mistakes
Hardware	Can function on CPU	Requires GPU / significant computing power
Time	Takes seconds to hours	Takes weeks
Uses	Forecasting, predicting and other simple applications	More complex applications like autonomous vehicles

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

The university has 5600 students. John's ID is number 1, he is 18 years old and already holds a B.Sc. degree. David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

Hard to predict patterns

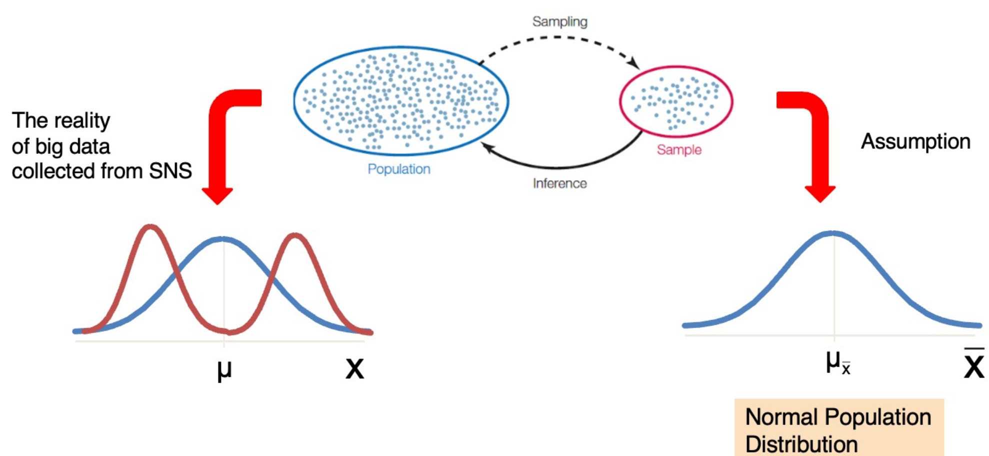
The Characteristics of Variables

```
df.head()
```

	PassengerId	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Google users will show more complex behavioral patterns.

Then, we cannot guarantee the Central Limit Theorem.



- We normally assume Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means are approximately normal as long as the sample size is large enough.
  - Real meaning implies how your data look like, meaning *the pattern of data .*

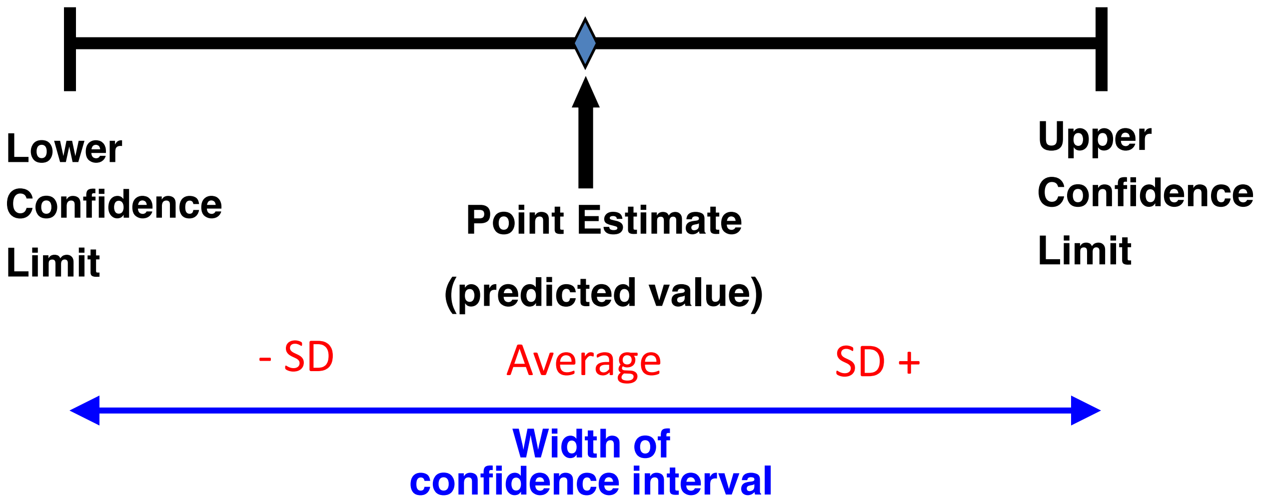
The Characteristics of Variables

- A point estimate is a single number.

Examples: sample mean, sample variance, sample proportion.

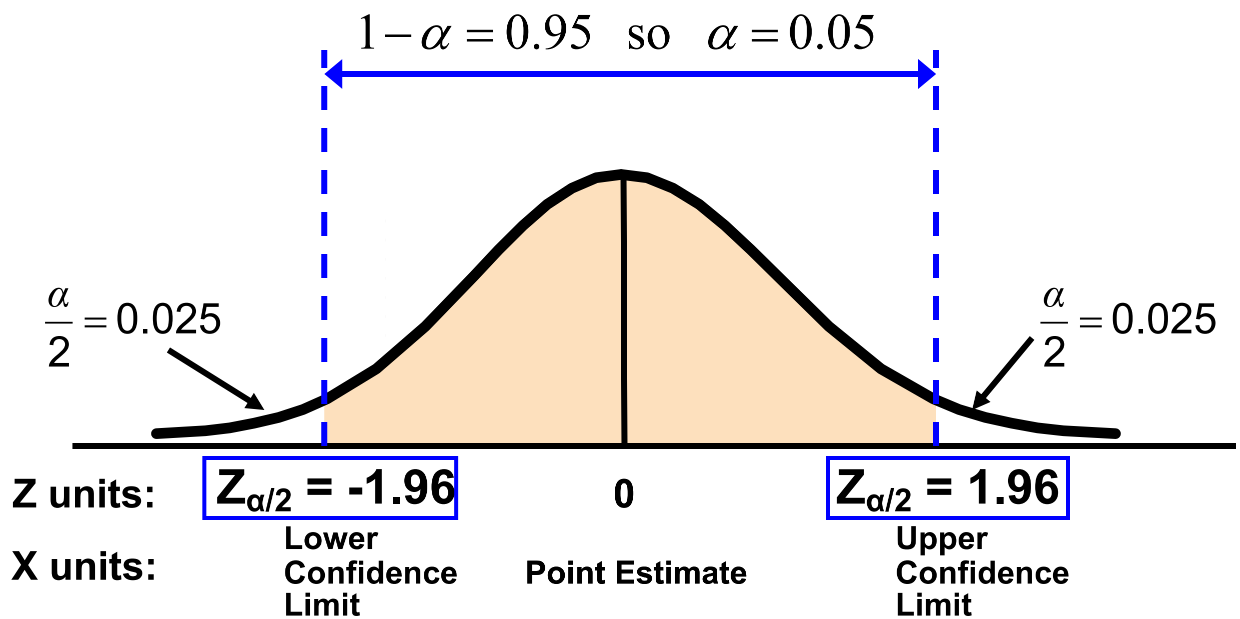
- A confidence interval provides additional information about the variability of the estimate (or predicted value)

- * This means that the predicted value will reflect the the ***average*** trend and associated ***error*** based on the pattern of given data.



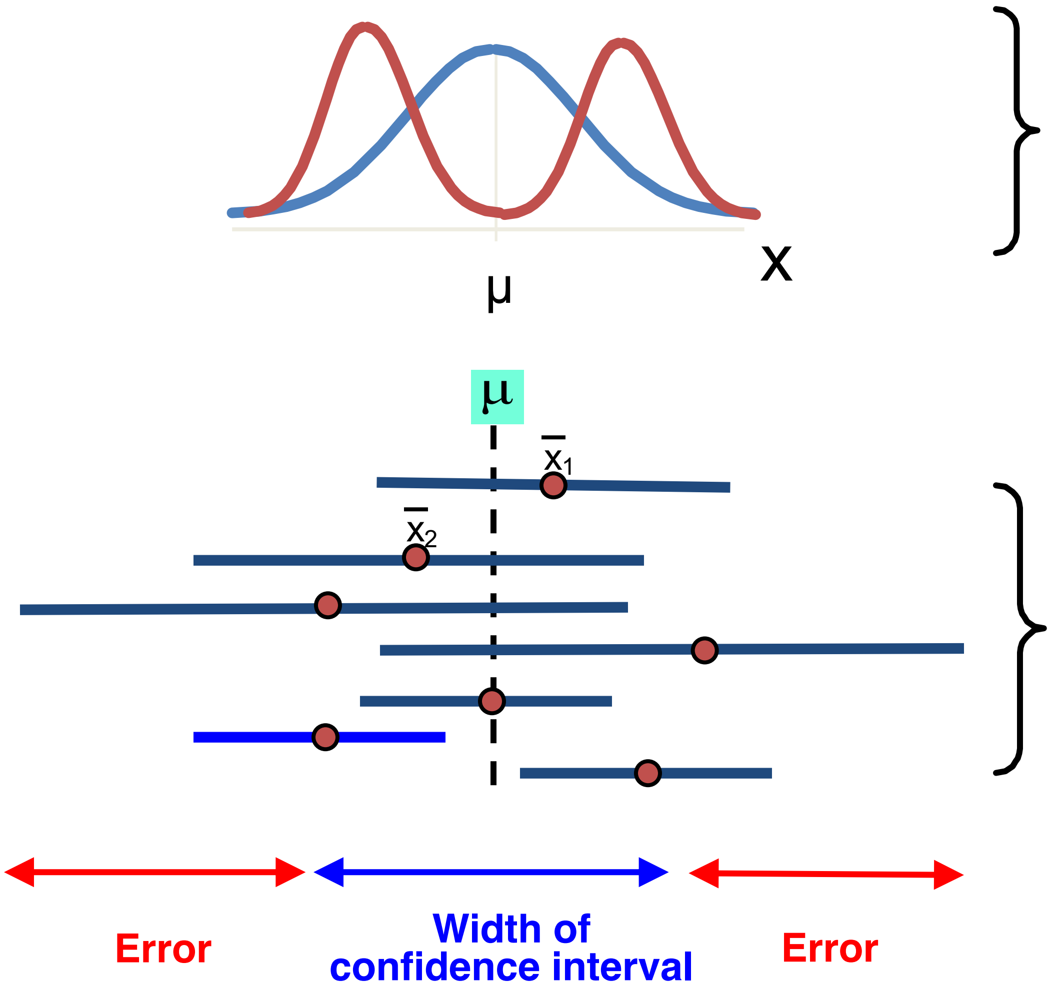
The Characteristics of Variables

- For a 95% confidence interval:



- The confidence coefficient (1-α) is the probability of not rejecting H_0 when it is true.
- * A 5% significance level implies that there is a 5% probability of observing the result due to random error – that is, 5 out of 100 predictions may be inaccurate by chance.

The Characteristics of Variables



The Distribution of
your Big Data

The pattern of
confidence interval
in your Big Data

* Meaning that
It is hard to predict the pattern.

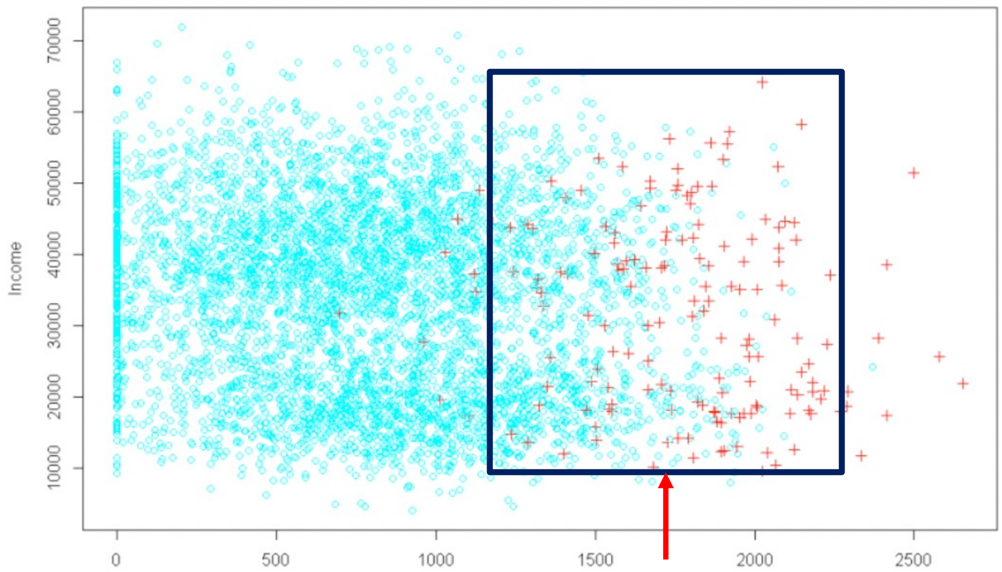
The Characteristics of Variables

- Big data example

  - Some customers prefer iced Americanos, while others enjoy hot lattes.

  - Suppose a bank wants to promote a Starbucks coupon to these customers.

  - However, customer preferences are not always consistent – people don't always choose the same type of coffee every time.



How can you categorize these customers?

The Characteristics of Variables

- Language data example

- Let’s suppose that we have a text prediction problem in sequence.

“ I was born in Paris. But I grew up in Tokyo. Therefore, I speak fluent.”

a. We process the words as a sequence, and the sequence of words captures the information required to make the prediction.

b. We have an innate ability for this, and it is trivial for us to know that the fluent language can be Japanese.

c. The language that we speak is sequential in nature, and the sequence conveys context and other subtle nuances.

- Take for example the following words.

“ is, am, I, on, the, building, today, on, fire.

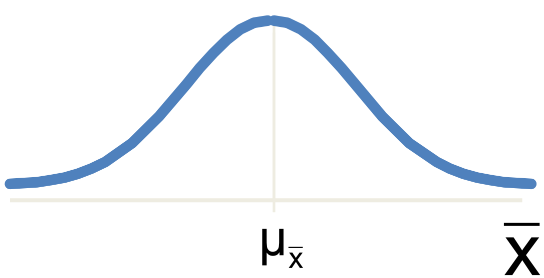
The Characteristics of Variables

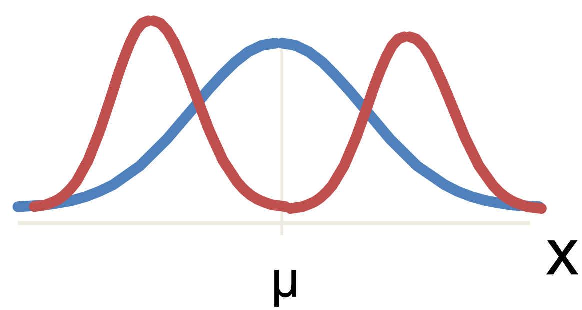
Golden Retriever?

No
(different color & size)

Golden Retriever?

No
(different size & ear)



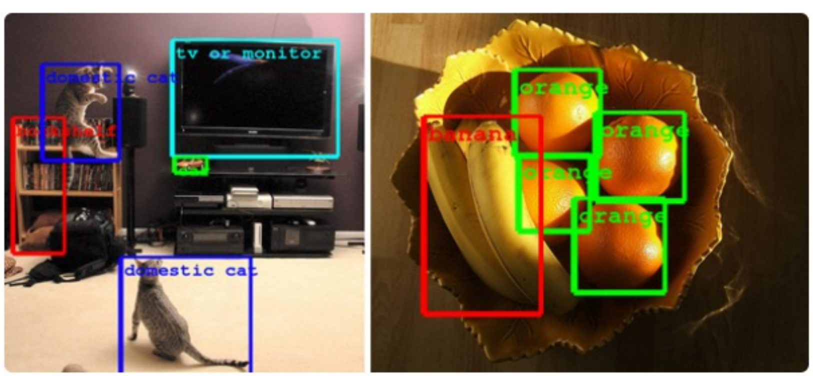


The Characteristics of Machine Learning

- The recent success examples for machine learning are as follows.
  - identification of spam emails
  - Segmentation of customer behavior for targeted advertising
  - Development of algorithms for auto-piloting drones and cars
  - Image and voice recognition

  ML

- Data mining, a stream of machine learning methods, is concerned with the generation of novel insight from big data.



The Characteristics of Machine Learning

- The recent success examples for machine learning are as follows.

  - identification of spam emails
  - Segmentation of customer behavior for targeted advertising
  - Development of algorithms for auto-piloting drones and cars
  - Image and voice recognition

ML

- Classification

  - A supervised learning task where the goal is to assign data points to predefined categories based on the input features.

- Clustering

  - An unsupervised learning task that groups data points into clusters based on their similarities – without using predefined labels.

The Characteristics of Machine Learning

- Eventually, what is the role of machine learning in the context of Titanic case?
- How can the Titanic case be extended to the case of Google or others?

$$Y = \boldsymbol{f}(X)$$

AI (or machine learning models)

```
df.head()
```

	PassengerId	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

- What do you think?

- The question asks whether a passenger is predicted to survive or not, based on the features (or variables such as age, sibsp, or parch) available in the dataset.

- Google case: Whether a customer is likely to respond positively to an ad or not

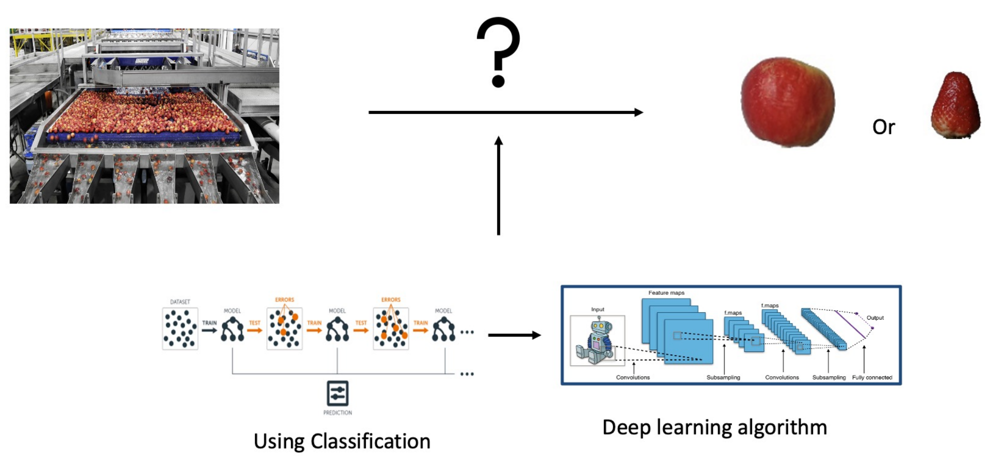
Machine Learning Model I - Classification

- Classification (or Classifier)
  - A method used to predict the categorical outcome Y based on a set of input features X, that is, to model and solve the relationship.

$$Y = \mathbf{f}(X)$$

→ A machine learning model called as classifier or classification

- Advanced: Deep Learning Model
  - Example: Can you predict if a fruit picture presents an apple or not?



Machine Learning Model I - Classification

- **Decision Tree** (Classification Tree, Regression Tree)

  Decision support tool that uses a tree-like graph of decisions and their possible consequences (Source: Wikipedia).

- Supervised data mining for classification & prediction

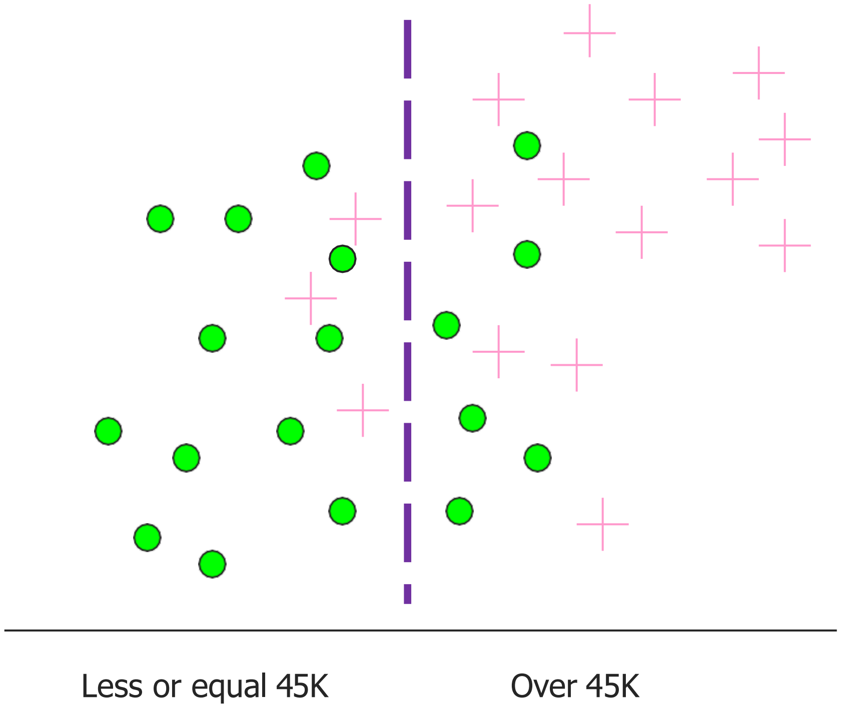
- The decision tree can be expressed

  by "**IF** X_1=… and X_2 =… and X_k =…, **THEN** Y = … (One of Predefined Groups)"

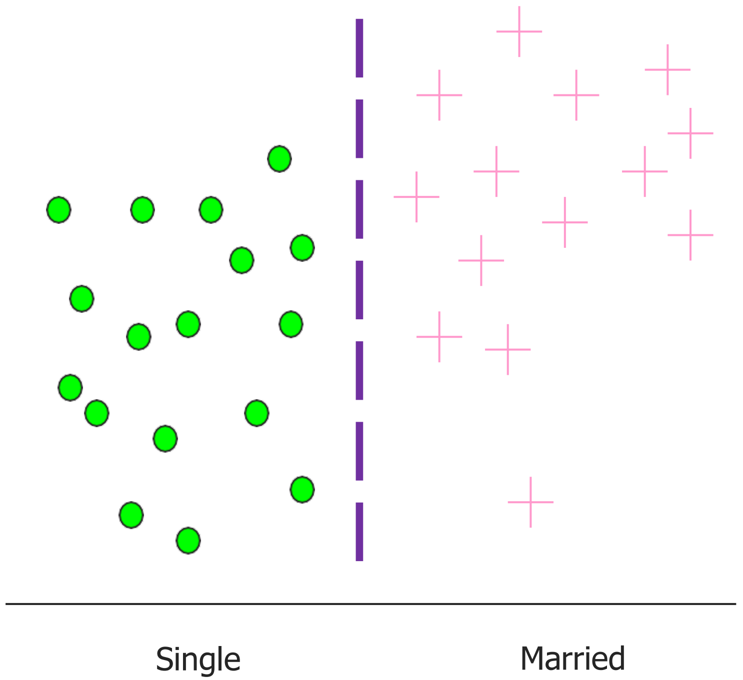
Machine Learning Model I - Classification

- How can we specify our prediction modeling to answer those questions in the case of machine learning (or data mining)?

Split over whether income exceeds 45K



Split over whether an applicant is married



Machine Learning Model I - Classification

Trace each path from the root node to a leaf node to generate a rule:

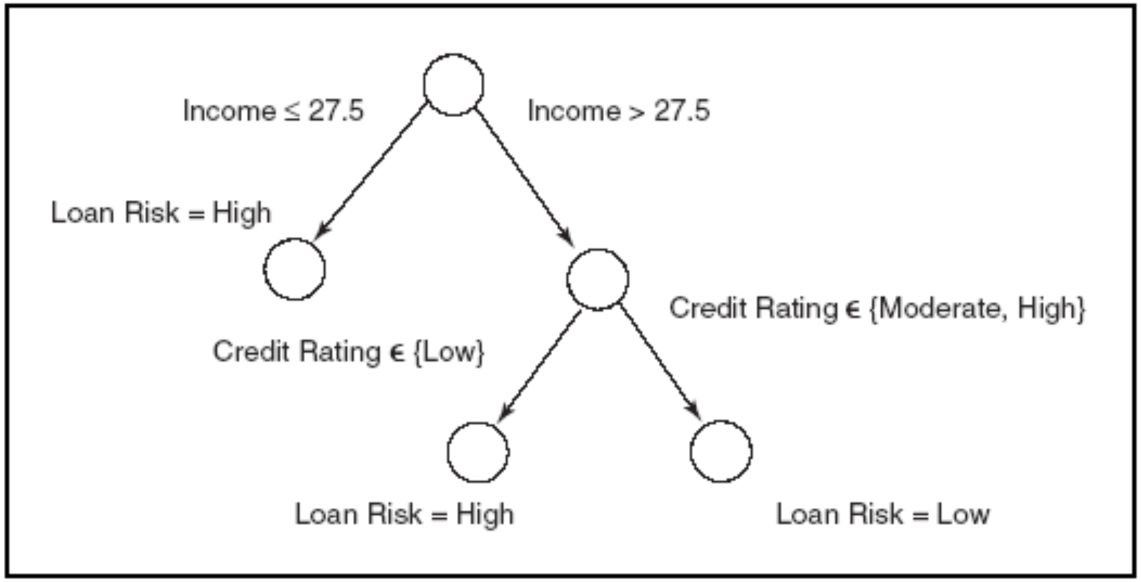


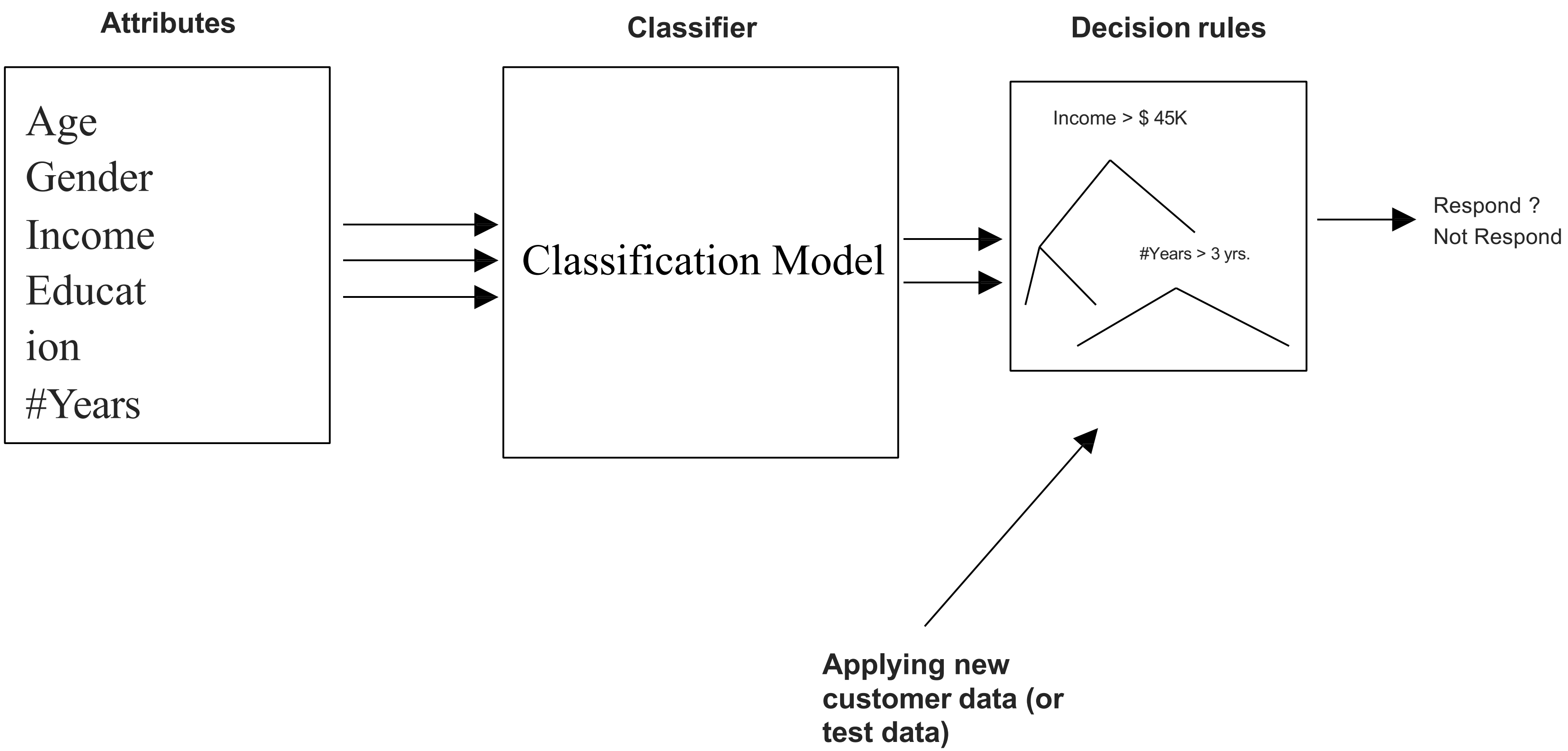
FIGURE 7.1 Decision Tree Using Gini Index for Split Criteria

If Income ≤ 27.5, then Loan-Risk = High

Else if Income > 27.5 and Credit-Rating = Low, then Loan-Risk = High

Else if Income > 27.5 and Credit-Rating = Moderate or High, then Loan-Risk = Low

Machine Learning Model I - Classification



Machine Learning Model I - Classification

- In the case of Bank, classification of data into two predefined groups:
  - Reject Class (Success=N): Customers with trouble paying back the loan
  - Accept Class (Success=Y): Customers with NO trouble paying back the loan

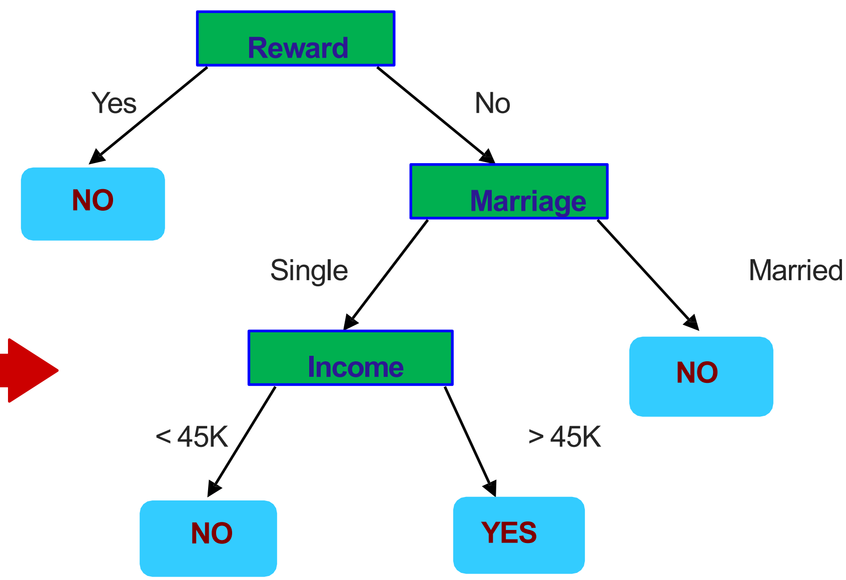
* Success = the outcome variable

Machine Learning Model I - Classification

In the case of Bank: Customer defection

categorical categorical continuous class

ID	Reward Use	Marital Status	Taxable Income	Defection
1	Yes	Married	120K	**No**
2	No	Married	110K	**No**
3	Yes	Single	90K	**No**
4	Yes	Married	130K	**No**
5	No	Single	45K	**Yes**
6	Yes	Married	50K	**No**
7	Yes	Single	220K	**No**
8	No	Single	75K	**Yes**
9	No	Single	25K	**No**



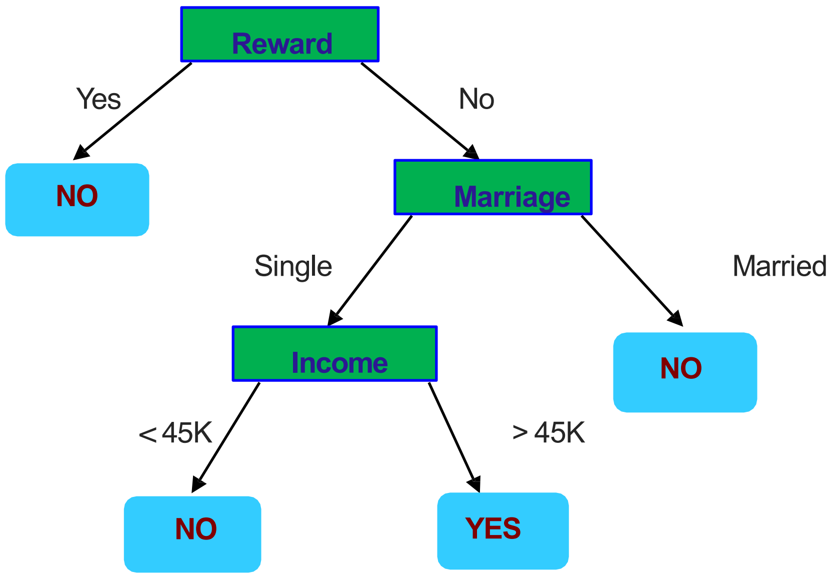
Model: Decision Tree

Training Data

Training Data: The set of all instances used for learning the model is called training set.

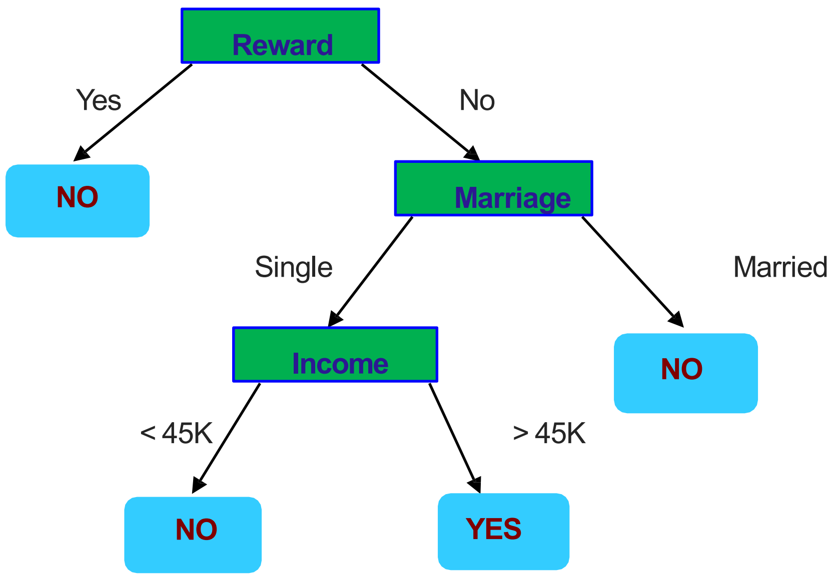
Machine Learning Model I - Classification

- Root Node
  - Has no incoming edges
  - Has many outgoing edges
- Internal Nodes
  - Has exactly one incoming edge and two or more outgoing edges
  - Contain conditions to separate records
- Leaf (terminal) Nodes
  - Has exactly one incoming edge and no outgoing edge
  - Each leaf node is assigned a class label



Model: Decision Tree

Machine Learning Model I - Classification

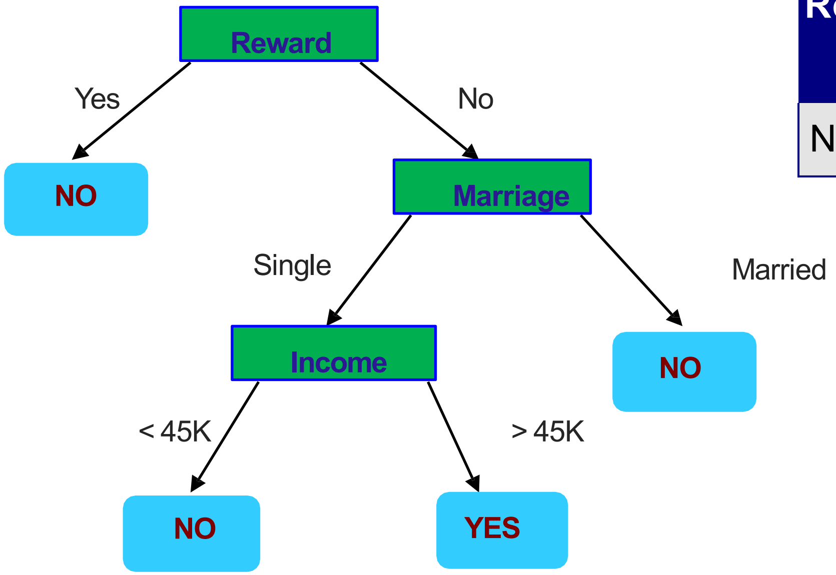


Model: Decision Tree

- If (*Reward* = Yes) Then No
- If (*Reward* = No) And (*Marriage* = Single) And (*Income* < 45K)

  Then Class = No
- Any other rules?

Machine Learning Model I - Classification



Model: Decision Tree

Input New Data

Reward	Marital Status	Taxable Income	Defection
No	Married	80K	?

The Principle of Classification

- ✓ How to choose splitting attributes/variables
- ✓ How to decide the order of variables for splitting
- ✓ How to prune the tree
- ✓ When to stop

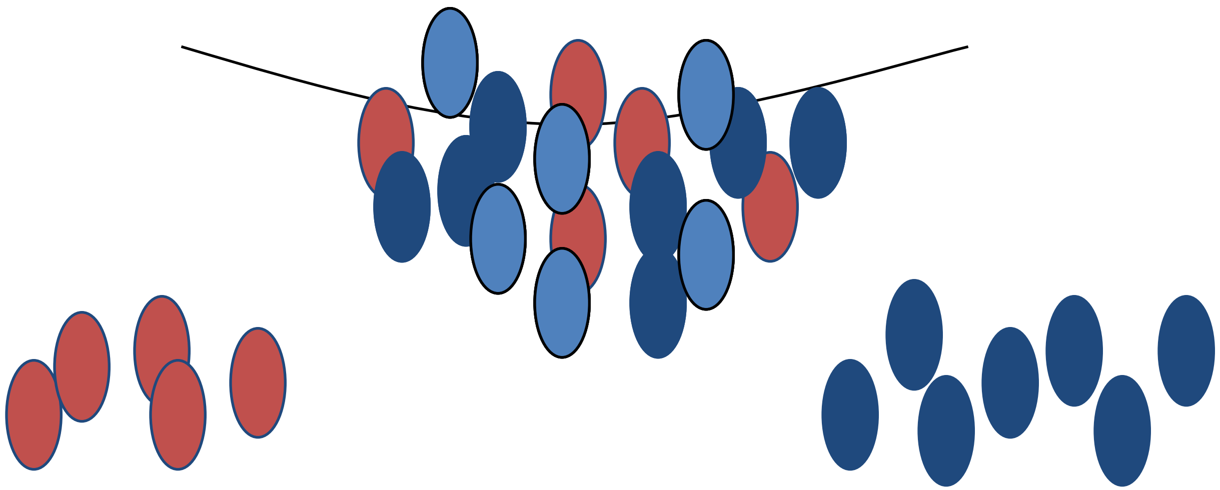
The Principle of Classification

- We want to find the most “useful” attribute in classifying a sample.

- Purity: the degree to which a subset of instances contains only a single class
  - If every instance of a node has the same value (class) for the target, then the node is pure
- Need a measure of node impurity
  - Gini Index
  - Entropy

The Principle of Classification

Decision Tree Induction is often based on Information Theory



- When all the marbles in the bowl are mixed up, uncertainty is high.
- When the marbles in the same bowl are same color, no uncertainty.

→ ***Split data in a way to reduce uncertainty!***

The Principle of Classification

- Gini Index for a given node:

$$\boxed{GINI = 1 - \sum_j [p(j)]^2}$$

  - *p(j)* is the relative frequency of class *j*
    - The proportion of instances into class j

The Principle of Classification

Gender	Car ownership	IncomeLevel	Transportation mode
Female	0	Low	Bus
Male	0	Low	Bus
Male	1	Medium	Bus
Male	1	Medium	Bus
Female	1	Medium	Train

Gender	Transportation mode
Female	Bus
Male	Bus
Male	Bus
Male	Bus
Female	Train

Car ownership	Transportation mode
0	Bus
0	Bus
1	Bus
1	Bus
1	Train

IncomeLevel	Transportation mode
Low	Bus
Low	Bus
Medium	Bus
Medium	Bus
Medium	Train

The Principle of Classification

$$p(j) = p(Bus) = 4/5 = 0.8$$

$$p(j) = p(Train) = 1/5 = 0.2$$

$$Gini = 1 - 0.8^2 - 0.2^2 = 0.32$$

Gender	Transportation mode
Female	Bus
Male	Bus
Male	Bus
Male	Bus
Female	Train

The Principle of Classification

- When a node p is split into k partitions (children), the quality of split is computed as,

$$\boxed{GINI_p = \sum_{i=1}^{k} \frac{n_i}{n} GINI(i)}$$

  - n_i = number of records at child i
  - n = number of records at node p

The Principle of Classification

- Female

$$p(j\,|\,t) = p(Bus\,/\,Female) = 1/2 = 0.5$$

$$p(j\,|\,t) = p(Train\,/\,Female) = 1/2 = 0.5$$

$$Gini(Female) = 1 - 0.5^2 - 0.5^2 = 0.5$$

- Male

$$p(j\,|\,t) = p(Bus\,/\,Male) = 3/3 = 1$$

$$p(j\,|\,t) = p(Train\,/\,Male) = 0/3 = 0$$

$$Gini(Male) = 1 - 1^2 - 0^2 = 0$$

Gender	Transportation mode
Female	Bus
Male	Bus
Male	Bus
Male	Bus
Female	Train

The Principle of Classification

$$GINI_p = \sum_{i=1}^{k} \frac{n_i}{n} GINI(i)$$

- Gini(Gender)
  - number of records at child female: 2
  - number of records at child male: 3

Gender	Transportation mode
Female	Bus
Male	Bus
Male	Bus
Male	Bus
Female	Train

$$Gini(Gender) = 2/5 * Gini(Female) + 3/5 * Gini(Male)$$

$$Gini(Gender) = 2/5 * 0.5 + 3/5 * 0 = 0.2$$

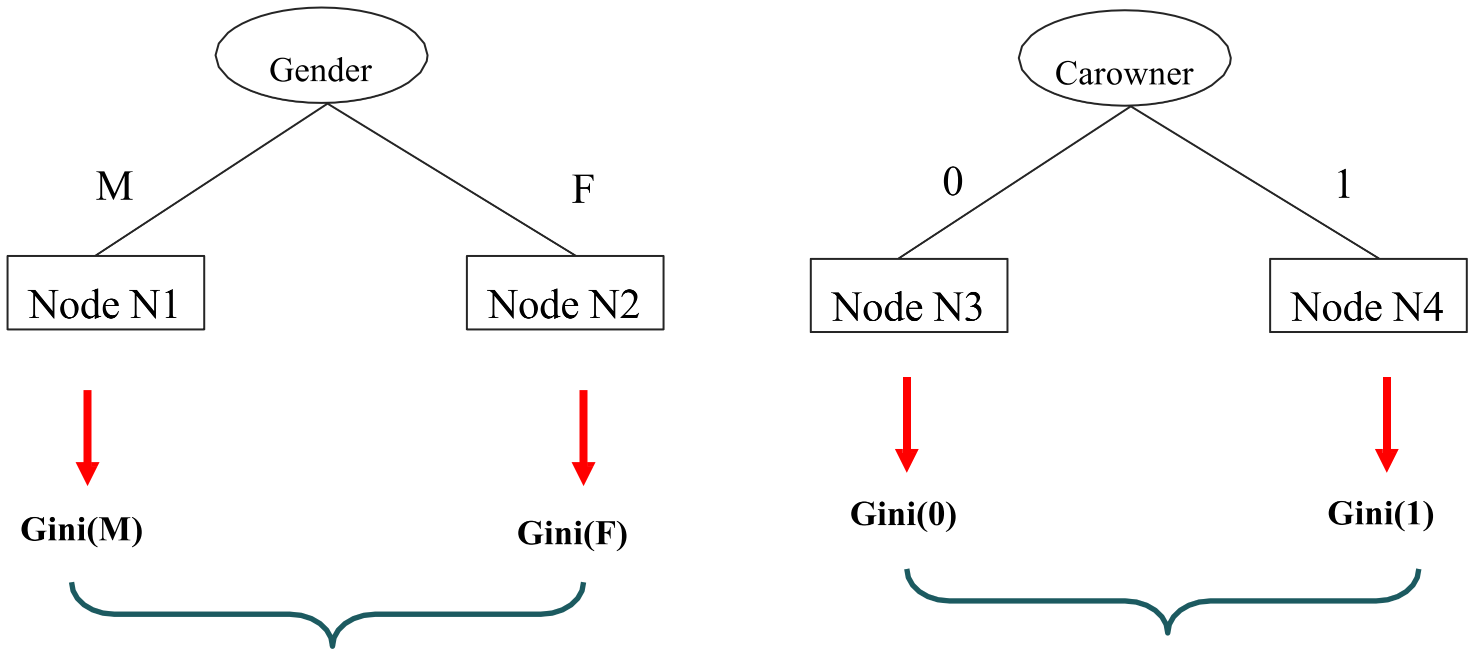
$$Gini(Car\ ownership) = 0.27$$
$$Gini(IncomeLevel) = 0.27$$

The Principle of Classification

- Compare the degree of impurity of the parent node (before splitting) with the degree of impurity of the child nodes (after splitting).
- The larger their difference, the better the test condition.
- Decision Tree algorithms often choose an attribute which maximizes the gain

The Principle of Classification

Before Splitting: Bus / Train → **Gini(Before)** = 0.32



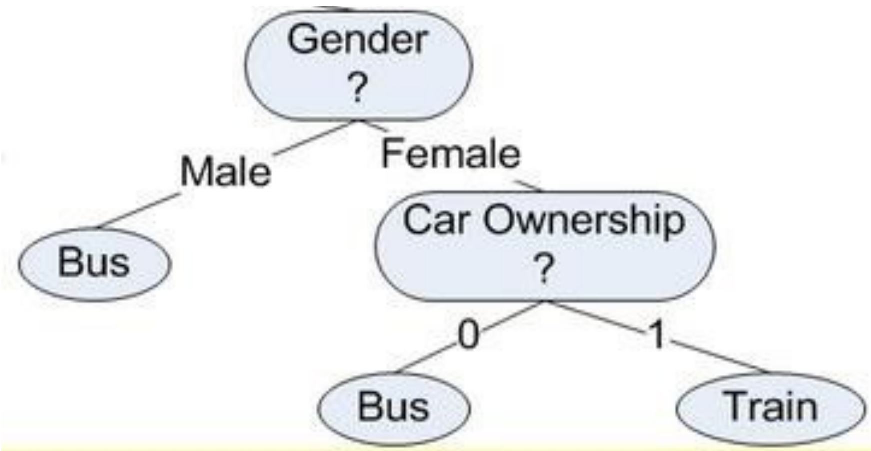
Gini(Gender) = 0.20

Gini(Carowner) = 0.27

Gain = Gini (Before) – Gini(Gender) vs. Gain = Gini(Before)– Gini(Carowner)

0.12 = (0.32 – 0.20) > 0.05 = (0.32 – 0.27)

The Principle of Classification



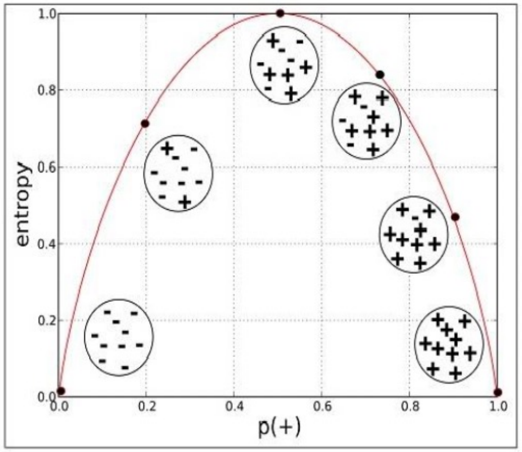
Gender	Car ownership	IncomeLevel	Transportation mode
Female	0	Low	Bus
Male	0	Low	Bus
Male	1	Medium	Bus
Male	1	Medium	Bus
Female	1	Medium	Train

The Principle of Classification

- Entropy at a given node t

$$\boxed{Entropy(t) = -\sum_j p(j \mid t) \log p(j \mid t)}$$

  - $p(j \mid t)$ is the relative frequency of class j at node t



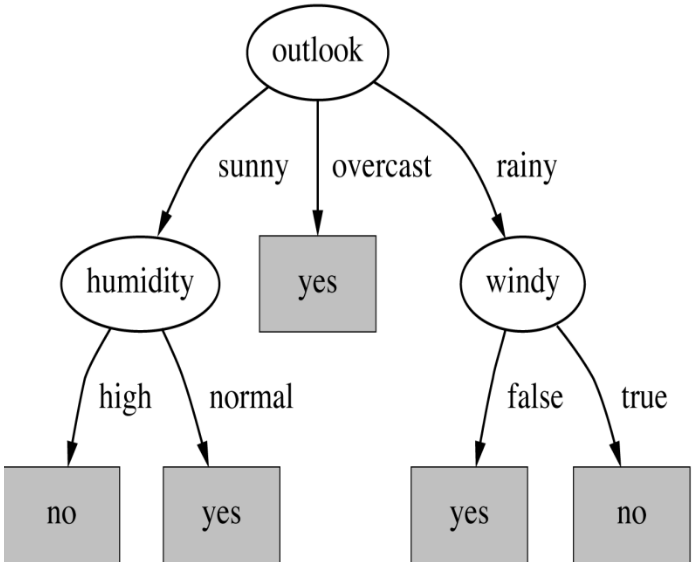
- Information Gain:

$$\boxed{GAIN_{split} = Entropy(p) - \left(\sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \right)}$$

  - Measures Reduction in Entropy achieved because of the split. Choose the split that achieves most reduction (maximizes GAIN)

The Principle of Classification

outlook	temp	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no



The Principle of Classification

- Model construction from weather data
  - If we choose outlook as an attribute to split on,

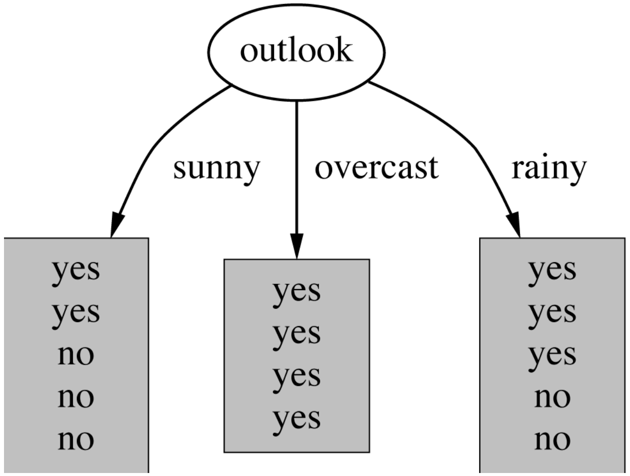
    the average $-\sum(p_i)\log(p_i)$ value is

$$5/14 * \{-(2/5)\log(2/5)-(3/5)\log(3/5)\}$$

$$+ 4/14 * \{-1\log(1)\}$$

$$+ 5/14 * \{-(3/5)\log(3/5)-(2/5)\log(2/5)\}$$

$$= 0.693$$



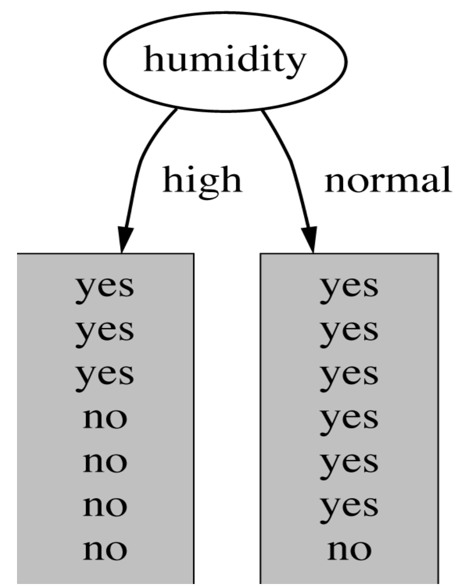
The Principle of Classification

- Model construction from weather data
  - If we choose humidity as an attribute to split on, the average $-\sum(p_i)\log(p_i)$ value is

$$7/14 * \{-(3/7)\log(3/7)-(4/7)\log(4/7)\}$$

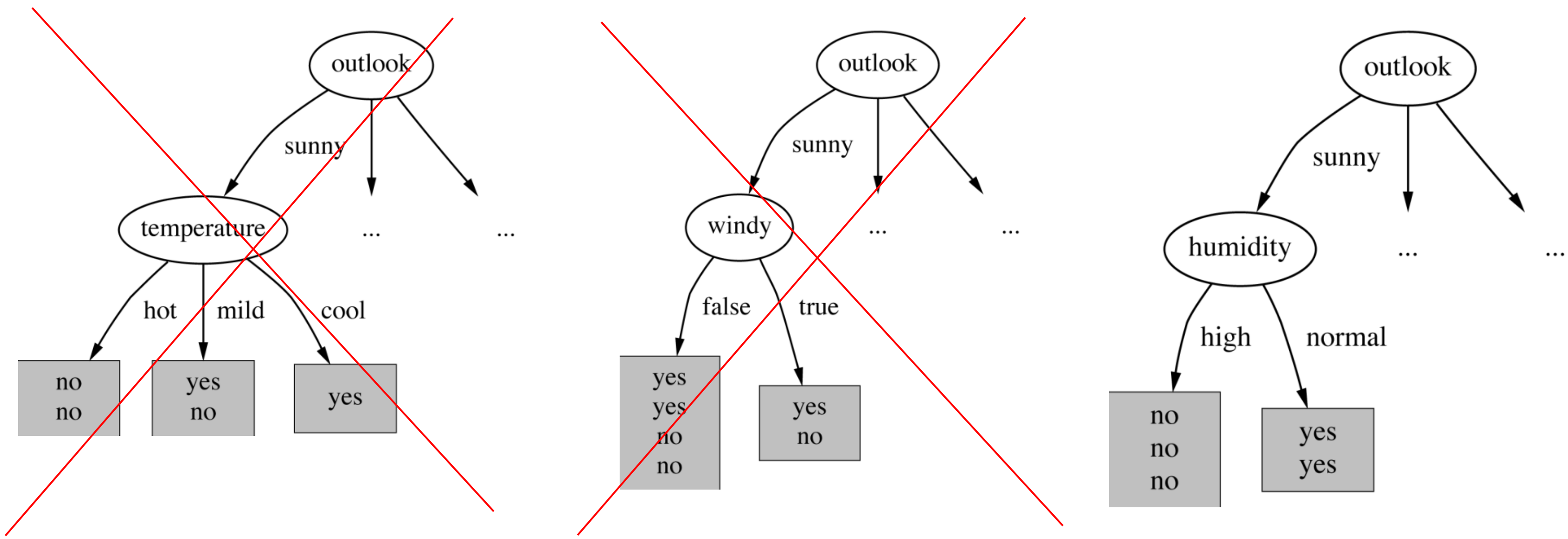
$$+ 7/14 * \{-(6/7)\log(6/7)-(1/7)\log(1/7)\}$$

$$= 0.788$$



The Principle of Classification

- Model construction from weather data
  - So, we finally choose outlook as the attribute to split on.
  - Continue the same process to get these:



Homework I (Due Date: 1pm Oct 21)

- Deliverables - A single Jupyter Notebook (.ipynb) with:

  - Name and ID

  - All code cells executed

  - Plots displayed

  - Try to answer to all questions in markdown cells

  (If you are unsure of an answer, leave the section blank.)

- Questions

  - Question 1: How many rows and columns does the dataset have?

  - Question 2: Which variables are numeric? Which are categorical?

  - Question 3: What is the mean and median of Age and Fare?

  - Question 4: Which passenger class (Pclass) had the most travelers?

  - Question 5: How did removing outliers affect the shape of Fare's distribution?

  - Question 6: Describe the shape of the age distribution.

  - Question 7: How do survival rates differ by passenger class?