Unit 1: Machine Learning
TCS 509

Best wishes to all readers ☺

Professor Jay Bhatnagar

mean
median
mode
range
outlier

7-2 Mean, Median, Mode, and Range

Helpful Hint

The mean is sometimes called the average.

The **mean** is the sum of the data values divided by the number of data items.

The **median** is the middle value of an odd number of data items arranged in order. For an even number of data items, the median is the average of the two middle values.

The **mode** is the value or values that occur most often. When all the data values occur the same number of times, there is no mode.

The **range** of a set of data is the difference between the greatest and least values. It is used to show the spread of the data in a data set.

Outlier

In Statistics, an outlier are data points that differ significantly from other observations (neighborhoods – DBSCAN density based spatial clustering of apps under noise, normalized deviation - z –score). An outlier may be due to variability in the measurement, an indication of novel data, or it may be the result of experimental error; the latter are sometimes excluded from the data set. An outlier can be an indication of exciting possibility but can also cause serious problems in statistical analyses.

Additional Example 1: Finding the Mean, Median, Mode, and Range of Data

Find the mean, median, mode, and range of the data set.

4, 7, 8, 2, 1, 2, 4, 2

mean:

$$\underbrace{4 + 7 + 8 + 2 + 1 + 2 + 4 + 2}_{\text{8 items}} = \underbrace{30}_{\text{sum}}$$

Add the values.

$30 \div 8 = 3.75$

Divide the sum by the number of items.

The mean is 3.75.

Additional Example 1 Continued

Find the mean, median, mode, and range of the data set.

4, 7, 8, 2, 1, 2, 4, 2

median:

~~1, 2, 2,~~ 2, 4, ~~4, 7, 8~~

Arrange the values in order.

2 + 4 = 6

6 ÷ 2 = 3

There are two middle values, so find the mean of these two values.

The median is 3.

Additional Example 1 Continued

Find the mean, median, mode, and range of the data set.

4, 7, 8, 2, 1, 2, 4, 2

mode:

1, 2, 2, 2, 4, 4, 7, 8 *The value 2 occurs three times.*

The mode is 2.

Check It Out: Example 1

Find the mean, median, mode, and range of the data set.

6, 4, 3, 5, 2, 5, 1, 8

mean:

$$\underbrace{6 + 4 + 3 + 5 + 2 + 5 + 1 + 8}_{\text{8 items}} = \underbrace{34}_{\text{sum}}$$

Add the values.

$$34 \div 8 = 4.25$$

Divide the sum by the number of items.

The mean is 4.25.

Check It Out: Example 1 Continued

Find the mean, median, mode, and range of the data set.

6, 4, 3, 5, 2, 5, 1, 8

median:

~~1, 2, 3~~, 4, 5, ~~5, 6, 8~~

Arrange the values in order.

4 + 5 = 9

9 ÷ 2 = 4.5

There are two middle values, so find the mean of these two values.

The median is 4.5.

Check It Out: Example 1 Continued

Find the mean, median, mode, and range of the data set.

6, 4, 3, 5, 2, 5, 1, 8

mode:

1, 2, 3, 4, 5, 5, 6, 8 *The value 5 occurs two times.*

The mode is 5.

Check It Out: Example 1 Continued

Find the mean, median, mode, and range of the data set.

6, 4, 3, 5, 2, 5, 1, 8

range:

1, 2, 3, 4, 5, 5, 6, 8

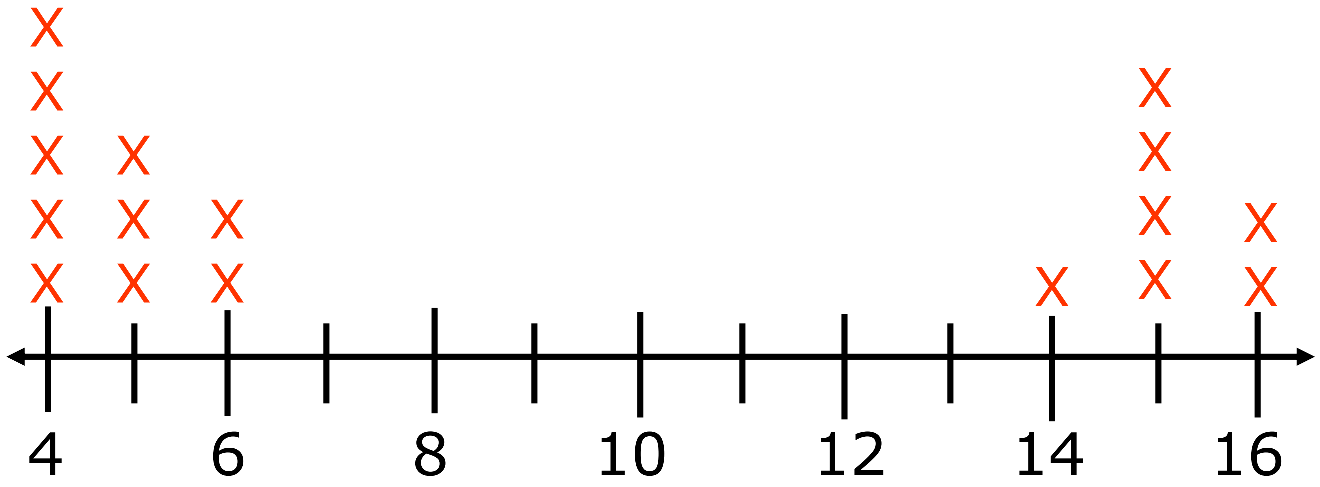
Subtract the least value from the greatest value.

8 – 1 = 7

The range is 7.

Additional Example 2: Choosing the Best Measure to Describe a Set of Data

The line plot shows the number of miles each of the 17 members of the cross-country team ran in a week. Which measure of central tendency best describes this data? Justify your answer.



Additional Example 2 Continued

The line plot shows the number of miles each of the 17 members of the cross-country team ran in a week. Which measure of central tendency best describes this data? Justify your answer.

mean:

$$\frac{4 + 4 + 4 + 4 + 4 + 5 + 5 + 5 + 6 + 6 + 14 + 15 + 15 + 15 + 15 + 16 + 16}{17}$$

$$= \frac{153}{17} = 9$$

The mean is 9. The mean best describes the data set because the data is clustered fairly evenly about two areas.

Additional Example 1 Continued

Find the mean, median, mode, and range of the data set.

4, 7, 8, 2, 1, 2, 4, 2

range:

1, 2, 2, 2, 4, 4, 7, 8

Subtract the least value from the greatest value.

8 – 1 = 7

The range is 7.

Additional Example 2 Continued

The line plot shows the number of miles each of the 17 members of the cross-country team ran in a week. Which measure of central tendency best describes this data? Justify your answer.

median:

4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 14, 15, 15, 15, 15, 16, 16

The median is 6. The median does not best describe the data set because many values are not clustered around the data value 6.

Additional Example 2 Continued

The line plot shows the number of miles each of the 17 members of the cross-country team ran in a week. Which measure of central tendency best describes this data? Justify your answer.

mode:

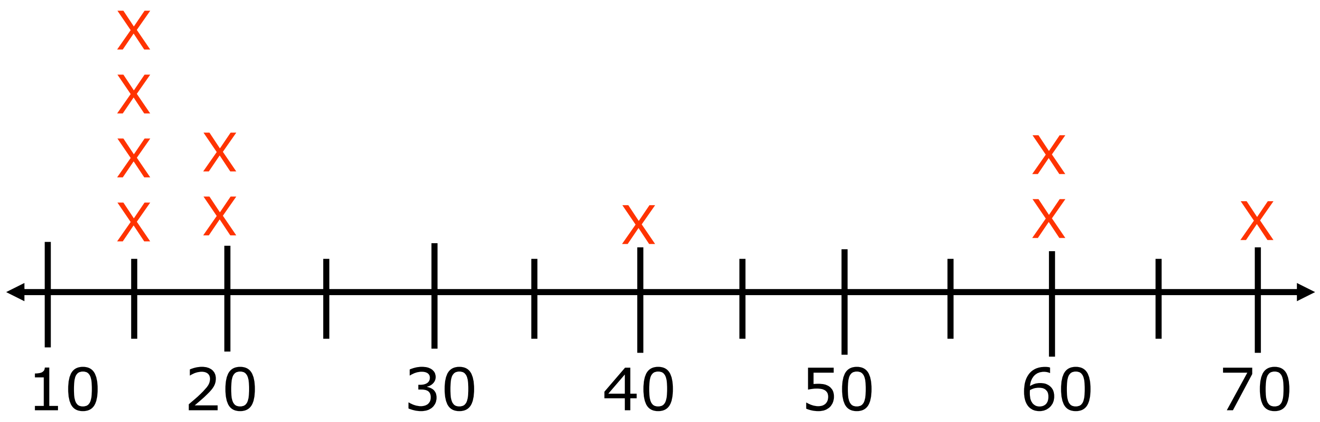
The greatest number of X's occur above the number 4 on the line plot.

The mode is 4.

The mode focuses on one data value and does not describe the data set.

Check It Out: Example 2

The line plot shows the number of dollars each of the 10 members of the cheerleading team raised in a week. Which measure of central tendency best describes this data? Justify your answer.



Check It Out: Example 2 Continued

The line plot shows the number of dollars each of the 10 members of the cheerleading team raised in a week. Which measure of central tendency best describes this data? Justify your answer.

mean:

$$\frac{15 + 15 + 15 + 15 + 20 + 20 + 40 + 60 + 60 + 70}{10}$$

$$= \frac{330}{10} = 33$$

The mean is 33. Most of the cheerleaders raised less than $33, so the mean does not describe the data set best.

Check It Out: Example 2 Continued

The line plot shows the number of dollars each of the 10 members of the cheerleading team raised in a week. Which measure of central tendency best describes this data? Justify your answer.

median:

15, 15, 15, 15, 20, 20, 40, 60, 60, 70

The median is 20. The median best describes the data set because it is closest to the amount most cheerleaders raised.

Check It Out: Example 2 Continued

The line plot shows the number of dollars each of the 10 members of the cheerleading team raised in a week. Which measure of central tendency best describes this data? Justify your answer.

mode:

The greatest number of X's occur above the number 15 on the line plot.

The mode is 15.

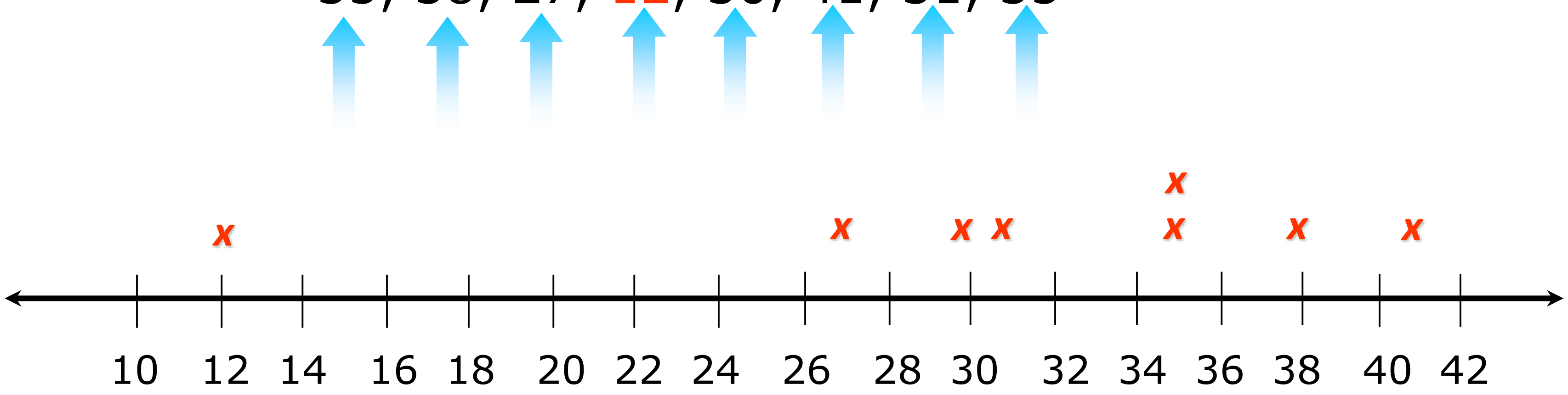
The mode focuses on one data value and does not describe the data set.

Measure	Most Useful When
mean	The data are spread fairly evenly
median	The data set has an outlier
mode	The data involve a subject in which many data points of one value are important, such as election results.

7-2 Mean, Median, Mode, and Range

In the data set below, the value 12 is much less than the other values in the set. An extreme value such as this is called an <u>outlier</u>.

35, 38, 27, 12, 30, 41, 31, 35



Additional Example 3: Exploring the Effects of Outliers on Measures of Central Tendency

The data shows Sara's scores for the last 5 math tests: 88, 90, 55, 94, and 89. Identify the outlier in the data set. Then determine how the outlier affects the mean, median, and mode of the data. Then tell which measure of central tendency best describes the data with the outlier.

55, 88, 89, 90, 94

outlier → 55

Additional Example 3 Continued

With the Outlier

55, 88, 89, 90, 94

outlier → 55

mean:	median:	mode:
55+88+89+90+94 = 416	55, 88, 89, 90, 94	
416 ÷ 5 = 83.2		
The mean is 83.2.	The median is 89.	There is no mode.

Additional Example 3 Continued

Without the Outlier

$\cancel{55}$, 88, 89, 90, 94

mean:	median:	mode:
88+89+90+94 = 361 361 ÷ 4 = 90.25 The mean is 90.25.	88, 89, 90, 94 $\frac{89+90}{2}$ = 89.5 The median is 89.5.	There is no mode.

Since all the data values occur the same number of times, the set has no mode.

Additional Example 3 Continued

	Without the Outlier	With the Outlier
mean	90.25	83.2
median	89.5	89
mode	no mode	no mode

Adding the outlier decreased the mean by 7.05 and the median by 0.5.

The mode did not change.

The median best describes the data with the outlier.

Check It Out: Example 3

Identify the outlier in the data set. Then determine how the outlier affects the mean, median, and mode of the data. The tell which measure of central tendency best describes the data with the outlier.

63, 58, 57, 61, 42

42, 57, 58, 61, 63

outlier → 42

Check It Out: Example 3 Continued

With the Outlier

42, 57, 58, 61, 63

outlier → 42

mean:	median:	mode:
42+57+58+61+63 = 281	42, 57, 58, 61, 63	
281 ÷ 5 = 56.2		
The mean is 56.2.	The median is 58.	There is no mode.

Check It Out: Example 3 Continued

Without the Outlier

~~42~~, 57, 58, 61, 63

mean:	median:	mode:
57+58+61+63 = 239 239 ÷ 4 = 59.75 The mean is 59.75.	57, 58, 61, 63 $\frac{58+61}{2}$ = 59.5 The median is 59.5.	There is no mode.

Check It Out: Example 3 Continued

	Without the Outlier	With the Outlier
mean	59.75	56.2
median	59.5	58
mode	no mode	no mode

Adding the outlier decreased the mean by 3.55 and decreased the median by 1.5.

The mode did not change.

The median best describes the data with the outlier.

Lesson Quiz: Part I

1. Find the mean, median, mode, and range of the data set. 8, 10, 46, 37, 20, 8, and 11

mean: 20; median: 11; mode: 8; range: 38

Lesson Quiz: Part II

2. Identify the outlier in the data set, and determine how the outlier affects the mean, median, and mode of the data. Then tell which measure of central tendency best describes the data with and without the outlier. Justify your answer.
85, 91, 83, 78, 79, 64, 81, 97

The outlier is 64. Without the outlier the mean is 85, the median is 83, and there is no mode. With the outlier the mean is 82, the median is 82, and there is no mode. Including the outlier decreases the mean by 3 and the median by 1, there is no mode. Because they have the same value and there is no outlier, the median and mean describes the data with the outlier. The median best describes the data without the outlier because it is closer to more of the other data values than the mean.

Measurement of dispersion tendency:

Absolute Deviation, Squared deviation, Mean Square error/ MS deviation aka Variance

Population *vs.* Sample: Designing to Sample

Measurement of dispersion tendency:

- The measures of central tendency are not adequate to describe data.
- Two data sets can have the same mean but they can be entirely different.
- Thus to describe data, one needs to know the extent of variability.
- Range, interquartile range, and standard deviation are the three commonly used measures of dispersion.

Average Deviation

- Average deviation is a statistical tool that provides the average of different variations from a data set.
- The purpose of average deviation is to measure the distance of a deviation from the data set's mean or median.
- The average deviation shows you the distance that one specific variable is from the mean of a data set.

Average deviation example: Here's an example to demonstrate the average deviation for a sport's player's seasonal scores:

- *A soccer player wants to determine the average deviation for the number of goals per game that she scored this season. Her data set is 5, 2, 7, 3, 1 and 4. First, she calculates the mean by adding up all the goals, then dividing by the total number of goals, which is six. To do so, she adds 5 + 2 + 7 + 3 + 1 + 4 = 22 goals. Then, she divides 22 / 6 = 3.6. This shows that she averaged 3.6 goals per game.*

- *Next she calculates the deviation, which is the value of each goal in the data set, from the mean.*

Exmple Cont….

- *The following equations to find the variation of each deviation:*
- *5 - 3.6 = 1.4*
- *2 - 3.6 = -1.6*
- *7 - 3.6 = 3.4*
- *3 - 3.6 = -0.6*
- *1 - 3.6 = -2.6*
- *4 - 3.6 = 0.4*

Example Cont…

- *Then, adds the sum of each variation, which is 1.4 + -1.6 + 3.4 + -0.6 + -2.6 + 0.4 = 0.4.*

- *The average deviation is the sum of all the deviations divided by the number of goals, which makes it 0.4 / 6 = 0.06.*
- *This means that the average deviation from the mean regarding the number of goals that she scored during the season is 0.06.*

Machine Learning

What is Learning?

- Herbert Simon: “Learning is any process by which a system improves performance from experience.”
- What is the task?
  - Classification
  - Categorization/clustering
  - Problem solving / planning / control
  - Prediction
  - others

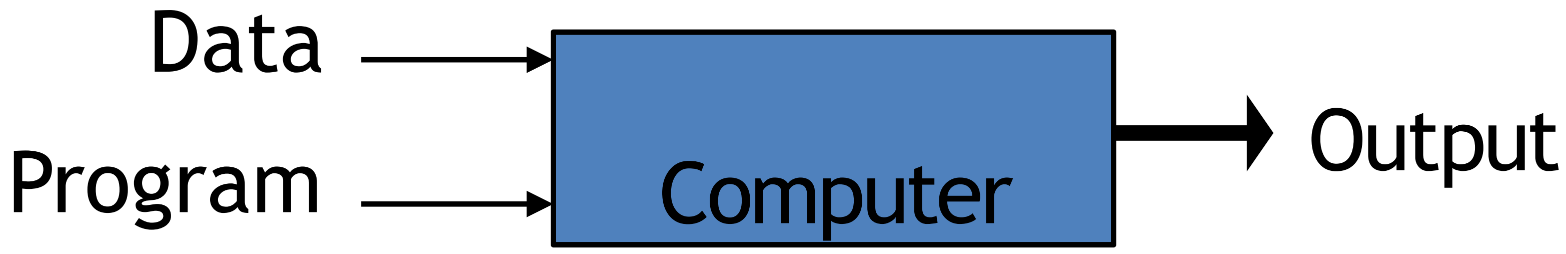
Why Study Machine Learning?
Developing Better Computing Systems

- Develop systems that are too difficult/expensive to construct manually because they require specific detailed skills or knowledge tuned to a specific task (***knowledge engineering bottleneck***).
- Develop systems that can automatically adapt and customize themselves to individual users.
  - Personalized news or mail filter
  - Personalized tutoring
- Discover new knowledge from large databases (***data mining***).
  - Market basket analysis (e.g. diapers and beer)
  - Medical text mining (e.g. migraines to calcium channel blockers to magnesium)

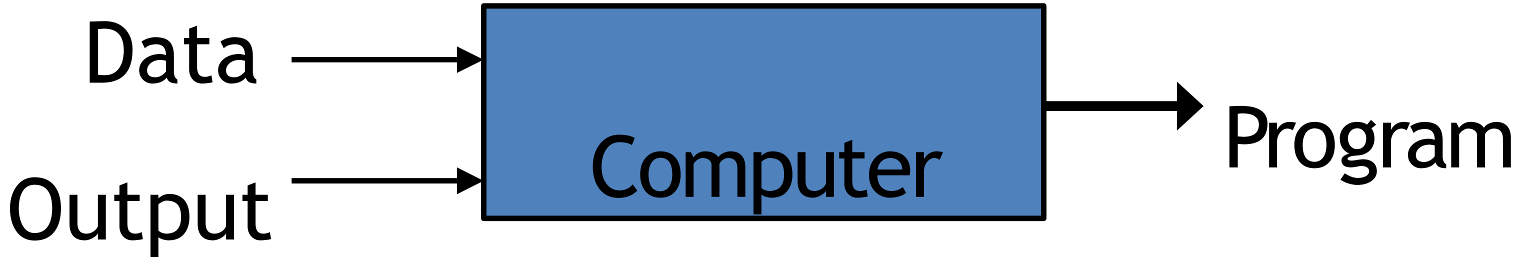
Related Disciplines

- Artificial Intelligence
- Data Mining
- Probability and Statistics
- Information theory
- Numerical optimization
- Computational complexity theory
- Control theory (adaptive)
- Psychology (developmental, cognitive)
- Neurobiology and many more

Traditional Programming



Machine Learning



Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging software
- [Your favorite area]

Human Learning

- It is a process of gaining information through observation, to solve/deal with real world scenarios.
- Human learning happens in one of the three ways:
  - An expert in the subject directly teaches us
  - We build our own notion indirectly based on what we have learnt in the past
  - We learn ourselves, may be after multiple attempts.

Types of Human Learning

- Learning under expert guidance
- Learning guided by knowledge gained from experts
- Learning by self

MACHINE LEARNING

A computer program is said to learn from experience ‘E’ with respect to some class of tasks ‘T’ and performance measure ‘P’, if its performance at tasks in ‘T’, as measured by ‘P’, improves with experience ‘E’.

- *Tom Mitchell*

History of Machine Learning

- 1950s
  - Automation, Machines (mechanical then arrived computation)
- 1960s:
  - Neural networks: Perceptron
  - Pattern recognition – Cover, Bradley, Brieman, Schapire, Nilsson, Hastie, Friedman, Tibshirani, Vapnik, Rumelhart, Schimdhuber, Hinton, Bengio, Blum
  - Learning in the limit theory
  - Minsky and Papert prove limitations of Perceptron

1970s:
  - Symbolic concept induction
  - Winston's arch learner
  - Expert systems and the knowledge acquisition bottleneck
  - Quinlan's ID3
  - Michalski's AQ and soybean diagnosis

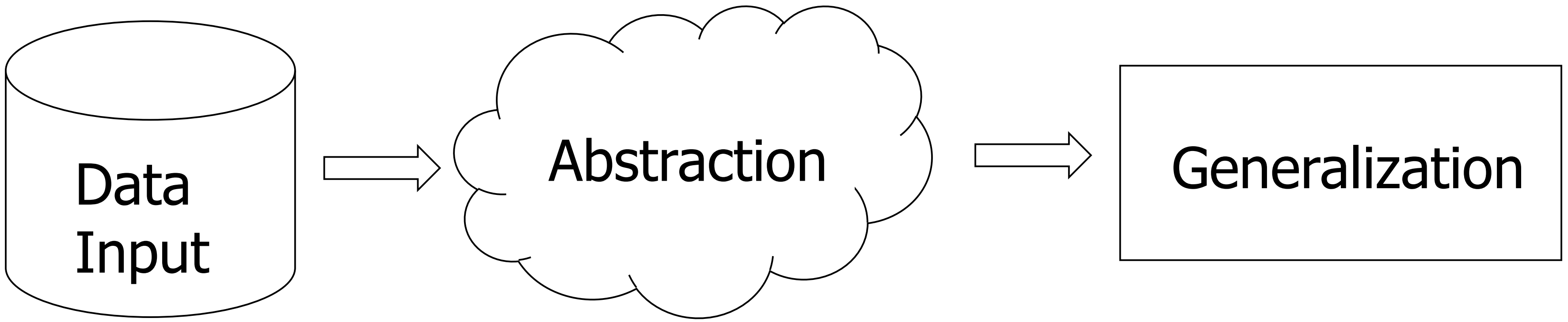
History of Machine Learning (cont.)

- 1980s:
  - Advanced decision tree and rule learning
  - Explanation-based Learning (EBL)
  - Learning and planning and problem solving
  - Utility problem, Analogy
  - Cognitive architectures
  - Resurgence of neural networks (connectionism, backpropagation)
- 1990s
  - Data mining
  - Adaptive software agents and web applications
  - Text learning
  - Reinforcement learning (RL) ….. 1950's Mendel, Fu, Richard Sutton
  - Inductive Logic Programming (ILP)
  - Ensembles: Bagging, Boosting, and Stacking
  - Bayes Net learning

History of Machine Learning (cont.)

- 2000s
  - Support vector machines
  - Kernel & Graphical models
  - Statistical relational and Transfer learning
  - Sequence labeling
  - Collective classification and structured outputs
  - Computer Systems Applications
    - Compilers
    - Debugging
    - Graphics
    - Security (intrusion, virus, and worm detection)
  - E-mail management
  - Personalized assistants that learn
  - Learning in robotics and vision

How do Machines Learn?



1. Data Input: Past data or information is utilized as a basis for future decision-making
2. Abstraction: The input data is represented in a broader way through the underlying algorithm.
3. Generalization: The abstracted representation is generalized to form a framework for making decisions.

Abstraction

- The data, given as input, cannot be used in the original form.
- **Abstraction helps in deriving a conceptual map based on the** input data.
- The model may be in any one of forms:
  - Computational blocks like if/else rules
  - Mathematical equations
  - Specific data structures like trees or graphs
  - Logical groupings of similar observations

Abstraction Cont’d

- The choice of the model is human specific.
- Selection of model is based on:
  - The type of problem to be solved.
  - Nature of the input data.
  - Domain of the problem.

Generalization

- Generalization is used for taking the decisions after training the model.
- The model is trained for a limited set of data. If we want to apply the model to take decision on a set of unknown data, we may encounter following problems:
  - The trained model is aligned with training data too much, hence may not represent the actual trend.
  - The test data possess certain characteristics apparently unknown to the training data.

Well-posed learning problem

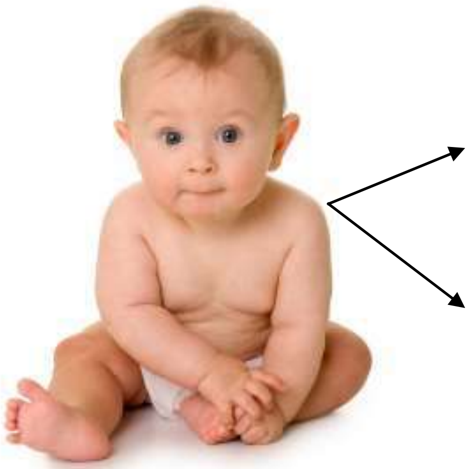
- A framework can be designed for deciding whether a problem can be solved using ML. The framework should answer:
  - What is the problem?
  - Why does the problem need to be solved?
  - How to solve the problem?

Types of Machine Learning

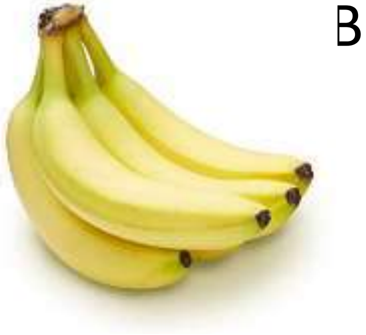
- Supervised learning: A machine predicts the class of unknown objects based on prior class-related information of similar objects. Also called predictive learning.
- Unsupervised/clustering learning: A machine finds patterns in unknown objects by grouping similar objects together. Also called descriptive learning.
- Reinforcement learning: A machine learns to act on its own to achieve the given goals.

SUPERVISED LEARNING

- The major motivation of supervised learning is to learn from past information.
- But how do machine learns?
    - By TRAINING DATA using labels.



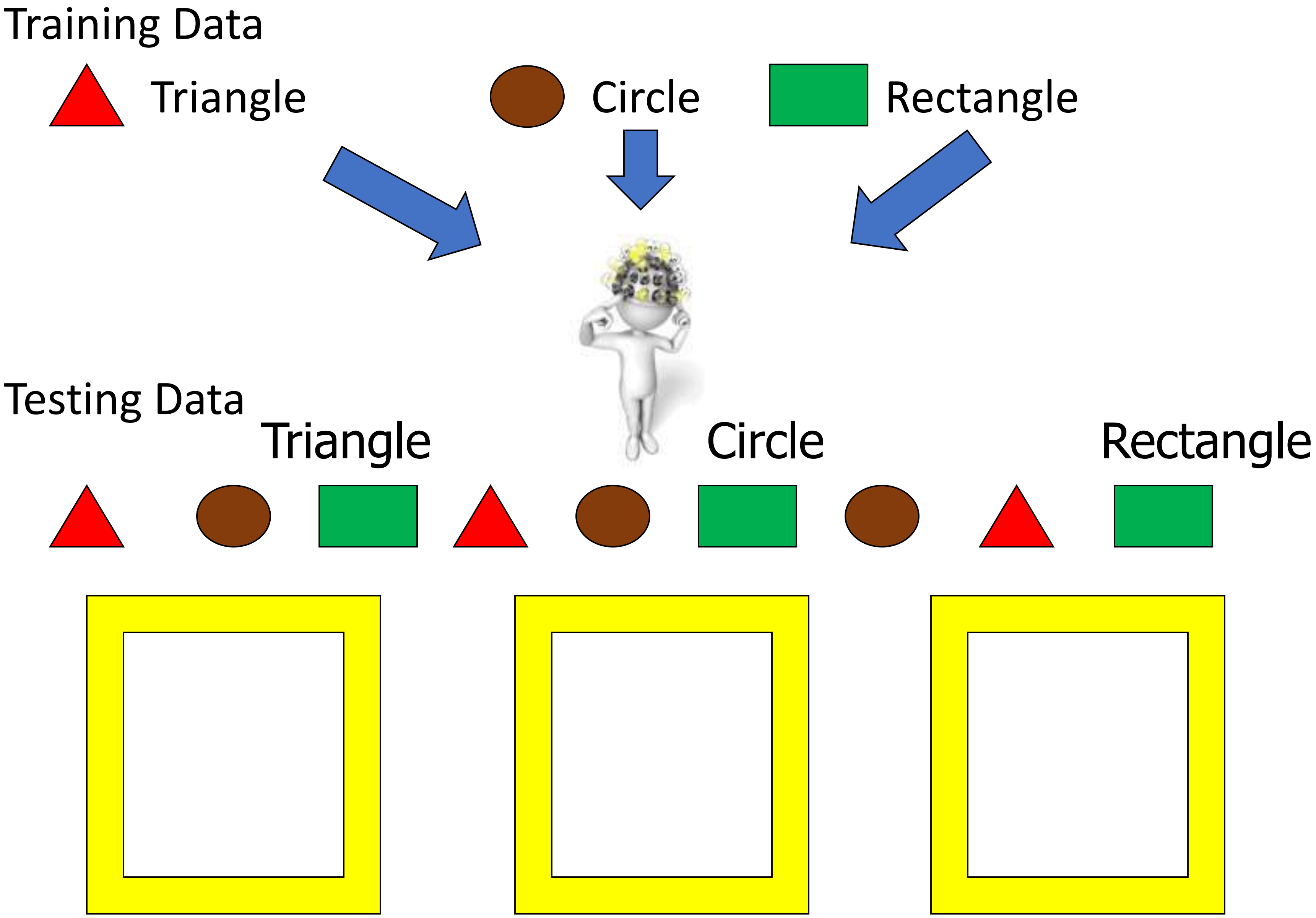
Apple



Banana

What's this?

It's an Apple


Training Data
Triangle
Circle
Rectangle
Testing Data
Triangle
Circle
Rectangle

Examples of Supervised Learning

- Predicting the results of a game.
- Predicting a tumour is malignant or benign.
- Predicting the piece of domains like real estate, stocks, etc.
- Classify texts such as classifying a set of emails as spam or non-spam.

Classification & Regression

- When we are trying to predict a categorical or nominal variable, the problem is known as a **classification** problem.
  - Ex: Identify Cat or Dog
- When we are trying to predict a real-valued variable, the problem falls under the category of **regression**.
  - Ex: Predict the value of a property.

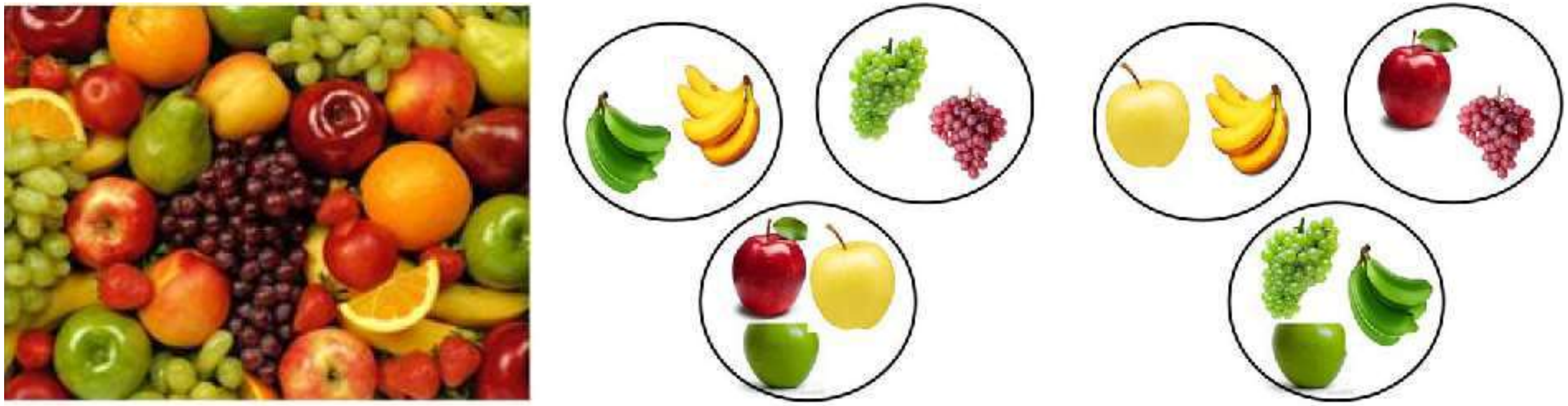
Unsupervised Learning

- Unsupervised learning(USL) is a type of self-organized learning that helps find previously unknown patterns in data set without pre-existing labels.
- The objective is to take a dataset as input and try to find natural grouping or patterns within the data elements or records.
- Hence, USL is termed as Descriptive Model and the process of USL is called Pattern or Knowledge Discovery.
- In unsupervised learning, the system is presented with unlabeled, uncategorized data and the system’s algorithms act on the data without prior training. The output is dependent upon the coded algorithms.

Unsupervised Learning

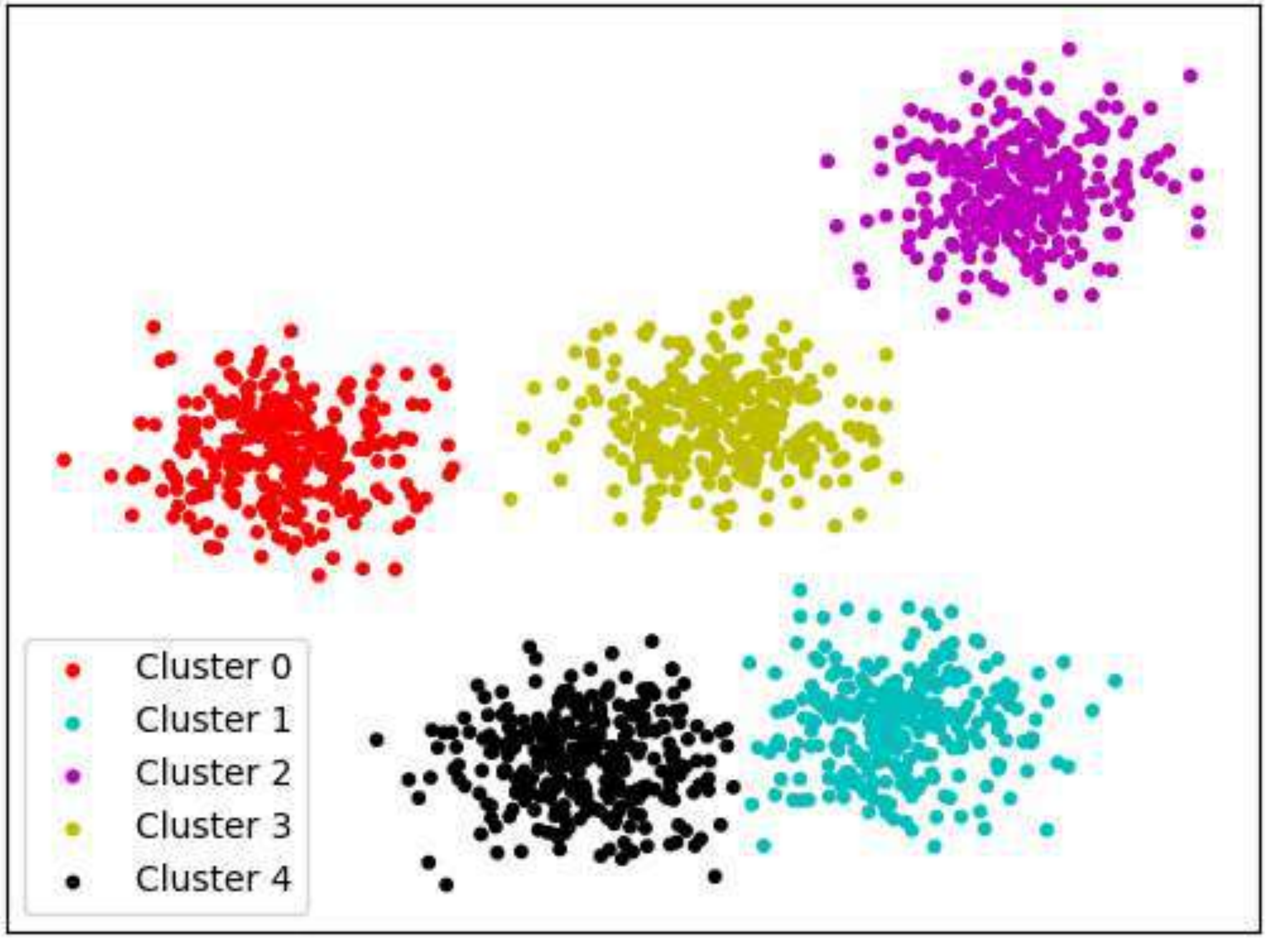
- Clustering is the main type of Unsupervised Learning.
- Clustering groups similar objects together.

Input Data | Cluster by type | Cluster by color

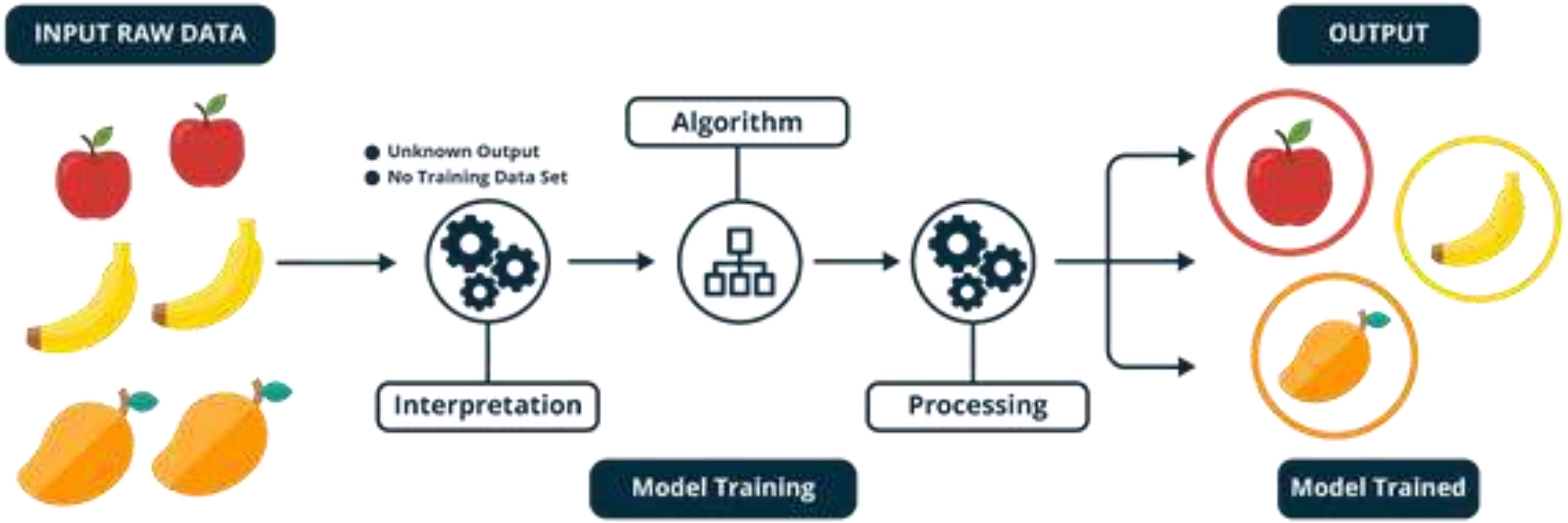


(a) (b) (c)

Sample Clustering



Unsupervised Learning



Unsupervised Learning Example: Categorize Cats and Dogs

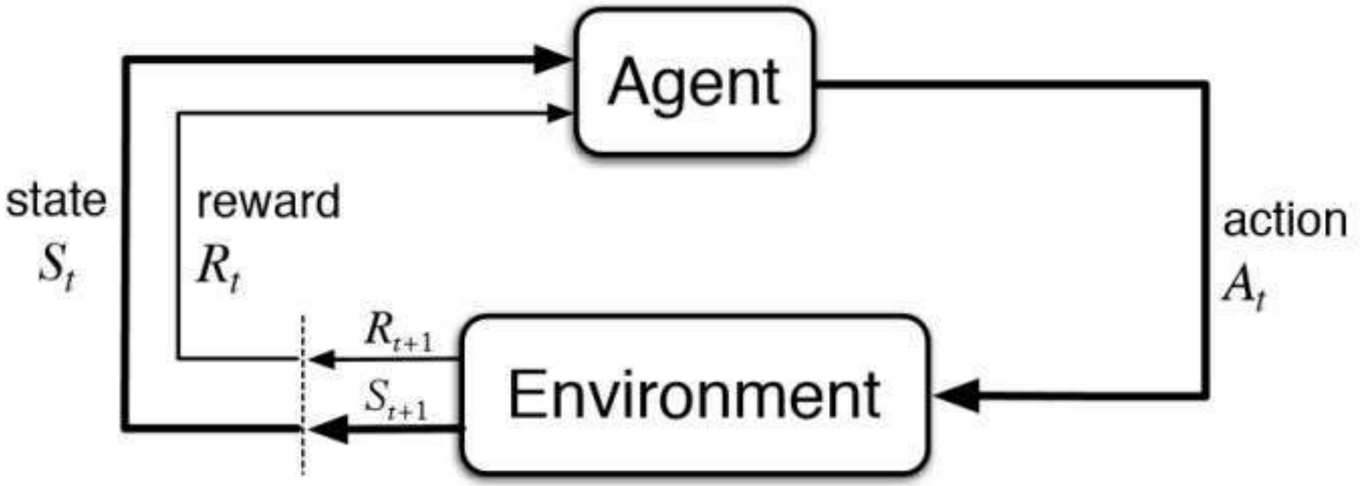
The image part with relationship ID rId2 was not found in the file.

Reinforcement Learning

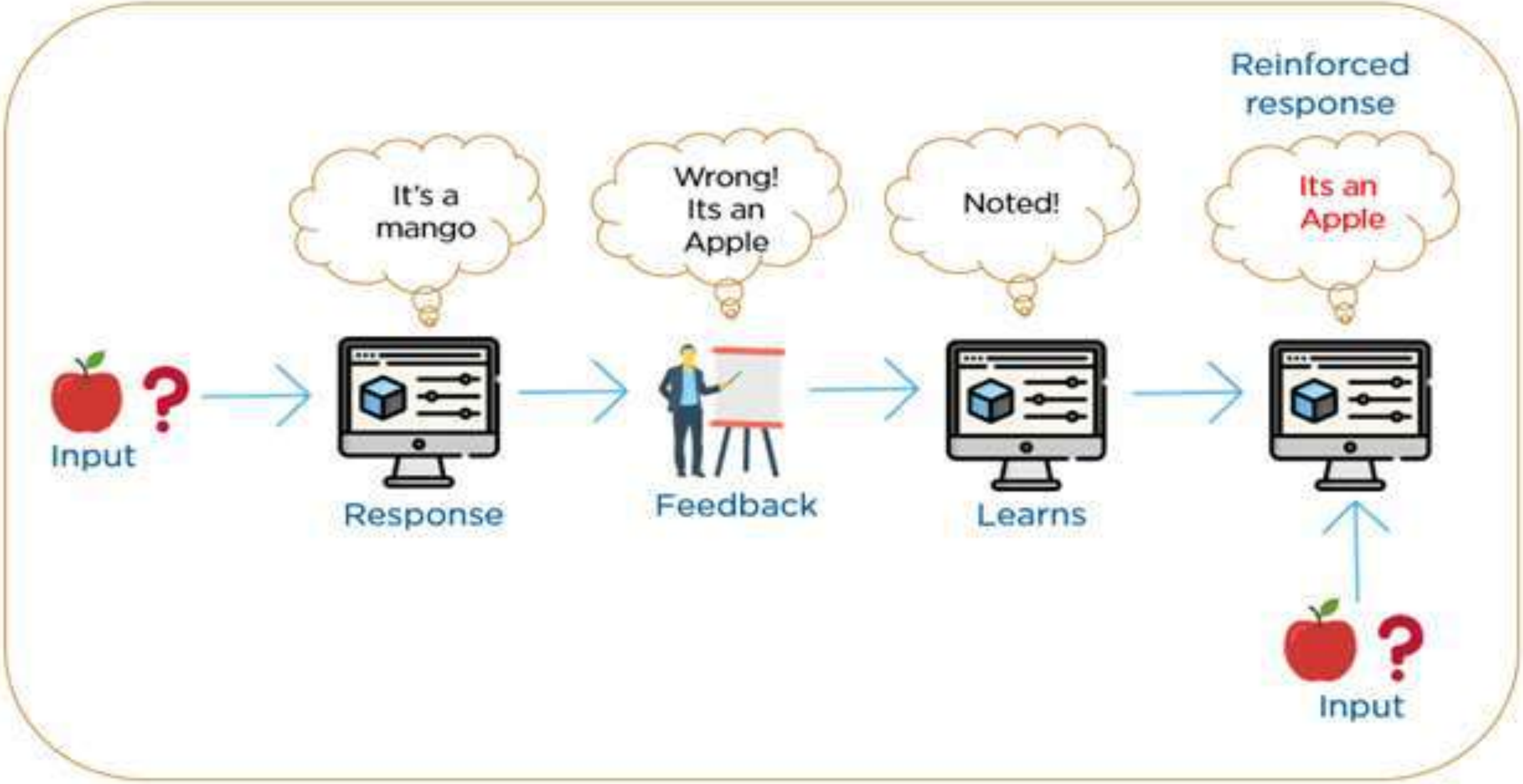
- It is about taking suitable action to maximize reward in a particular situation.
- It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation.
- Close to human learning.

Cont’d

- Algorithm learns a policy of how to act in a given environment.
- Every action has some impact in the environment, and the environment provides rewards that guides the learning algorithm.



Reinforcement Learning



Different Varieties of Machine Learning

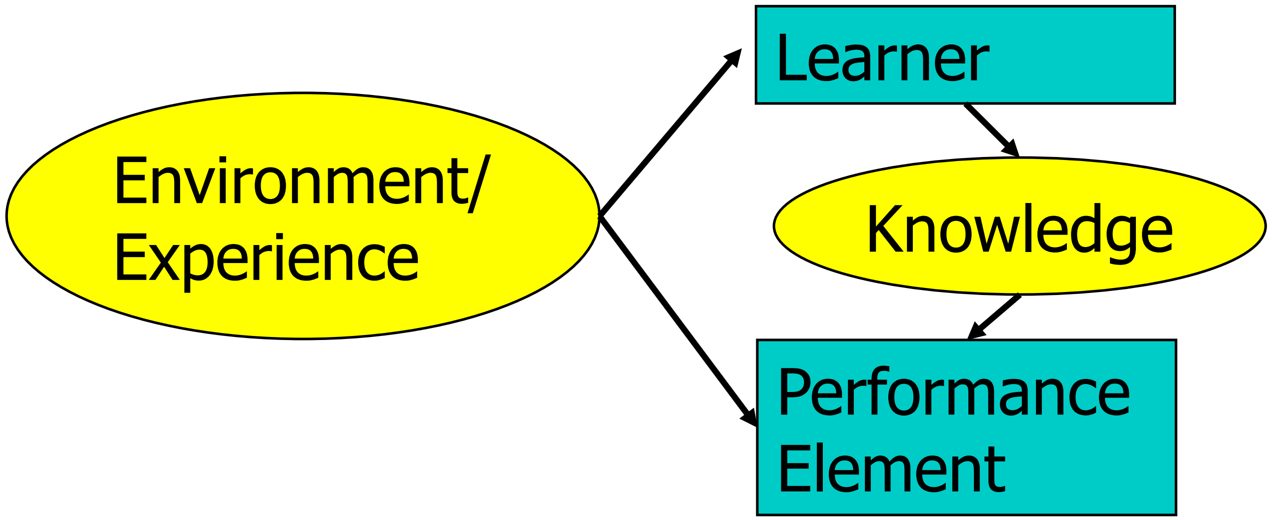
- Concept Learning
- Clustering Algorithms
- Connectionist Algorithms
- Genetic Algorithms
- Explanation-based and Transformation-based Learning
- Reinforcement and Case-based Learning
- Macro Learning
- Evaluation Functions
- Cognitive Learning Architectures
- Constructive Induction
- Discovery Systems

Languages or Tools for Machine Learning

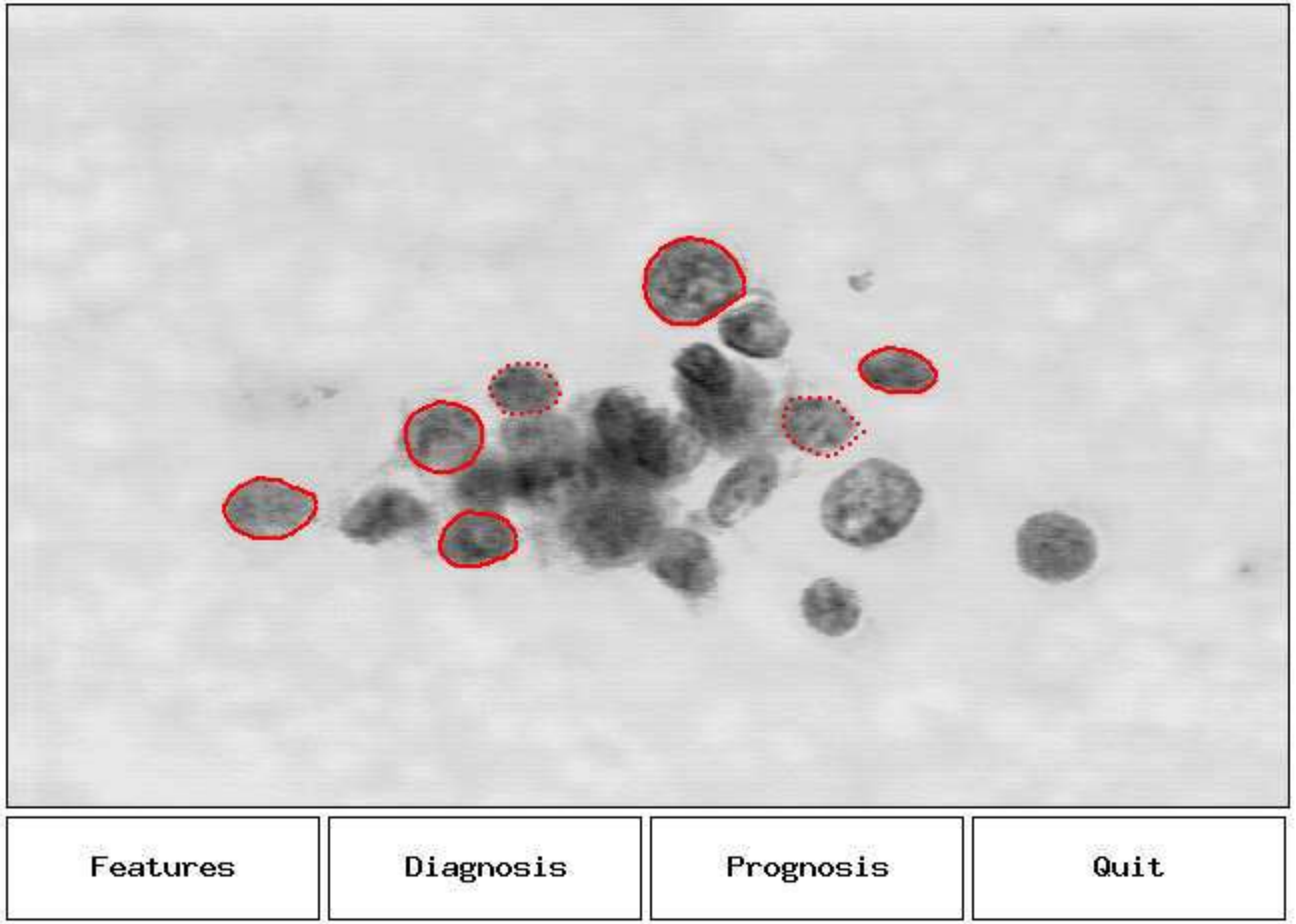
- Python – Open source programming language adopted for machine learning.
- R – Open source software. Used for statistical computing and data analysis
- Matlab - Developed by MathWorks. Licensed version. Used for variety of applications.
- SAS – Statistical Analysis System, was developed and licensed by SAS Institute provides strong support for ML.
- Others-
  - SPSS(Statistical Package for the Social Sciences) – IBM
  - Julia – MIT(Massachusetts Institute of Technology)

Designing a Learning System

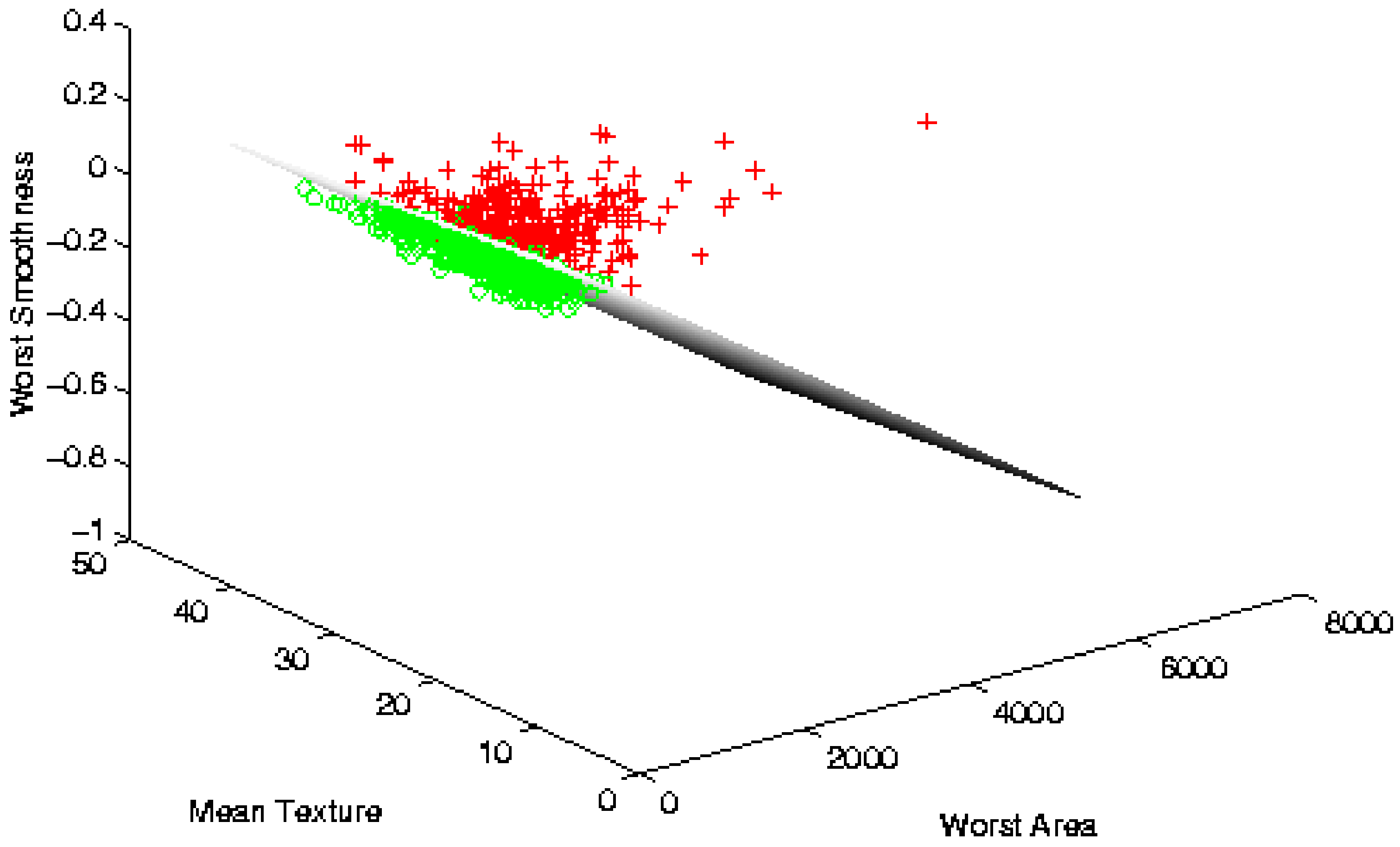
- Choose the training experience
- Choose exactly what is too be learned, i.e. the ***target function***.
- Choose how to represent the target function.
- Choose a learning algorithm to infer the target function from the experience.



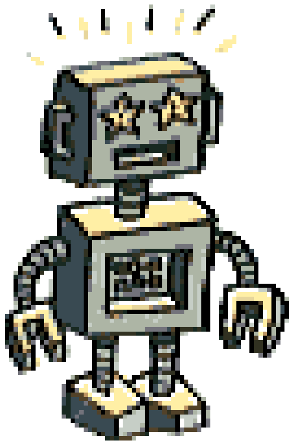
Application: Cancer Diagnosis



Cancer Diagnosis Separation



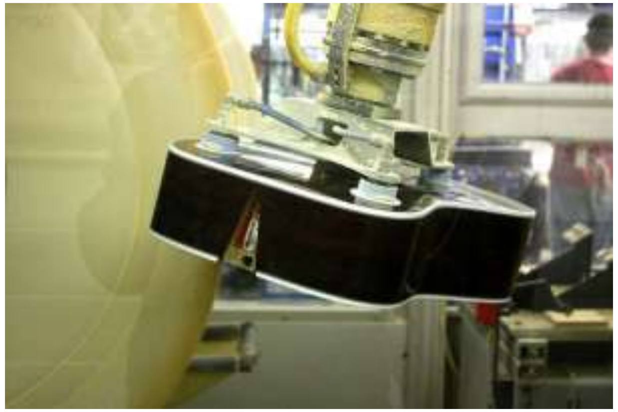
Robotics and ML



- Areas that robots are used:
  - Industrial robots
  - Military, government and space robots
  - Service robots for home, healthcare, laboratory
- Why are robots used?
  - Dangerous tasks or in hazardous environments
  - Repetitive tasks
  - High precision tasks or those requiring high quality
  - Labor savings
- Control technologies:
  - Autonomous (self-controlled), tele-operated (remote control)

Industrial Robots

- Uses for robots in manufacturing:
  - Welding
  - Painting
  - Cutting
  - Dispensing
  - Assembly
  - Polishing/Finishing
  - Material Handling
    - Packaging, Palletizing
    - Machine loading



Space Robots

- Mars Rovers – Spirit and Opportunity
  - Autonomous navigation features with human remote control and oversight

Service Robots

- Many uses…
  - Cleaning & Housekeeping
  - Humanitarian Demining
  - Rehabilitation
  - Inspection
  - Agriculture & Harvesting
  - Lawn Mowers
  - Surveillance
  - Mining Applications
  - Construction
  - Automatic Refilling
  - Fire Fighters
  - Search & Rescue

iRobot Roomba vacuum cleaner robot

Issues in Machine Learning

- What algorithms can approximate functions well and when?
  - How does the number of training examples influence accuracy
- Problem representation / feature extraction
- Intention/independent learning
- Integrating learning with systems
- What are the theoretical limits of learnability
- Transfer learning
- Continuous learning

Scaling issues in ML

- Number of
  - Inputs
  - Outputs
  - Batch vs realtime
  - Training vs testing

Machine Learning vs. Human Learning

- Some ML behavior can challenge the performance of human experts (e.g., playing chess)
- Although ML sometimes matches human learning capabilities, it is not able to learn as well as humans or in the same way that humans do
- There is no claim that machine learning can be applied in a truly creative way
- Formal theories of ML systems exist but are often lacking (why a method succeeds or fails is not clear)
- ML success is often attributed to manipulation of symbols (rather than mere numeric information)

Questions

- How does ML affect information science?
- Natural vs artificial learning – which is better?
- Is ML needed in all problems?
- What are the future directions of ML?

► The set $\mathbf{R}^n$ with operations of componentwise addition and scalar multiplication is an example of a **vector space,** and its elements are called **vectors.**

We shall henceforth interpret R^n to be a vector space.

(We say that $\mathbf{R}^n$ is **closed** under addition and scalar multiplication).

► In general, if **u** and **v** are vectors in the same vector space, then **u + v** is the diagonal of the **parallelogram** defined by **u** and **v**.

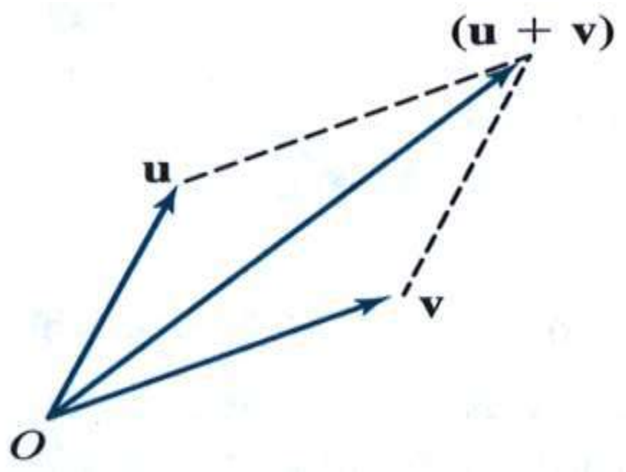


Figure 4.1

Example 2

Let **u** = (–1, 4, 3) and **v** = (–2, –3, 1) be elements of $\mathbf{R}^3$.

Find **u** + **v** and 3**u**.

Solution: **u** + **v** = (–1, 4, 3) + (– 2, –3, 1) = (-3 ,1 ,4)

3**u** = **3** (–1, 4, 3) = (-3 ,12 ,9)

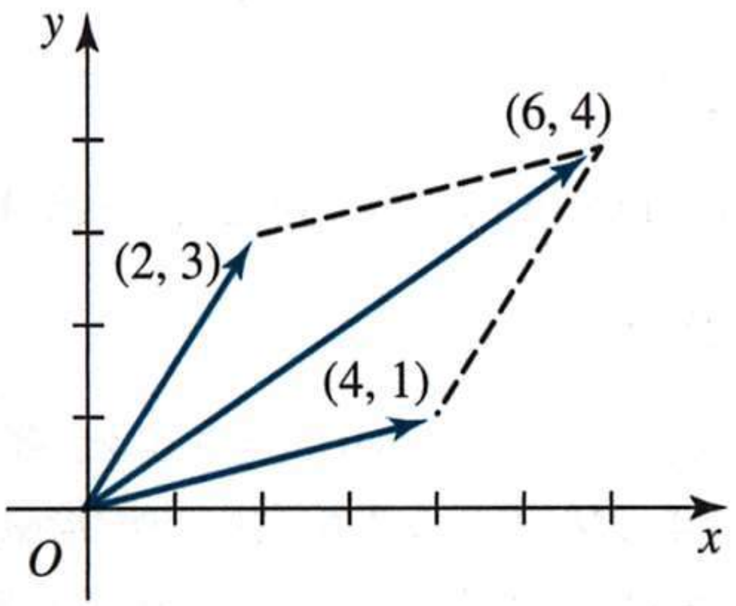
Example 3

In R^2 , consider the two elements (4, 1) and (2, 3).

Find their sum and give a geometrical interpretation of this sum.

we get (4, 1) + (2, 3) = (6, 4).

The vector (6, 4), the sum, is the diagonal of the parallelogram.

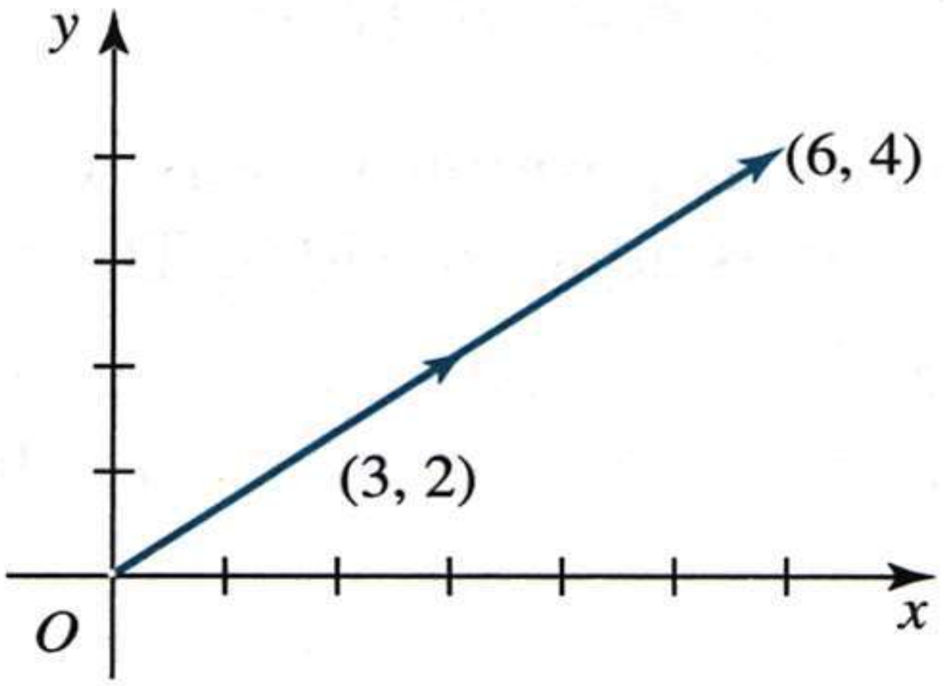


Example 4

Consider the scalar multiple of the vector (3, 2) by 2, we get

$$2(3, 2) = (6, 4)$$

Observe in Figure 4.3 that (6, 4) is a vector in the same direction as (3, 2), and 2 times it in length.

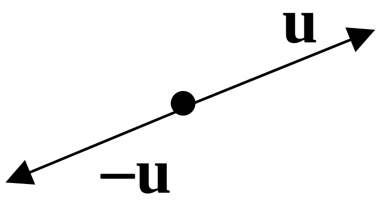


Zero Vector

The vector (0, 0, …, 0), having *n* zero components, is called the **zero vector** of $\mathbf{R}^n$ and is denoted **0**.

Negative Vector

The vector (–1)**u** is writing –**u** and is called **the negative of u**. It is a vector having the same length (or magnitude) as **u**, but lies in the opposite direction to **u**.



Subtraction

Subtraction is performed on element of $\mathbf{R}^n$ by subtracting corresponding components.

Theorem 4.1

Let **u**, **v**, and **w** be vectors in $\mathbf{R}^n$ and let c and d be scalars.

(a) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(b) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$

(c) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(d) $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$

(e) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(f) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(g) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(h) $1\mathbf{u} = \mathbf{u}$

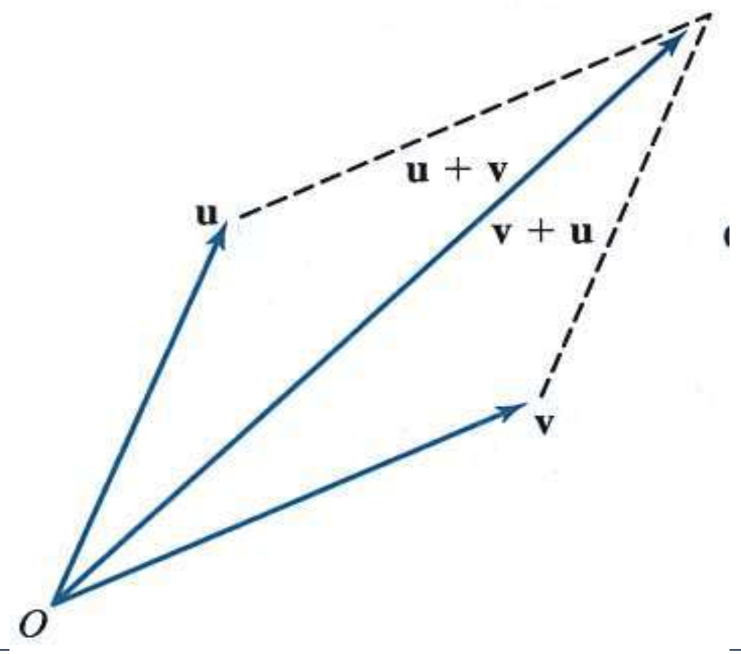


Figure 4.4

Commutativity of vector addition

$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

Example 5

Show that the following pairs of vectors are orthogonal.

(a) (1, 0) and (0, 1).

(b) (2, –3, 1) and (1, 2, 4).

Solution

(a) $\mathbf{(1, 0).(0, 1) = (1 \times 0) + (0 \times 1) = 0.}$

The vectors are orthogonal.

(b) $\mathbf{(2, -3, 1).(1, 2, 4) = (2 \times 1) + (-3 \times 2) + (1 \times 4) = 2 - 6 + 4 = 0.}$

The vectors are orthogonal.

Note

- (1, 0), (0,1) are <u>orthogonal</u> <u>unit</u> vectors in $\mathbf{R}^2$.
- (1, 0, 0), (0, 1, 0), (0, 0, 1) are orthogonal unit vectors in $\mathbf{R}^3$.
- (1, 0, …, 0), (0, 1, 0, …, 0), …, (0, …, 0, 1) are orthogonal unit vectors in $\mathbf{R}^n$.

Theorem 4.3

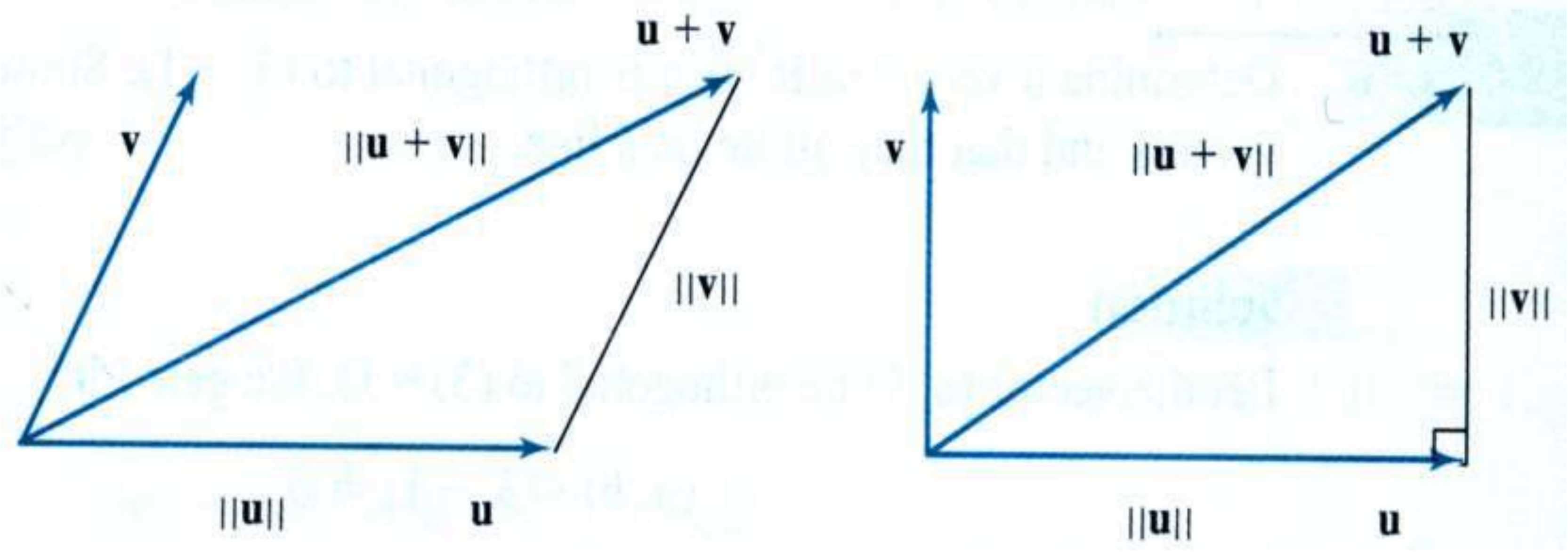
Let **u** and **v** be vectors in $\mathbf{R}^n$.

(a) Triangle Inequality:

$$\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|.$$

(a) Pythagorean theorem :

$$\text{If } \mathbf{u}.\mathbf{v} = 0 \text{ then } \|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$



Homework

- Exercise set 1.5 pages 47 to 48:
  3, 7, 8, 9, 11, 13, 16, 17, 26.

Exercise 36

Let **u** and **v** be vectors in $\mathbf{R}^n$.
Prove that $\|\mathbf{u}\| = \|\mathbf{v}\|$ *if and only if* $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ are orthogonal.

4.3 General Vector Spaces

Our aim in this section will be to focus on the algebraic properties of $\mathbf{R}^n$.

Definition

A **vector space** is a set V of elements called **vectors,** having operations of ***addition*** and ***scalar multiplication*** defined on it that satisfy the following conditions.

Let u, v, and w be arbitrary elements of V, and c and d are scalars.

- **Closure Axioms**

1. The sum $\mathbf{u} + \mathbf{v}$ exists and is an element of V. (V is closed under addition.)
2. $c\mathbf{u}$ is an element of V. (V is closed under scalar multiplication.)

Example 1

(1) V={ …, –3, –1, 1, 3, 5, 7, … }

V is not closed under addition because $1+3=4 \notin V$.

(2) Z={ …, –2, –1, 0, 1, 2, 3, 4, … }

Z is closed under addition because

for any $a, b \in Z$, $a + b \in Z$.

Z is not closed under scalar multiplication because

½ is a scalar, for any odd $a \in Z$, $(½)a \notin Z$.

Definition of Vector Space (continued)

- **Addition Axioms**

3. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (commutative property)
4. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (associative property)
5. There exists an element of V, called the **zero vector**, denoted $\mathbf{0}$, such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
6. For every element $\mathbf{u}$ of V there exists an element called the **negative** of $\mathbf{u}$, denoted $-\mathbf{u}$, such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

- **Scalar Multiplication Axioms**

7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$
10. $1\mathbf{u} = \mathbf{u}$

Vector Spaces of Functions

Prove that $\boldsymbol{F} = \{ f \mid f: R \to R \}$ is a vector space.

For example: $f: R \to R$, $f(x)=2x$,
$g: R \to R$, $g(x)=x^2+1$.

Let $f, g \in \boldsymbol{F}$, $c \in R$.

Axiom 1:

$f + g$ is defined by $(f + g)(x) = f(x) + g(x)$.

$\Rightarrow f + g : R \to R$

$\Rightarrow f + g \in \boldsymbol{F}$. Thus $\boldsymbol{F}$ is <u>closed under addition</u>.

Axiom 2:

cf is defined by $(cf)(x) = c \cdot f(x)$.

$\Rightarrow cf : R \to R$

$\Rightarrow cf \in \boldsymbol{F}$. Thus $\boldsymbol{F}$ is <u>closed under scalar multiplication</u>.

Theorem 4.4 (useful properties)

Let V be a vector space, $\mathbf{v}$ a vector in V, $\mathbf{0}$ the **zero vector** of V, c a scalar, and 0 the zero scalar. Then

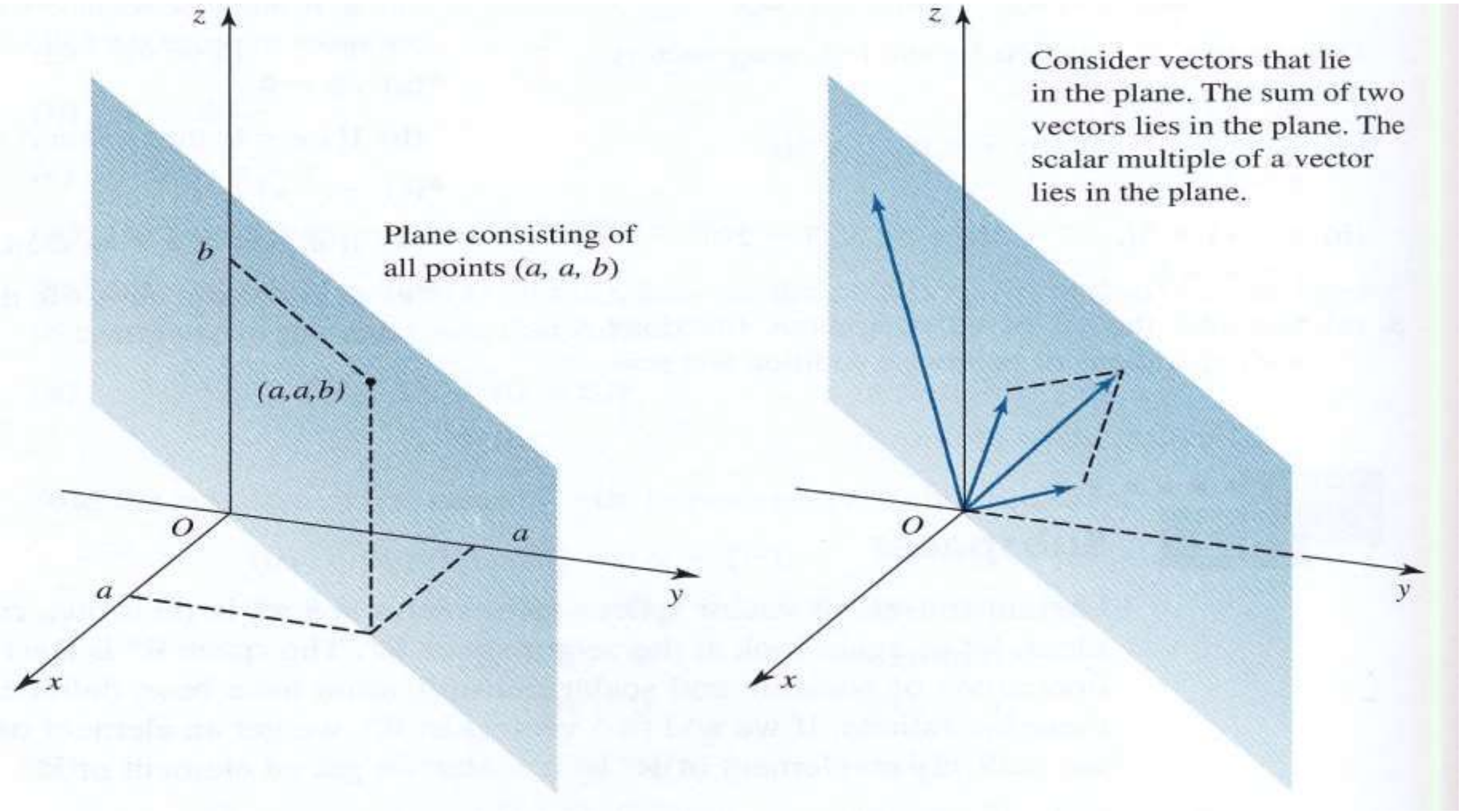
(a) $0\mathbf{v} = \mathbf{0}$

(b) $c\mathbf{0} = \mathbf{0}$

(c) $(-1)\mathbf{v} = -\mathbf{v}$

(d) If $c\mathbf{v} = \mathbf{0}$, then either $c = 0$ or $\mathbf{v} = \mathbf{0}$.

4.4 Subspaces



Note:

► In general, a subset of a vector space may or may not satisfy the closure axioms.

► However, any subset that is closed under both of these operations satisfies all the other vector space properties.

Definition

Let V be a vector space and U be a *nonempty* subset of V.

U is said to be a **subspace** of V if it is closed under addition and under scalar multiplication.

Example 1

Let U be the subset of $\mathbf{R}^3$ consisting of all vectors of the form $(a, 0, 0)$ (with zeros as second and third components and $a \in \mathbf{R}$),
i.e., $U = \{(a, 0, 0) \in \mathbf{R}^3 \}$.
Show that U is a subspace of $\mathbf{R}^3$.

Solution

Let $(a, 0, 0), (b, 0, 0) \in U$, and let $k \in \mathbf{R}$.
We get

$$(a, 0, 0) + (b, 0, 0) = (\mathrm{a} + b, 0, 0) \in U$$
$$k(a, 0, 0) = (k\, a, 0, 0) \in U$$

The sum and scalar product are in U.
Thus U is a subspace of $\mathbf{R}^3$. #

Geometrically, *U is the set of vectors that lie on the x-axis.*

Example 2

Let V be the set of vectors of of $\mathbf{R}^3$ of the form (a, a^2, b), namely
$V = \{(a, a^2, b) \in \mathbf{R}^3 \}$.
Show that V is not a subspace of $\mathbf{R}^3$.

Solution

Let $(a, a^2, b), (c, c^2, d) \in V$.

$$\begin{aligned}(a, a^2, b) + (c, c^2, d) &= (a + c, a^2 + c^2, b + d) \\ &\neq (a + c, (a + c)^2, b + d),\end{aligned}$$

since $a^2 + c^2 \neq (a + c)^2$.

Thus $(a, a^2, b) + (c, c^2, d) \notin V$.

V is not closed under addition.

V is not a subspace.

Theorem 4.5 (Very important condition)

Let U be a subspace of a vector space V.

U contains the zero vector of V.

Note. Let $\mathbf{0}$ be the zero vector of V.

If $\mathbf{0} \notin U \Rightarrow U$ is not a subspace of V.

If $\mathbf{0} \in U \Rightarrow (+)(\cdot)$ hold $\Rightarrow U$ is a subspace of V.

$(+)(\cdot)$ failed $\Rightarrow U$ is not a subspace of V.

Caution. ***This condition is necessary but not sufficient.***

(See, for instance, *Example 2* above and *Example 5* below)

Example 5

Let W be the set of vectors of the form $(a, a, a+2)$.
Show that W is not a subspace of $\mathbf{R}^3$.

Solution

If $(a, a, a+2) = (0, 0, 0)$, then $a = 0$ and $a + 2 = 0$.

This system is inconsistent it has no solution.

Thus $(0, 0, 0) \notin W$. (The necessary condition does not hold)

$\Rightarrow$ W is not a subspace of $\mathbf{R}^3$.

Homework

- Exercise set 4.1, pages 207-208:
  19, 21, 23, 25, 27, 29, 31, 33.
- Exercise set 1.3, page 32: 11, 12, 13, 15.

Exercise

Let $\boldsymbol{F} = \{ f \mid f\colon \mathbf{R} \to \mathbf{R} \}$ the vector space of functions on $\mathbf{R}$.
Which of the following are subspaces of $\boldsymbol{F}$?

(a) $\boldsymbol{W}_1 = \{ f \mid f\colon \mathbf{R} \to \mathbf{R},\ f(0)=0 \}$.

(b) $\boldsymbol{W}_2 = \{ f \mid f\colon \mathbf{R} \to \mathbf{R},\ f(0)=3 \}$.

(c) $\boldsymbol{W}_3 = \{ f \mid f\colon \mathbf{R} \to \mathbf{R},$ for some $c \in \mathbf{R},$ $f(x)=c$ for every $x\}$.

4.5 Linear Combinations of Vectors

$W=\{(a, a, b) \mid a,b \in \mathbf{R}\} \subseteq \mathbf{R}^3$

$(a, a, b) = a\,(1,1,0) + b\,(0,0,1)$

$\therefore W$ is generated by (1,1,0) and (0,0,1).

e.g., $(2, 2, 3) = 2\,(1,1,0) + 3\,(0,0,1)$

$(-1, -1, 7) = -1\,(1,1,0) + 7\,(0,0,1)$.

Definition

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ be vectors in a vector space V.

We say that $\mathbf{v}$, a vector of V, is a **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$, if there exist scalars $c_1, c_2, \ldots, c_m$ such that $\mathbf{v}$ can be written $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_m\mathbf{v}_m$.

Example 1

The vector (5, 4, 2) is a linear combination of the vectors (1, 2, 0), (3, 1, 4), and (1, 0, 3), since it can be written

(5, 4, 2) = (1, 2, 0) + 2(3, 1, 4) – 2(1, 0, 3)

Spanning Sets

Definition

The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are said to **span** a vector space if every vector in the space can be expressed as a *linear combination* of these vectors.

In this case $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is called a **spanning set**.

Theorem 4.6

Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be vectors in a vector space V. Let U be the set consisting of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.
Then U is a subspace of V spanned by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$.
U is said to be the vector space **generated** by $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

Proof

(+) Let $\mathbf{u}_1 = a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m$ and $\mathbf{u}_2 = b_1\mathbf{v}_1 + \ldots + b_m\mathbf{v}_m \in U$.

Then $\mathbf{u}_1 + \mathbf{u}_2 = (a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m) + (b_1\mathbf{v}_1 + \ldots + b_m\mathbf{v}_m)$
$= (a_1 + b_1)\,\mathbf{v}_1 + \ldots + (a_m + b_m)\,\mathbf{v}_m$

$\Rightarrow \mathbf{u}_1 + \mathbf{u}_2$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

$\Rightarrow \mathbf{u}_1 + \mathbf{u}_2 \in U$.

$\Rightarrow U$ is closed under vector addition.

(·)Let $c \in \mathbf{R}$. Then

$$c\mathbf{u}_1 = c(a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m)$$
$$= ca_1\mathbf{v}_1 + \ldots + ca_m\mathbf{v}_m$$

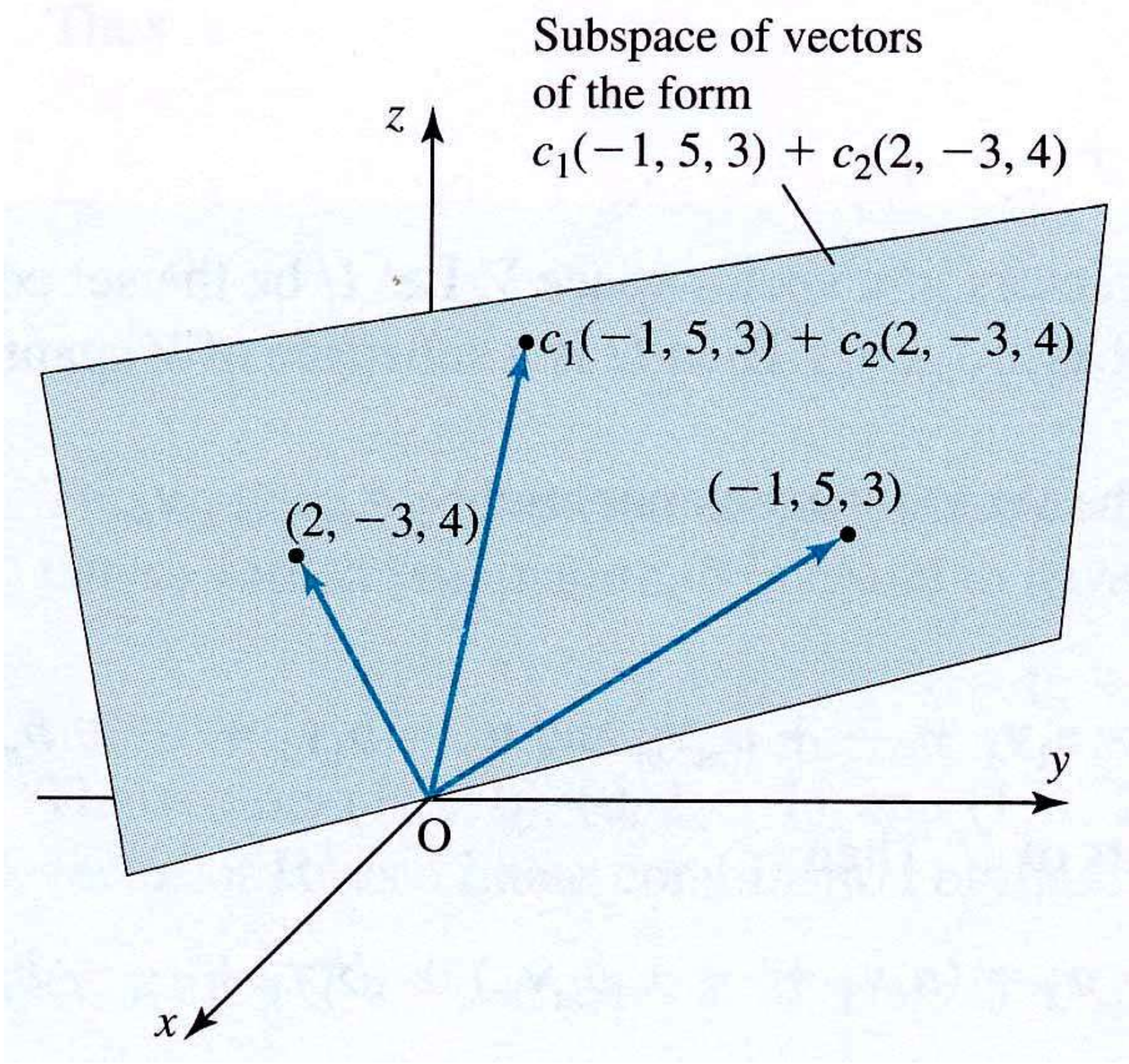
$\Rightarrow c\mathbf{u}_1$ is <u>a linear combination</u> of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.
$\Rightarrow c\mathbf{u}_1 \in U$.

$\Rightarrow U$ is <u>closed under scalar multiplication</u>.

Thus U is a subspace of V.

By the definition of U, every vector in U can be written as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

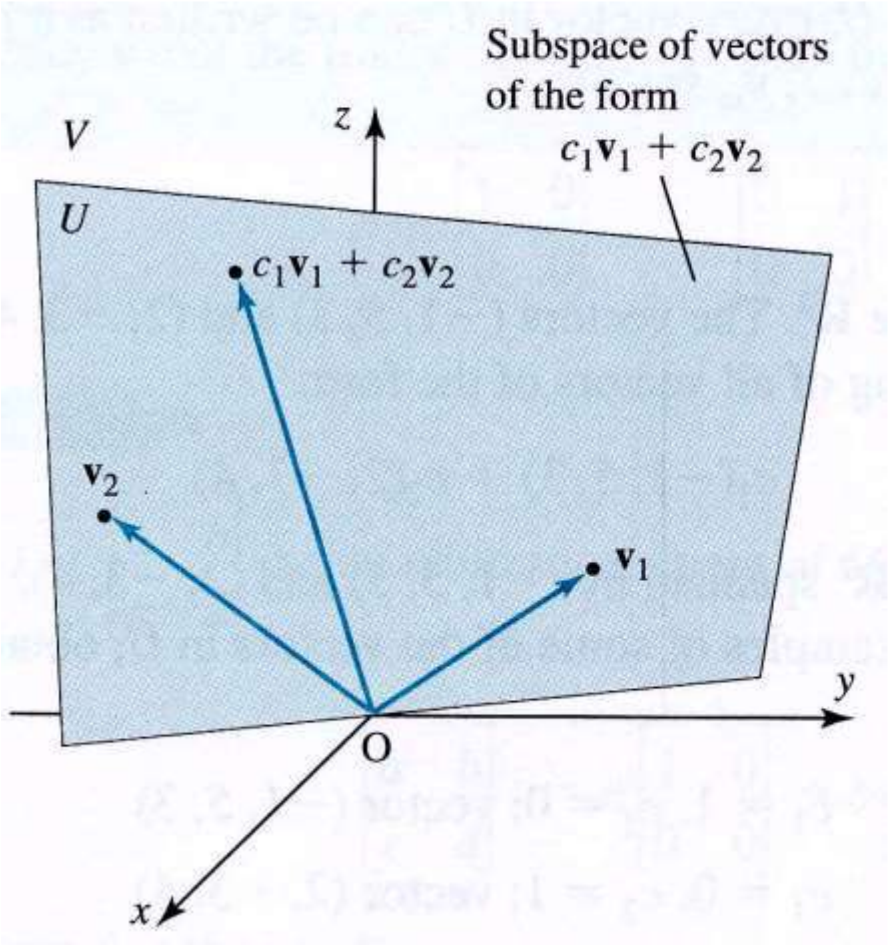
Thus $\mathbf{v}_1, \ldots, \mathbf{v}_m$ span U.


Subspace of vectors
of the form
$c_1(-1, 5, 3) + c_2(2, -3, 4)$
$c_1(-1, 5, 3) + c_2(2, -3, 4)$
$(2, -3, 4)$
$(-1, 5, 3)$
z
y
x
O

We can generalize this result. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be vectors in the space $\mathbf{R}^3$.

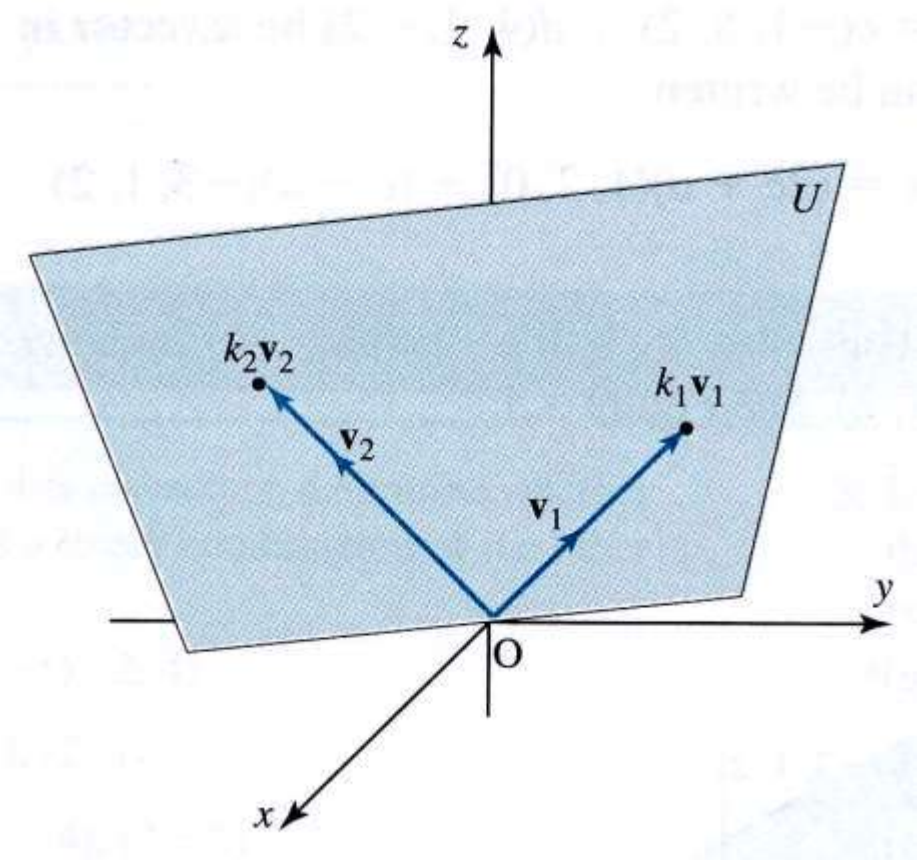
The subspace U generated by $\mathbf{v}_1$ and $\mathbf{v}_2$ is the set of all vectors of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$.

If $\mathbf{v}_1$ and $\mathbf{v}_2$ are not colinear, then U is the plane defined by $\mathbf{v}_1$ and $\mathbf{v}_2$.



If $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors in $\mathbf{R}^3$ that are not colinear, then we can visualize U as a plane in three dimensions.

$k_1\mathbf{v}_1$ and $k_2\mathbf{v}_2$ will be vectors on the same lines as $\mathbf{v}_1$ and $\mathbf{v}_2$.



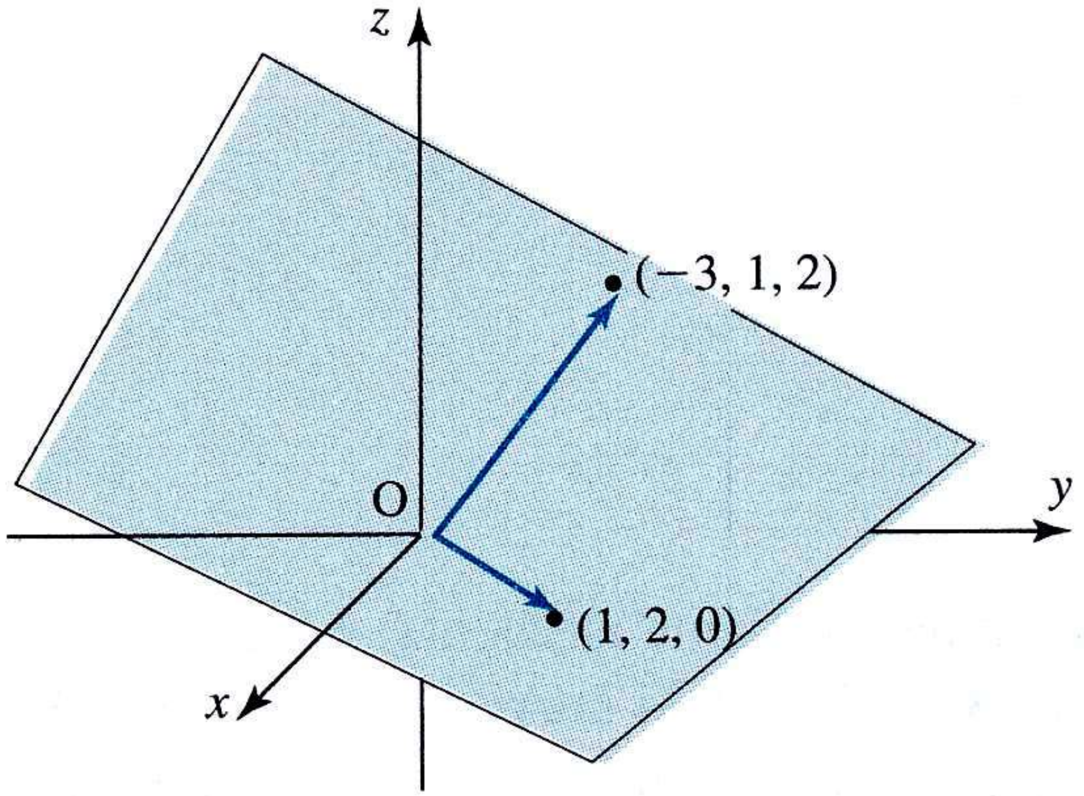


Figure 4.13

Example 11

Let U be the vector space generated by the functions $f(x) = x + 1$ and $g(x) = 2x^2 - 2x + 3$. Show that the function $h(x) = 6x^2 - 10x + 5$ lies in U.

Solution

h will be in the space generated by f and g if there exist scalars a and b such that

$$a(x + 1) + b(2x^2 - 2x + 3) = 6x^2 - 10x + 5$$

This given

$$2bx^2 + (a - 2b)x + a + 3b = 6x^2 - 10x + 5$$

$$\Rightarrow \quad \begin{aligned} 2b &= 6 \\ a - 2b &= -10 \\ a + 3b &= 5 \end{aligned}$$

This system has the unique solution $a = -4$, $b = 3$.

Thus $\quad -4(x + 1) + 3(2x^2 - 2x + 3) = 6x^2 - 10x + 5$

4.6 Linear Dependence and Independence

The concepts of dependence and independence of vectors are useful tools in constructing "efficient" spanning sets for vector spaces – sets in which there are no redundant vectors.

Definition

(a) The set of vectors $\{ \mathbf{v}_1, \ldots, \mathbf{v}_m \}$ in a vector space V is said to be **linearly dependent** if there exist scalars $c_1, \ldots, c_m$, not all zero, such that $c_1\mathbf{v}_1 + \ldots + c_m\mathbf{v}_m = 0$

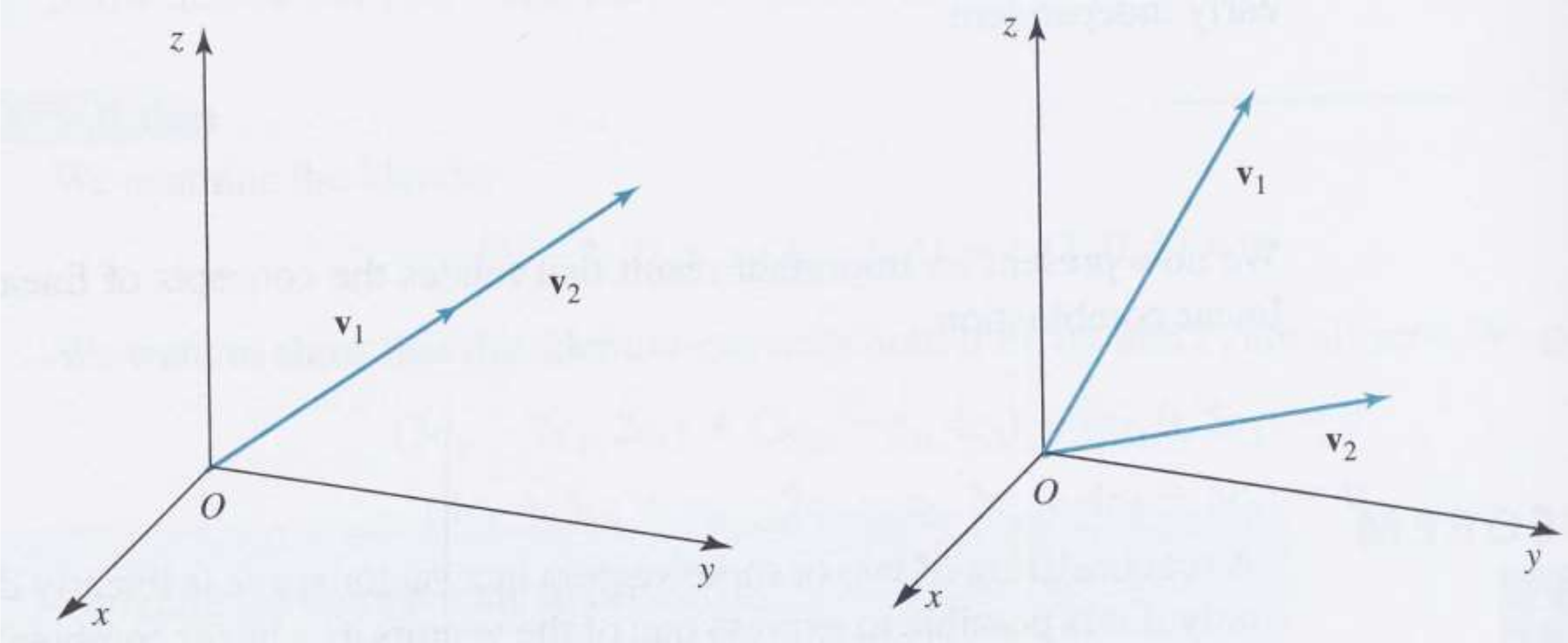
(b) The set of vectors $\{ \mathbf{v}_1, \ldots, \mathbf{v}_m \}$ is **linearly independent** if $c_1\mathbf{v}_1 + \ldots + c_m\mathbf{v}_m = 0$ can only be satisfied when $c_1 = 0, \ldots, c_m = 0$.

It can be shown that this system of three equations has the unique solution

$$c_1 = 0, c_2 = 0, c_3 = 0$$

Thus $c_1 f + c_2 g + c_3 h = \mathbf{0}$ implies that $c_1 = 0$, $c_2 = 0$, $c_3 = 0$.
The set $\{ f, g, h \}$ is linearly independent.

Linear Dependence of $\{v_1, v_2\}$

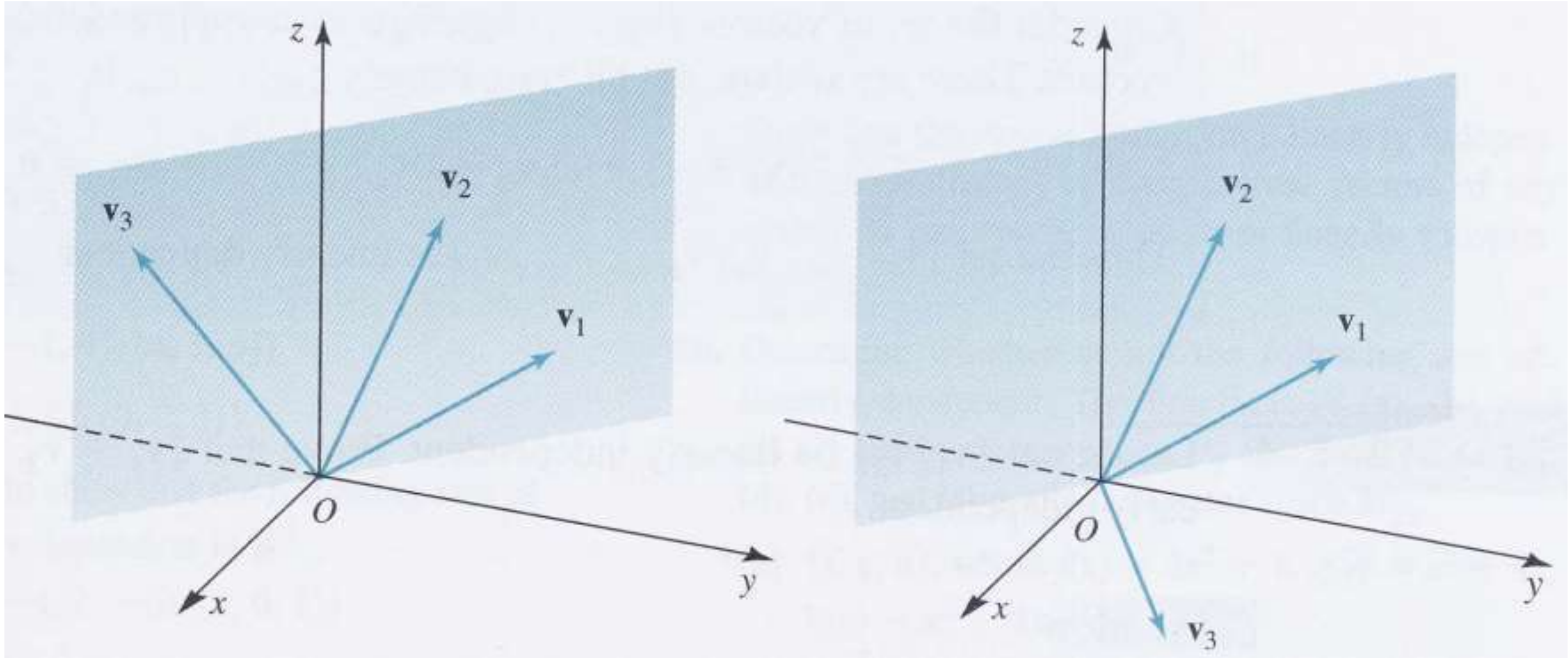


$\{\mathbf{v}_1, \mathbf{v}_2\}$ linearly dependent; vectors lie on a line

$\{\mathbf{v}_1, \mathbf{v}_2\}$ linearly independent; vectors do not lie on a line

Figure 4.14 Linear dependence and independence of $\{\mathbf{v}_1, \mathbf{v}_2\}$ in $\mathbf{R}^3$.

Linear Dependence of $\{v_1, v_2, v_3\}$



$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly dependent; vectors lie in a plane

$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly independent; vectors do not lie in a plane

Figure 4.15 Linear dependence and independence of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in $\mathbf{R}^3$.

Homework

- Exercise set 4.3 pages 219-220: 1, 3, 7, 8, 9, 13, 15, 17.

4.7 Bases and Dimension

Definition

A finite set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is called a **basis** for a vector space V if the set spans V and is linearly independent.

Standard Basis

The set of n vectors

$$\{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, \ldots, 1)\}$$

is a basis for $\mathbf{R}^n$. This basis is called the **standard basis** for $\mathbf{R}^n$.

How to prove it?

Theorem 4.11

Any two bases for a vector space V consist of the same number of vectors.

Proof

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be two bases for V.

By Theorem 4.10,

$$m \le n \text{ and } n \le m$$

Thus $n = m$.

Definition

If a vector space V has a basis consisting of n vectors, then the **dimension** of V is said to be n. We write **dim(V)** for the dimension of V.

- V is **finite dimensional** if such a finite basis exists.
- V is **infinite dimensional** otherwise.

Theorem 4.12

(a) The origin is a subspace of $\mathbf{R}^3$. The dimension of this subspace is defined to be zero.

(b) The one-dimensional subspaces of $\mathbf{R}^3$ are lines through the origin.

(c) The two-dimensional subspaces of $\mathbf{R}^3$ are planes through the origin.

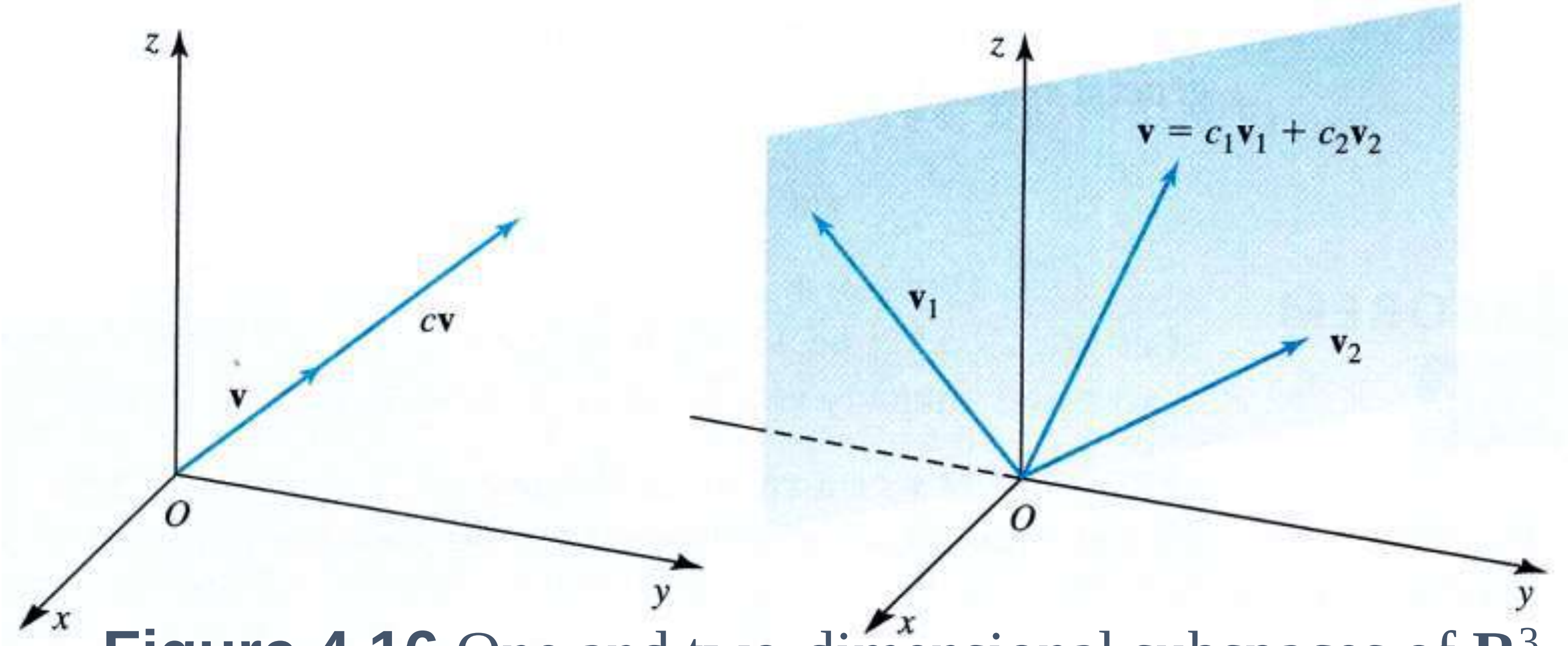


Figure 4.16 One and two-dimensional subspaces of $\mathbf{R}^3$

Proof

(a) Let V be the set $\{(0, 0, 0)\}$, consisting of a single elements, the zero vector of $\mathbf{R}^3$. Let c be the arbitrary scalar. Since

$$(0, 0, 0) + (0, 0, 0) = (0, 0, 0) \text{ and } c(0, 0, 0) = (0, 0, 0)$$

V is closed under addition and scalar multiplication. It is thus a subspace of $\mathbf{R}^3$. The dimension of this subspaces is defined to be zero.

(b) Let $\mathbf{v}$ be a basis for a one-dimensional subspace V of $\mathbf{R}^3$. Every vector in V is thus of the form $c\mathbf{v}$, for some scalar c. We know that these vectors form a line through the origin.

(c) Let $\{\mathbf{v}_1, \mathbf{v}_2\}$be a basis for a two-dimensional subspace V of $\mathbf{R}^3$. Every vector in V is of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. V is thus a plane through the origin.

Theorem 4.14

Let V be a vector space of dimension n.

(a) If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a set of n linearly independent vectors in V, then S is a basis for V.

(b) If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a set of n vectors V that spans V, then S is a basis for V.

Let V be a vector space, $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a set of vectors in V.

(a) $\dim(V) = |S|$.

(b) S is a linearly independent set.

(c) S spans V.

S is a basis of V.

Theorem 4.15

Let V be a vector space of dimension n. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a set of m linearly independent vectors in V, where $m < n$.

Then there exist vectors $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ such that

$\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$ is a basis of V.

Example 5

State (with a brief explanation) whether the following statements are true or false.

(a) The vectors (1, 2), (–1, 3), (5, 2) are linearly dependent in $\mathbf{R}^2$.

(b) The vectors (1, 0, 0), (0, 2, 0), (1, 2, 0) span $\mathbf{R}^3$.

(c) {(1, 0, 2), (0, 1, -3)} is a basis for the subspace of $\mathbf{R}^3$ consisting of vectors of the form $(a, b, 2a - 3b)$.

(d) Any set of two vectors can be used to generate a two-dimensional subspace of $\mathbf{R}^3$.

Solution

(a) True: The dimension of $\mathbf{R}^2$ is two. Thus any three vectors are linearly dependent.

(b) False: The three vectors are linearly dependent. Thus they cannot span a three-dimensional space.

(c) True: The vectors span the subspace since

$$(a, b, 2a - 3b) = a(1, 0, 2) + b(0, 1, -3)$$

The vectors are also linearly independent since they are not colinear.

(d) False: The two vectors must be linearly independent.

Inverse of Matrix

Prepared by Vince Zaccone

For Campus Learning
Assistance Services at UCSB