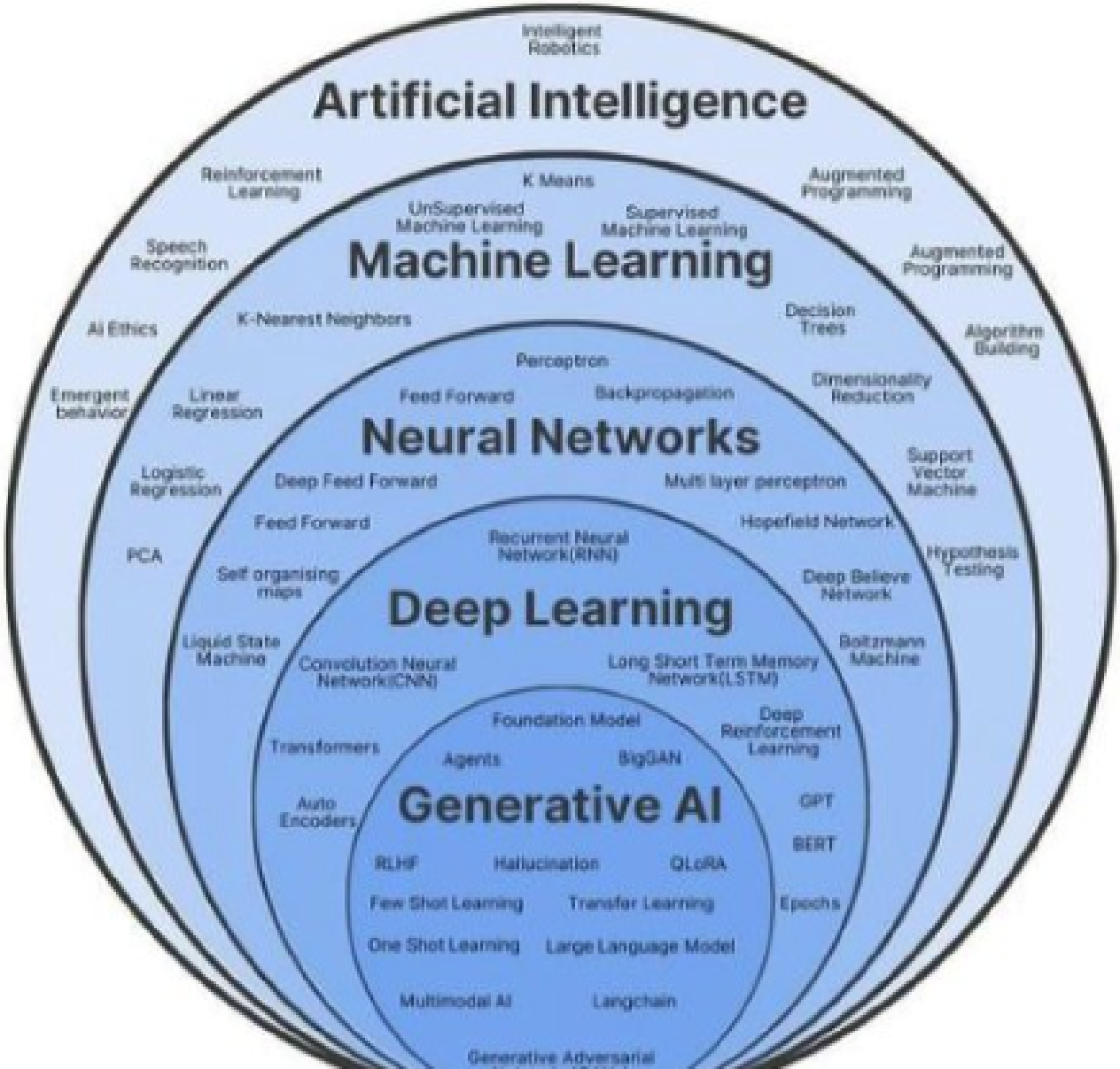
CS5805 : Machine Learning I
Lecture #1

Reza Jafari, Ph.D

Collegiate Associate Professor
rjafari@vt.edu

World of Artificial Intelligence



Can you read this?

http://www.mrc-cbu.cam.ac.uk/~mattd/Cmabrigde/ (January 2008)

Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.

Machine Learning

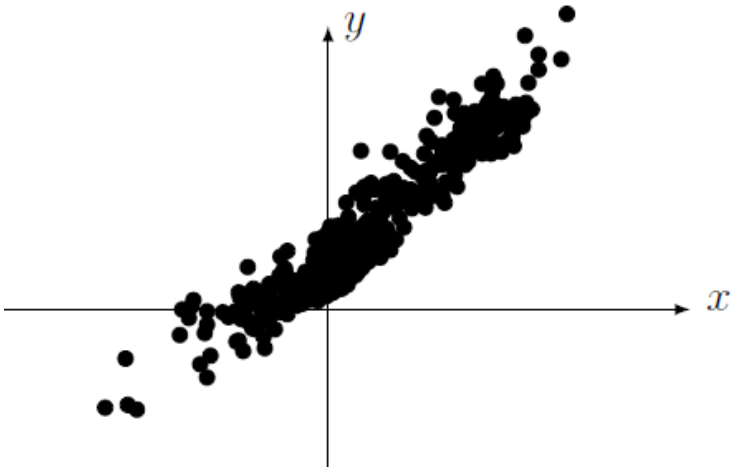
- Imagine looking outside of the window during the morning of a winter day in Finland.
- We can down load the recordings of minimum and maximum daytime temperature for most days

$$\mathbb{D} = \{\mathbf{z}^{(1)}, ..., \mathbf{z}^{(m)}\}$$

Machine Learning

- Each data point $\mathbf{z}^{(i)} = (x^{(i)}, y^{(i)})$ for $i = 1, ..., n$ represents some previous day min and max daytime temperature $x^{(i)}$ and $y^{(i)}$.
- Machine Learning learn a hypothesis $h(x)$ reads min temperature and delivers a prediction $\hat{y} = h(x)$.
- Every practical ML method uses a particular hypothesis space out of which the hypothesis h is chosen.

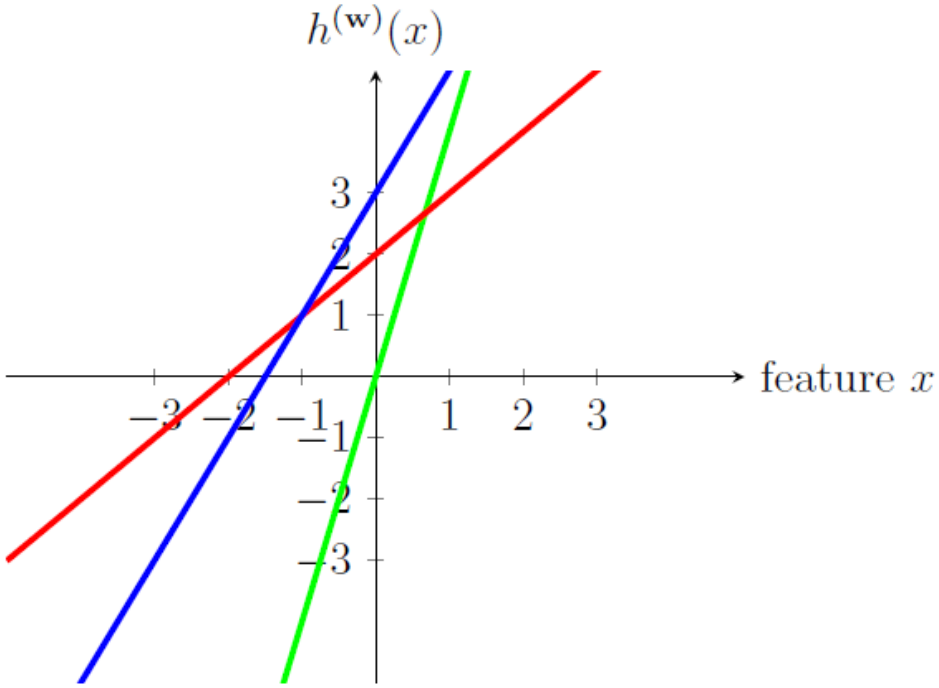
$$\boxed{h(x) := w_1 x + w_0}, w_1 \in \mathbb{R}_+, w_0 \in \mathbb{R}$$



Loss function

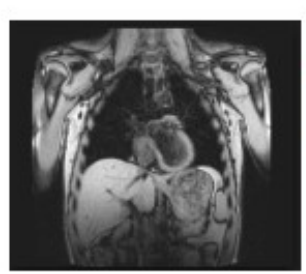
- The loss function quantify the difference between actual data and predicted data.
- Or Mean of Squared of Errors[MSE]

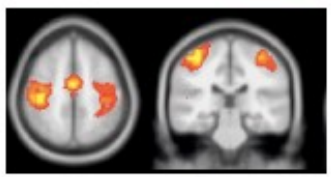
$$(1/n)\sum_{i=1}^{n}(y^{(i)} - h(x^{(i)}))^2$$

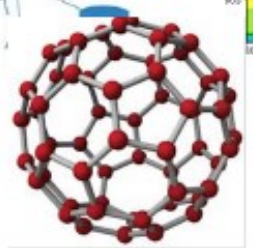


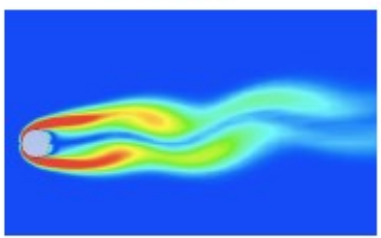
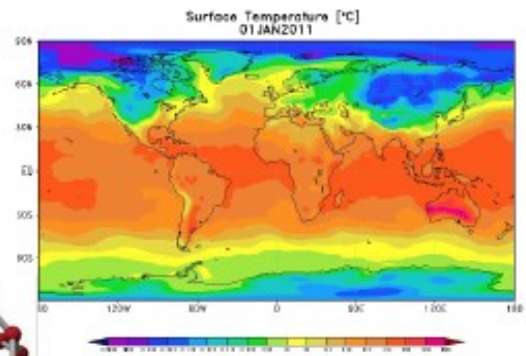
Large-scale Data is Everywhere!

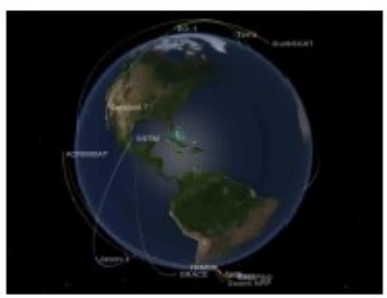
- There has been enormous growth of data in both commercial and scientific arena due to advances in data generation, storage and retrieval technologies.







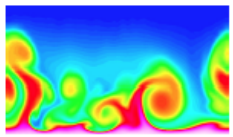


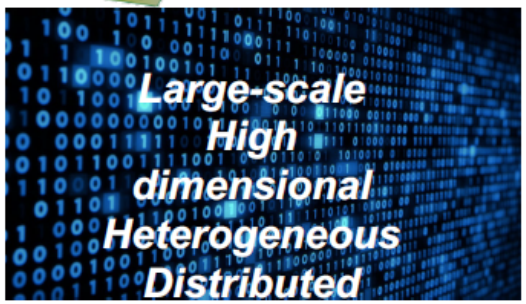
Large-scale Data is Everywhere!

- There has been enormous growth of data in both commercial and scientific arena due to advances in data generation, storage and retrieval technologies.
- Every day, $\approx$ 2.5 quintillion($\times 10^{18}$) bytes of data is created.

Golden Age of Data Science

Data → ***Data Science*** → **Knowledge**





Machine Learning
Artificial Intelligence
Pattern Recognition
Data Mining / Data Analytics

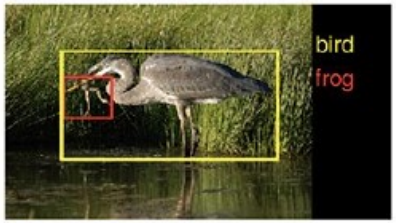
- **Patterns**
- **Models**
- **Relationships**

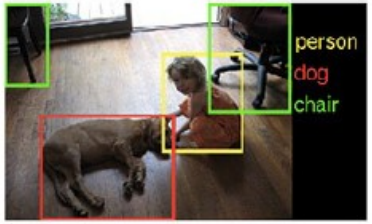
Automated tools for knowledge extraction from large volumes of data

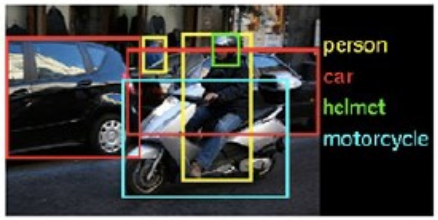
Why Data Mining? Commercial Viewpoint

- Lots of data is being collected and warehoused.
- Competitive pressure is strong.

IMAGENET







DeepMind

NETFLIX

WATSON VS HUMANS: THE JEOPARDY FINALE LIVE

facebook

Google AI algorithm masters ancient game of Go

Google Ads

Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.

Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.

Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.
    - NASA EOSDIS archives over petabytes of earth science data / year

Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.
    - NASA EOSDIS archives over petabytes of earth science data / year
  - Telescope scanning the skies.

Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.
    - NASA EOSDIS archives over petabytes of earth science data / year
  - Telescope scanning the skies.
  - High-throughput biological data.

Why Data Mining? Scientific Viewpoint

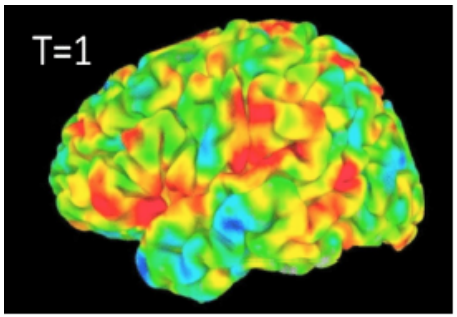
- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.
    - NASA EOSDIS archives over petabytes of earth science data / year
  - Telescope scanning the skies.
  - High-throughput biological data.
  - Scientific simulations

Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.
    - NASA EOSDIS archives over petabytes of earth science data / year
  - Telescope scanning the skies.
  - High-throughput biological data.
  - Scientific simulations
    - terabytes of data generated in a few hours

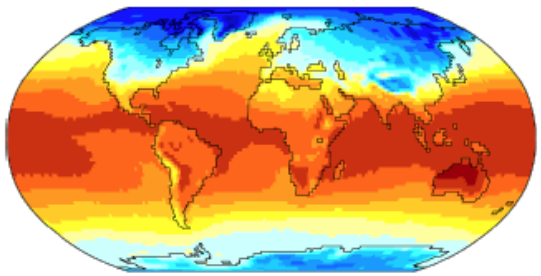
Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.
    - NASA EOSDIS archives over petabytes of earth science data / year
  - Telescope scanning the skies.
  - High-throughput biological data.
  - Scientific simulations
    - terabytes of data generated in a few hours
- Data mining helps scientists:



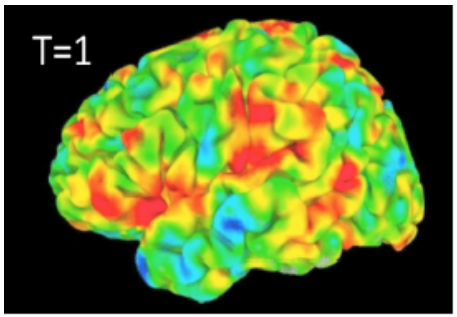
fMRI Data from Brain


Sky Survey Data


Surface Temperature of Earth

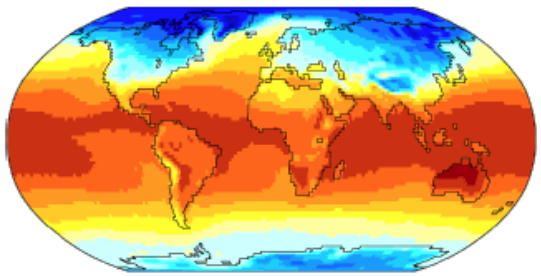
Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.
    - NASA EOSDIS archives over petabytes of earth science data / year
  - Telescope scanning the skies.
  - High-throughput biological data.
  - Scientific simulations
    - terabytes of data generated in a few hours
- Data mining helps scientists:
  - In automated analysis of massive datasets



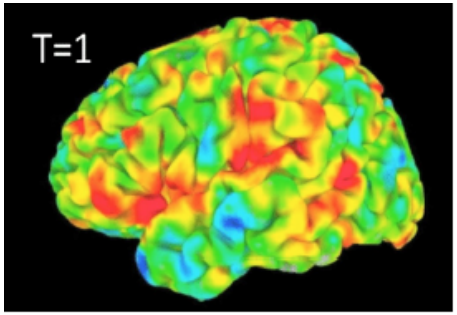
fMRI Data from Brain


Sky Survey Data


Surface Temperature of Earth

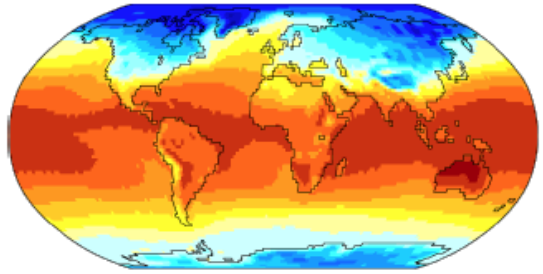
Why Data Mining? Scientific Viewpoint

- Data collected and stored at enormous speeds.
  - Remote sensors on a satellite.
    - NASA EOSDIS archives over petabytes of earth science data / year
  - Telescope scanning the skies.
  - High-throughput biological data.
  - Scientific simulations
    - terabytes of data generated in a few hours
- Data mining helps scientists:
  - In automated analysis of massive datasets
  - In hypothesis formation.



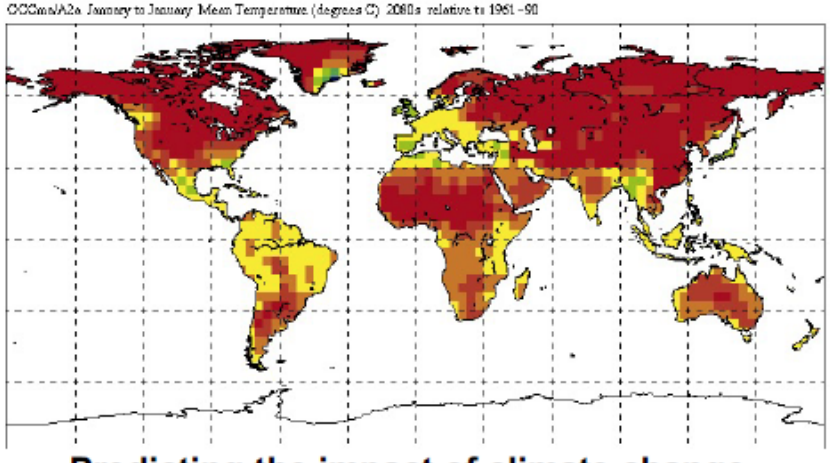
fMRI Data from Brain


Sky Survey Data


Surface Temperature of Earth

Great Opportunities to Solve Society's Major Problems

Improving health care and reducing costs



Predicting the impact of climate change

Finding alternative/ green energy sources

Reducing hunger and poverty by increasing agriculture production

What is Not Data Mining?

- What is not Data Mining?
  - Look up phone number in phone directory

- What is 'Data Mining?

What is Not Data Mining?

- What is not Data Mining?
  - Look up phone number in phone directory
  - Query a Web search engine for information about "Amazon"

- What is 'Data Mining?

What is Not Data Mining?

- What is not Data Mining?
  - Look up phone number in phone directory
  - Query a Web search engine for information about "Amazon"

- What is 'Data Mining?
  - Credit card application [Approval/Denial]

What is Not Data Mining?

- What is not Data Mining?
  - Look up phone number in phone directory
  - Query a Web search engine for information about "Amazon"

- What is 'Data Mining?
  - Credit card application [Approval/Denial]
  - Loan application.

What is Not Data Mining?

- What is not Data Mining?
  - Look up phone number in phone directory
  - Query a Web search engine for information about "Amazon"

- What is 'Data Mining?
  - Credit card application [Approval/Denial]
  - Loan application.
  - Playing tennis based on the weather features.

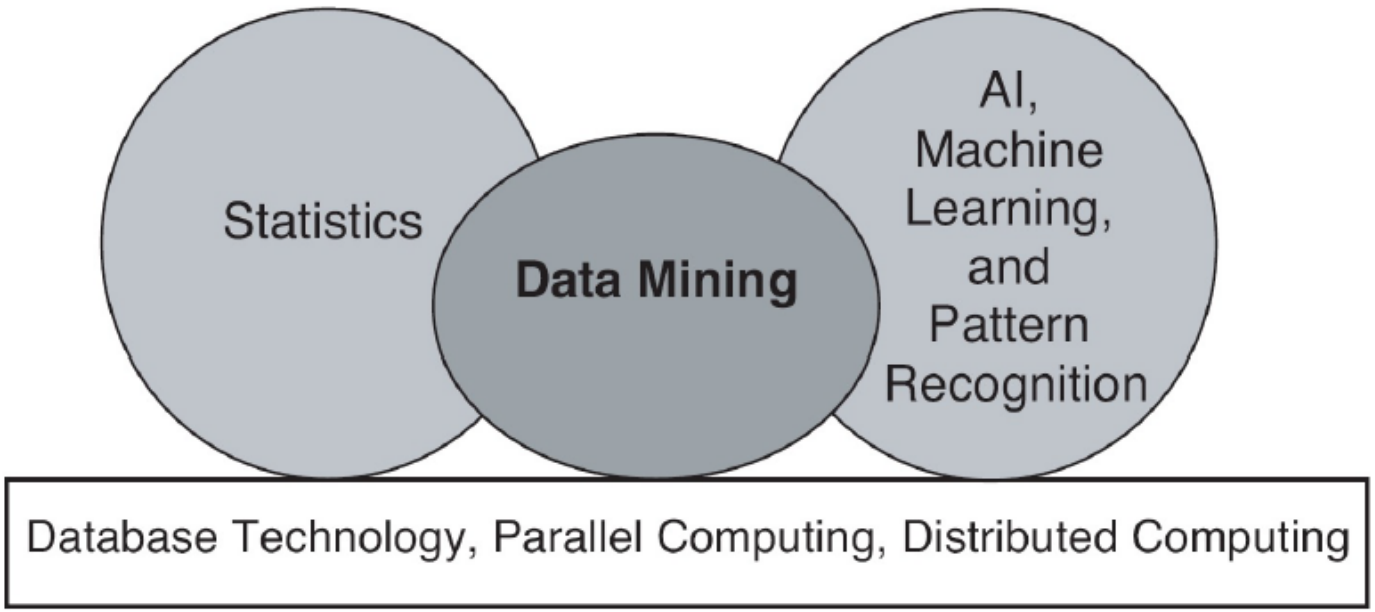
What is Not Data Mining?

- What is not Data Mining?
  - Look up phone number in phone directory
  - Query a Web search engine for information about "Amazon"

- What is 'Data Mining?
  - Credit card application [Approval/Denial]
  - Loan application.
  - Playing tennis based on the weather features.
  - Diabetes classification based on patient information

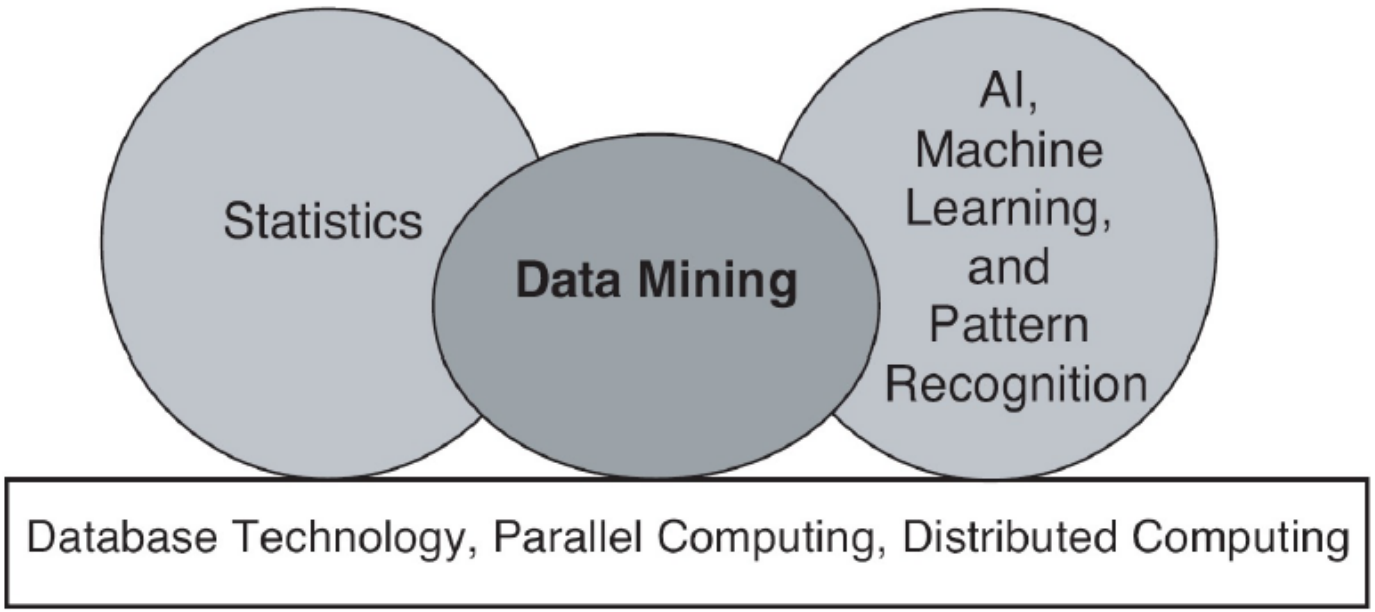
Origin of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems.



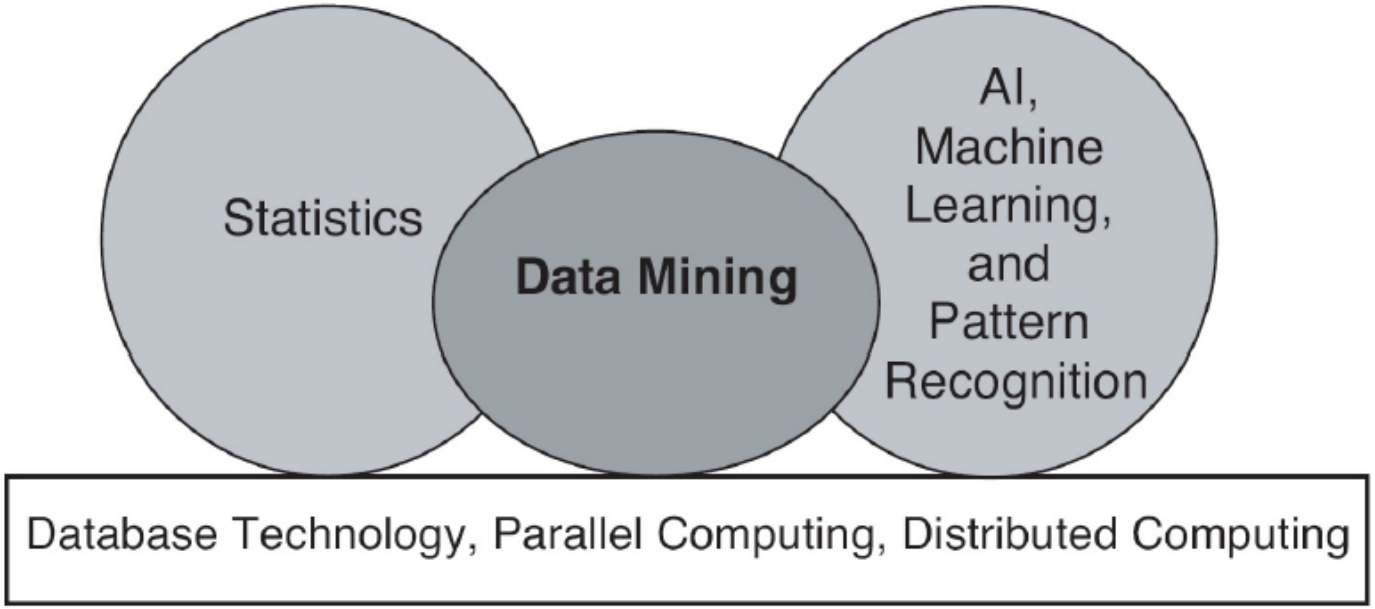
Origin of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems.
- Traditional techniques may be unsuitable due to data that is



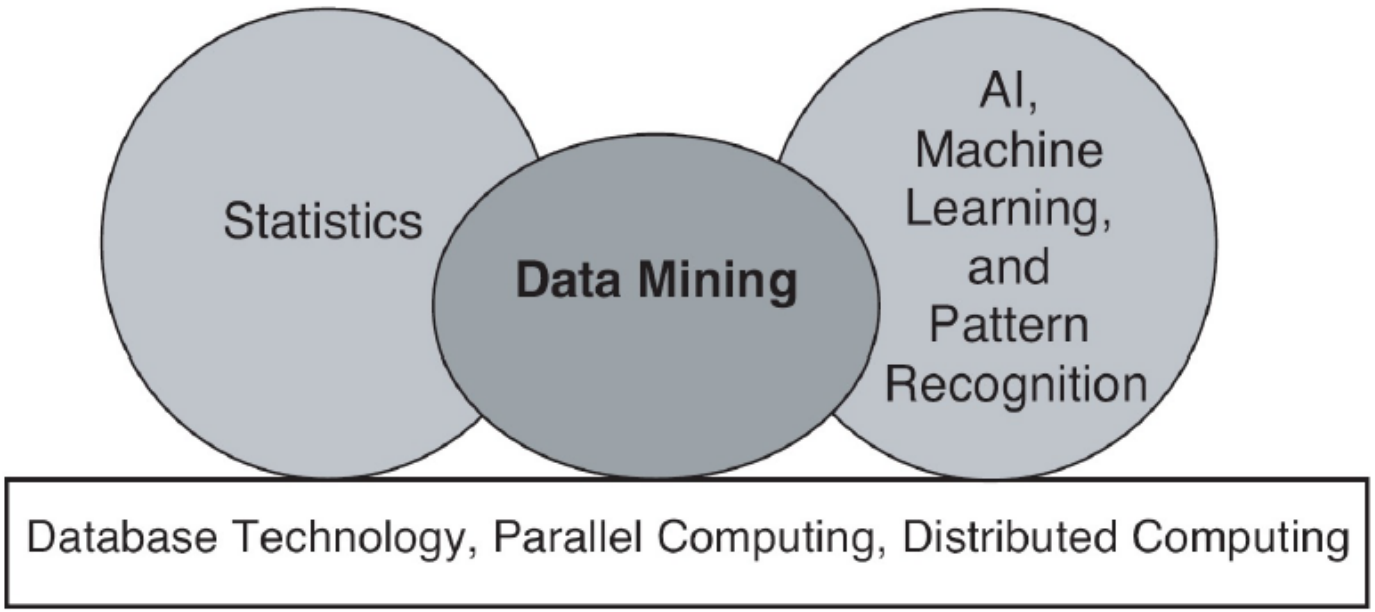
Origin of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems.
- Traditional techniques may be unsuitable due to data that is
  - Large-scale



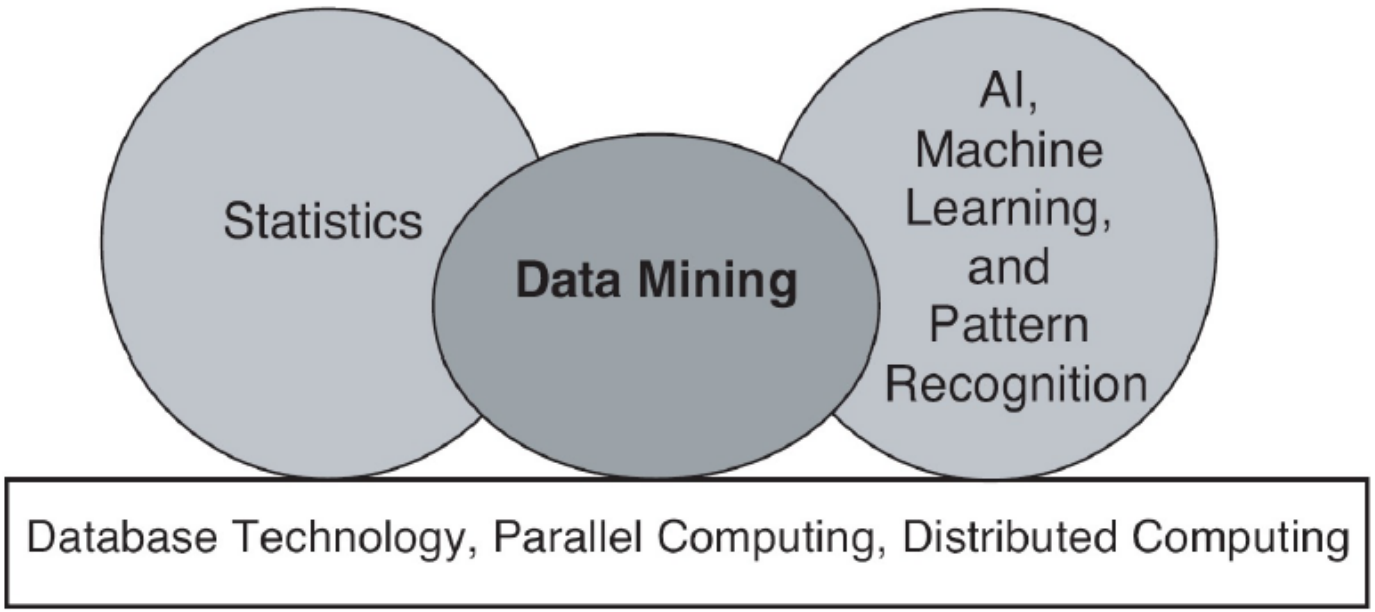
Origin of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems.
- Traditional techniques may be unsuitable due to data that is
  - Large-scale
  - High dimensional



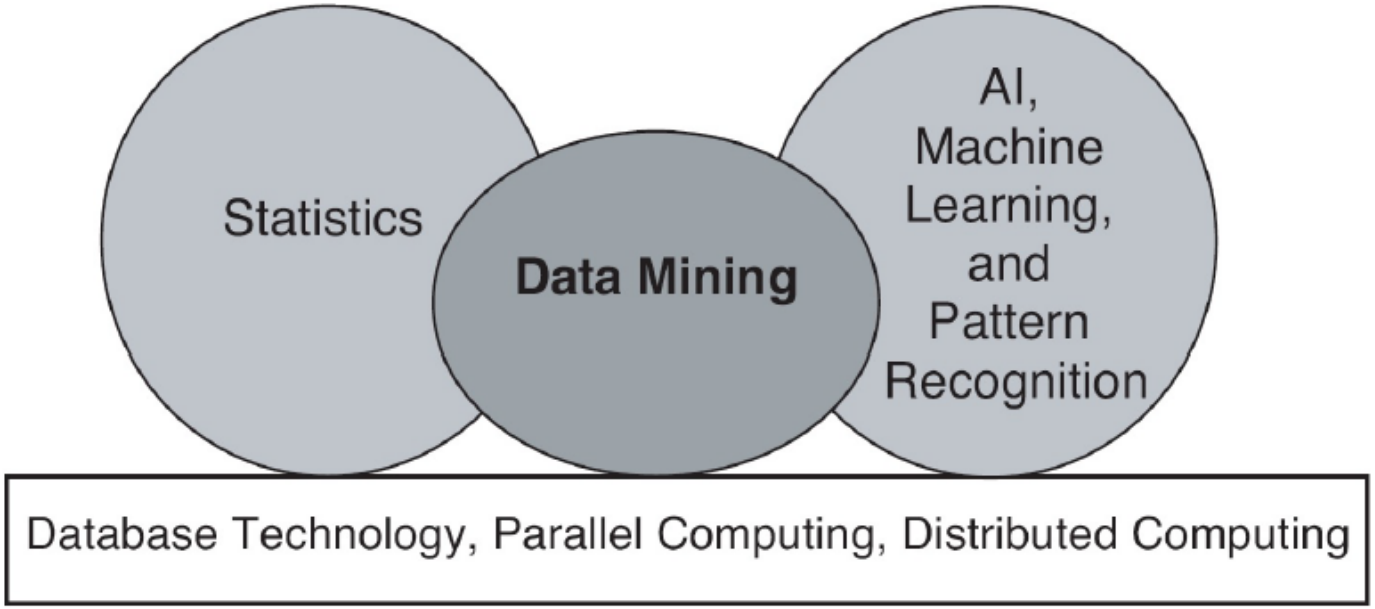
Origin of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems.
- Traditional techniques may be unsuitable due to data that is
  - Large-scale
  - High dimensional
  - Heterogeneous



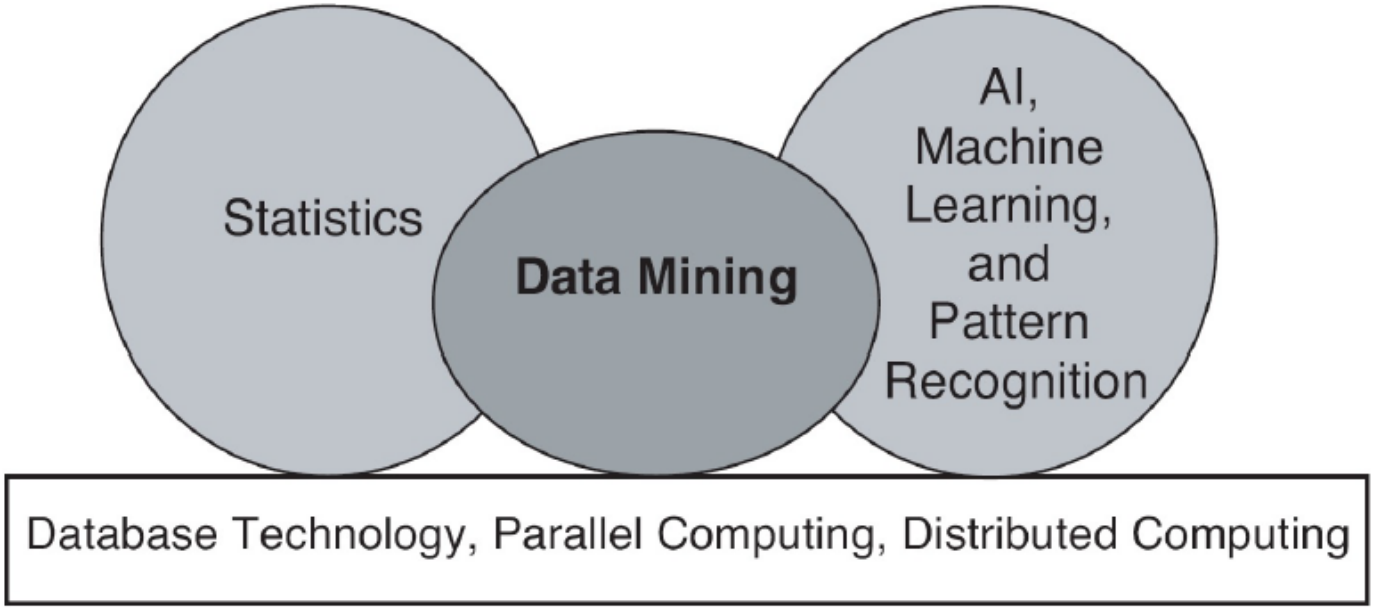
Origin of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems.
- Traditional techniques may be unsuitable due to data that is
  - Large-scale
  - High dimensional
  - Heterogeneous
  - Complex



Origin of Data Mining

- Draws ideas from machine learning/AI, pattern recognition, statistics, and database systems.
- Traditional techniques may be unsuitable due to data that is
  - Large-scale
  - High dimensional
  - Heterogeneous
  - Complex
  - Distributed

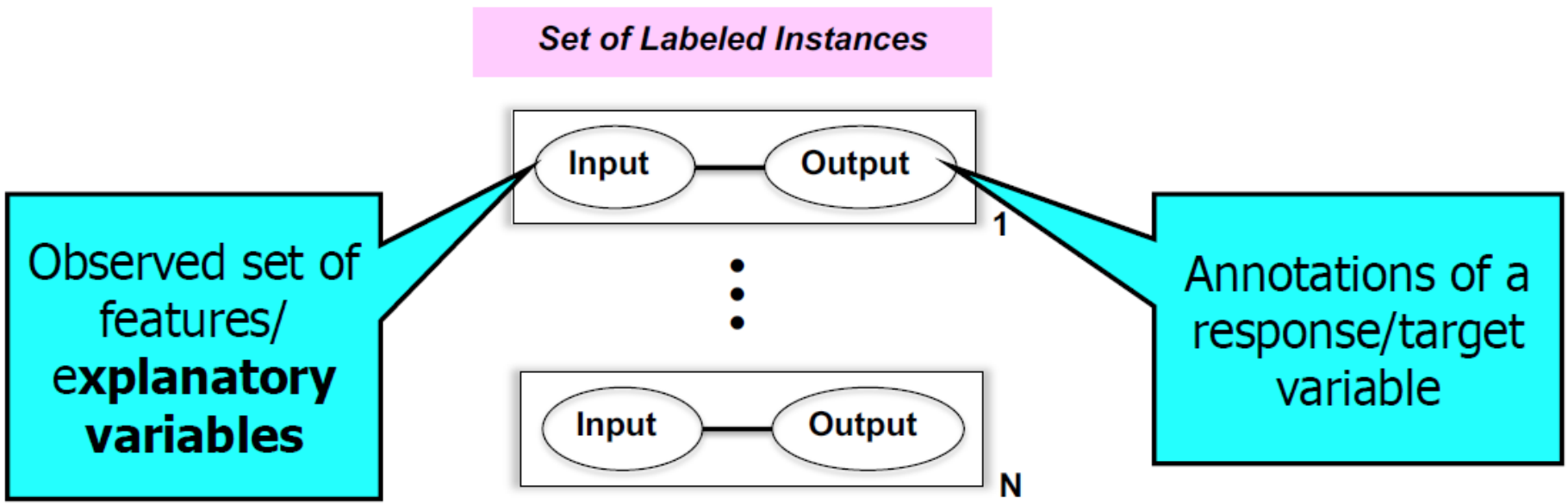


Key Areas of Data Mining

- Predictive Modeling / Supervised Learning

Big Goal

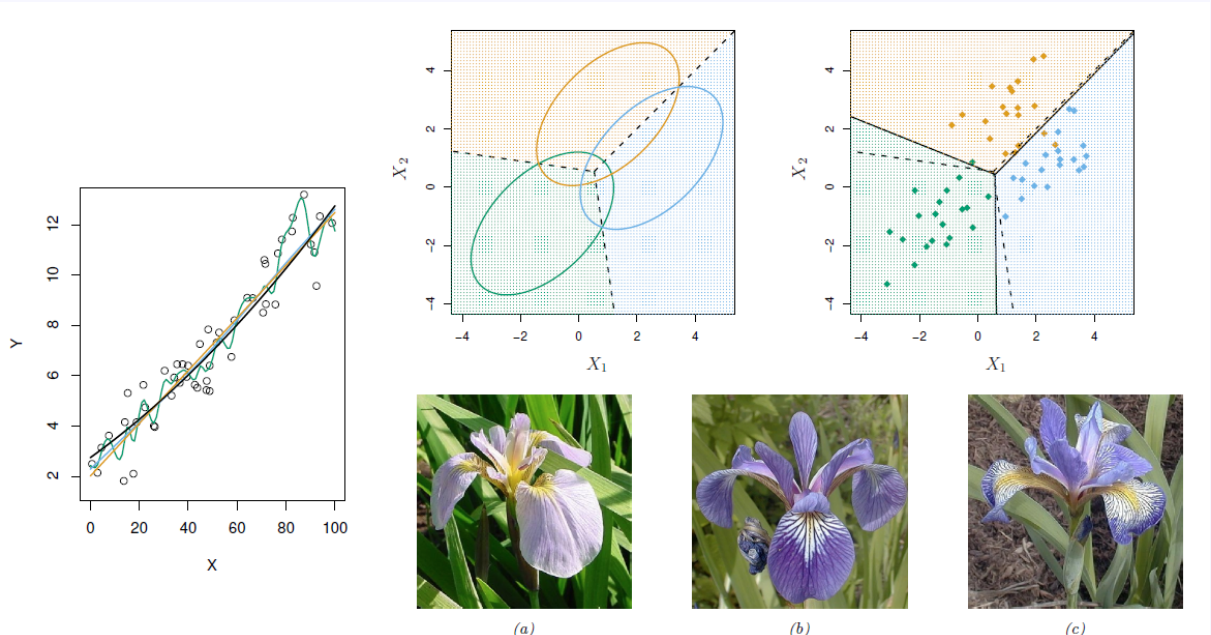
Model relationship between input and output variables to predict the output on unseen (new) instances



Key Areas of Data Mining

1- Predictive Modeling

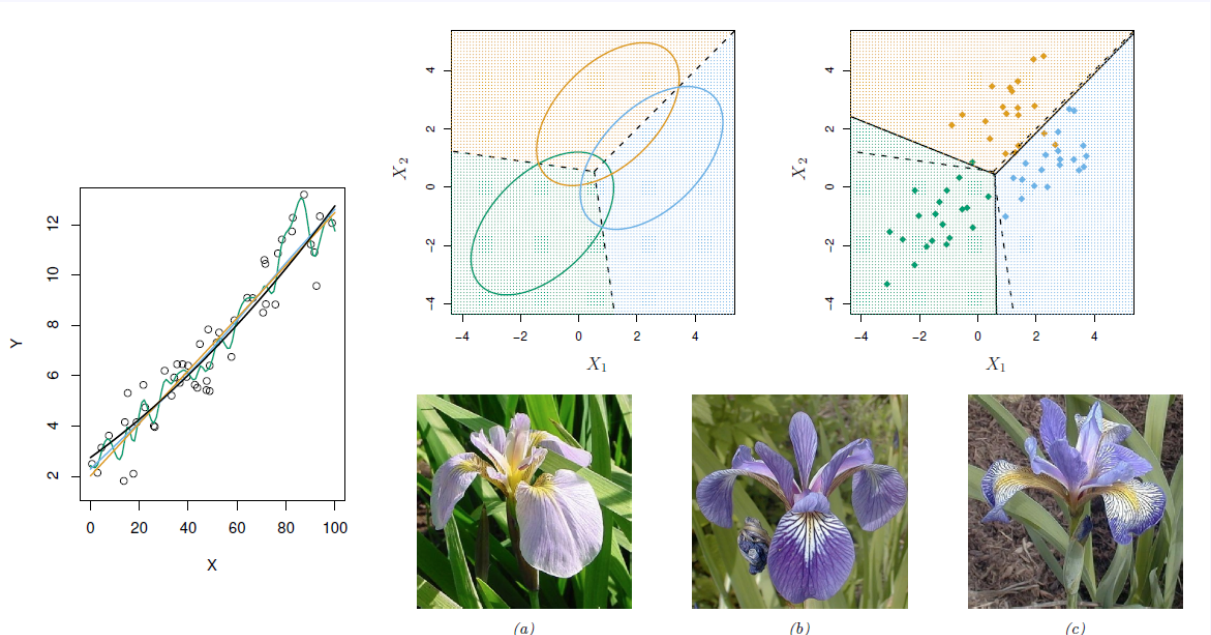
- Regression



Key Areas of Data Mining

1- Predictive Modeling

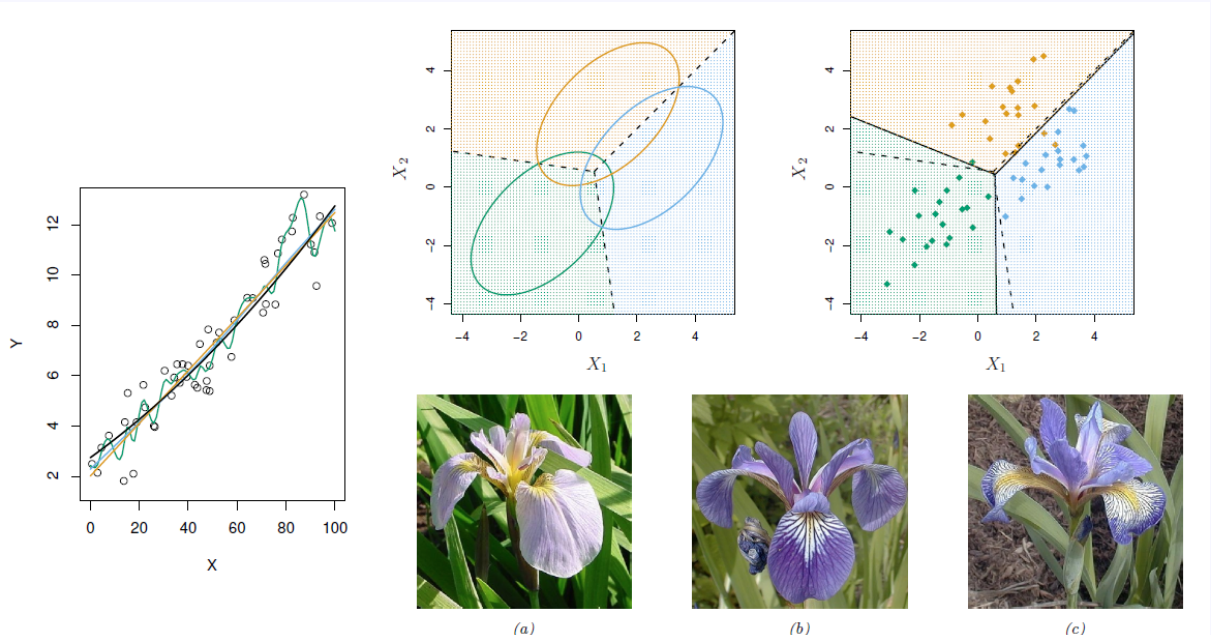
- Regression
  - Target takes continuous values



Key Areas of Data Mining

1- Predictive Modeling

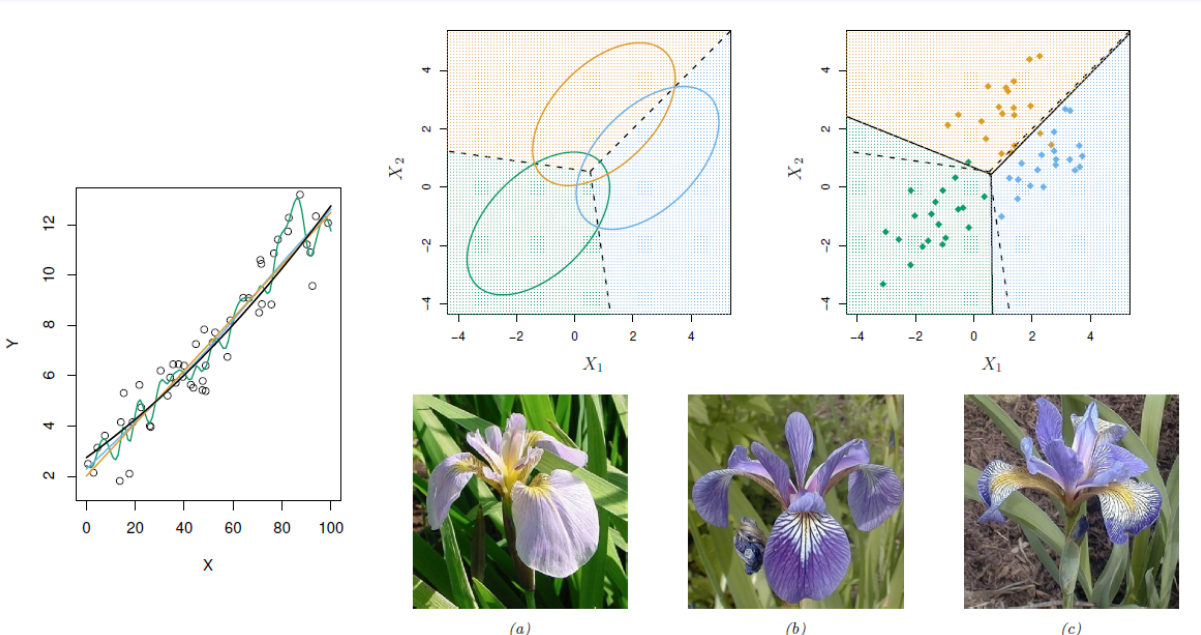
- Regression
  - Target takes continuous values
- Classification



Key Areas of Data Mining

1- Predictive Modeling

- Regression
  - Target takes continuous values
- Classification
  - Target takes discrete values/label : $\{setosa, versicolor, virginica\}$

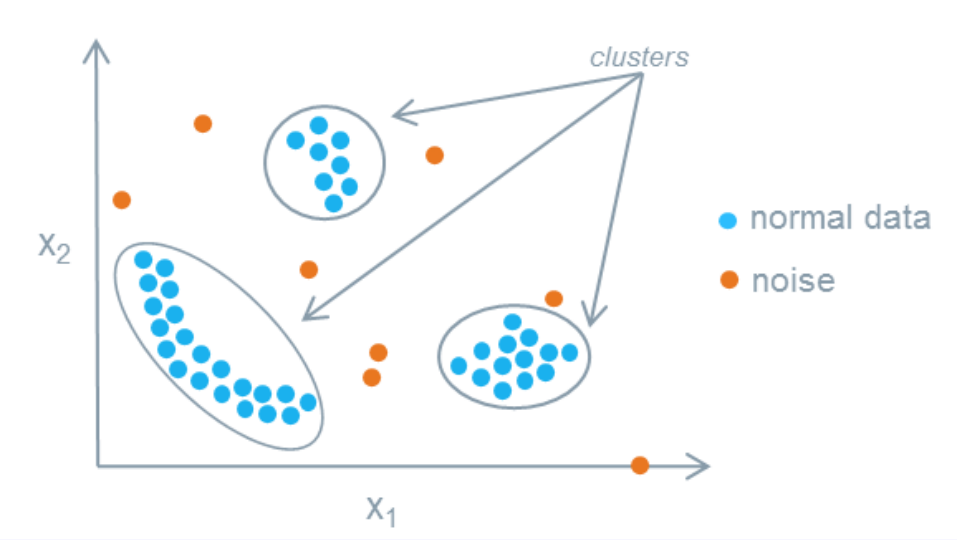


Key Areas of Data Mining

2- Descriptive Modeling/ Unsupervised Learning

Find human-interpretable patterns from "unlabeled" data

- Clustering

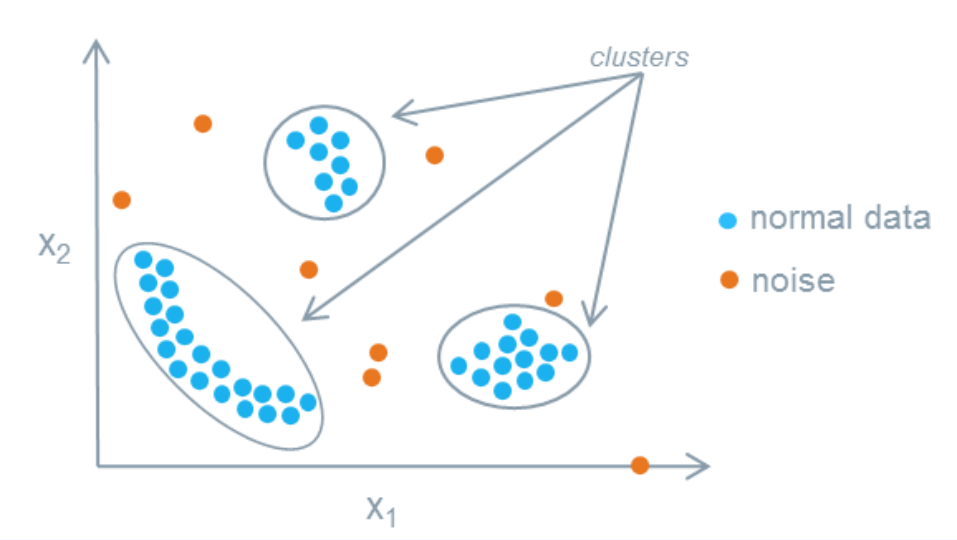


Key Areas of Data Mining

2- Descriptive Modeling/ Unsupervised Learning

Find human-interpretable patterns from "unlabeled" data

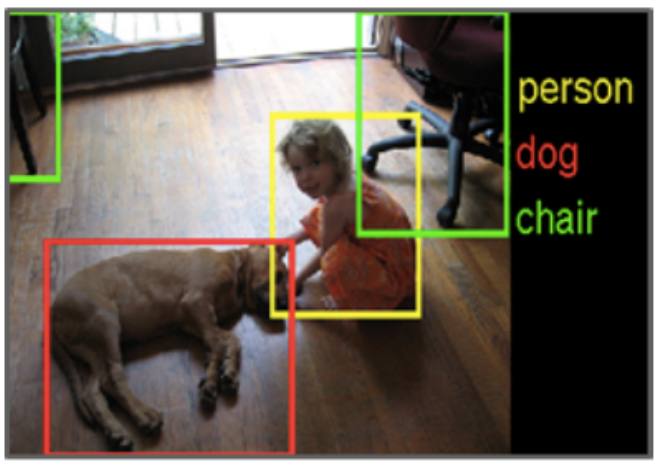
- Clustering
  - Find groups with similar properties
- Anomaly Detection
  - Find unusual instances

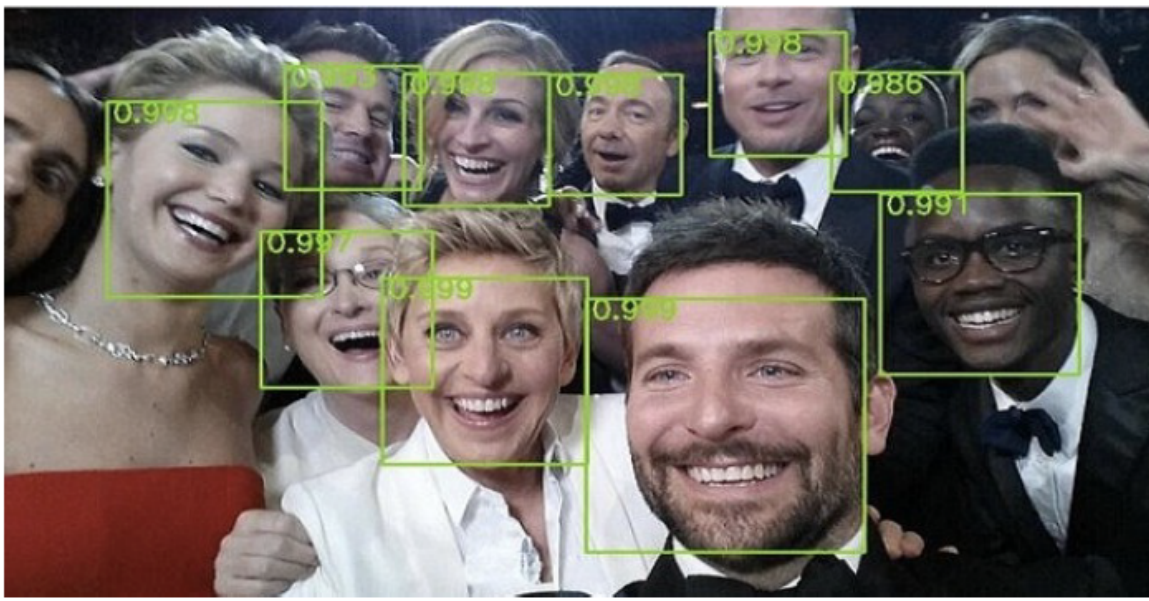


Classification: Illustrative Examples

Image Recognition

Given the pixel values of an image region (features), identify the type of object (class)





Classification: Illustrative Examples

Spam Filtering

Given the message header and content of an email (features), classify spam or no spam (class)

Land Cover Mapping

Given the multi-spectral values (features), classify land cover: water, vegetation, urban, etc. (class)

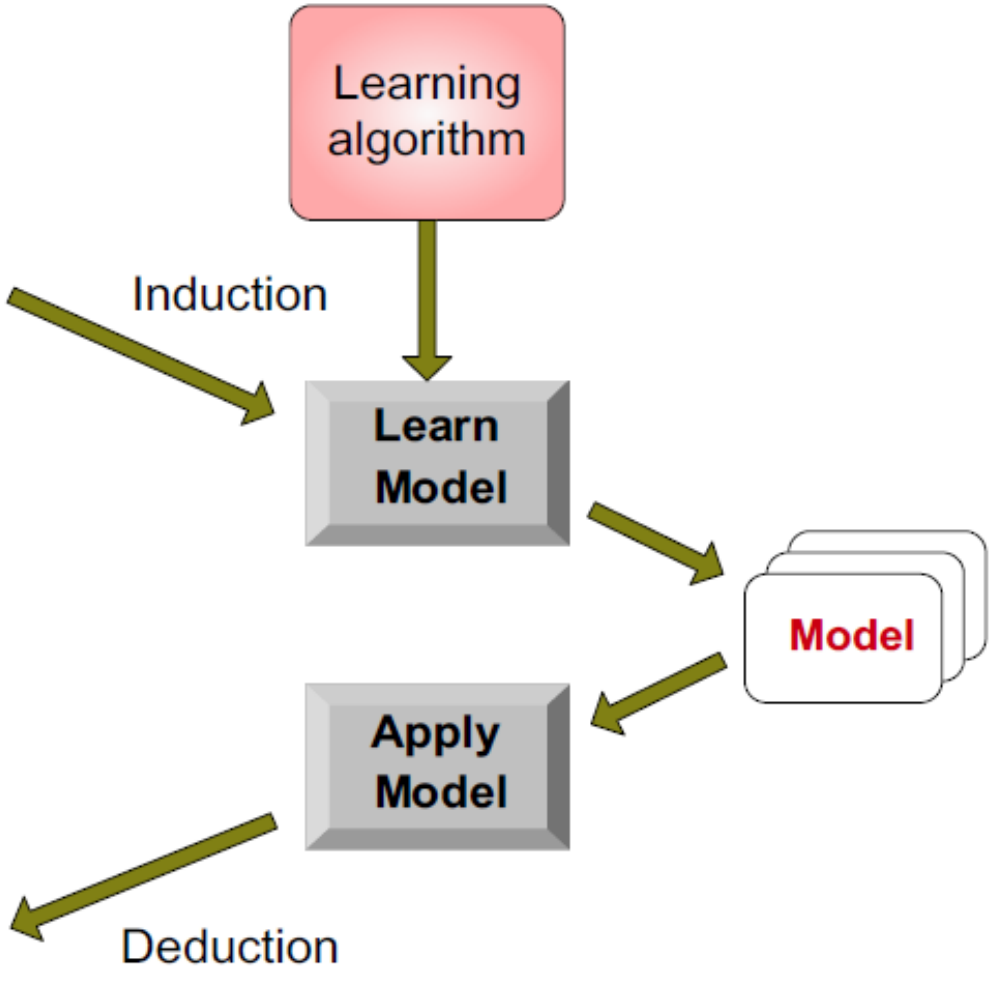
Predictive Modeling: General Approach

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



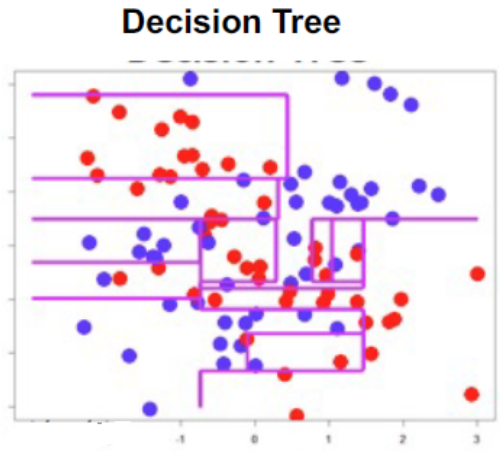
Classification Models

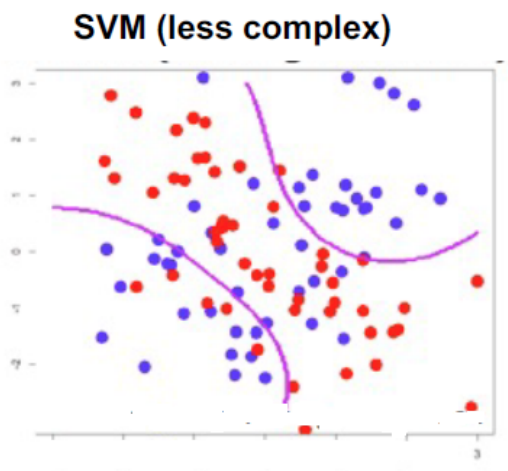
Decision Trees

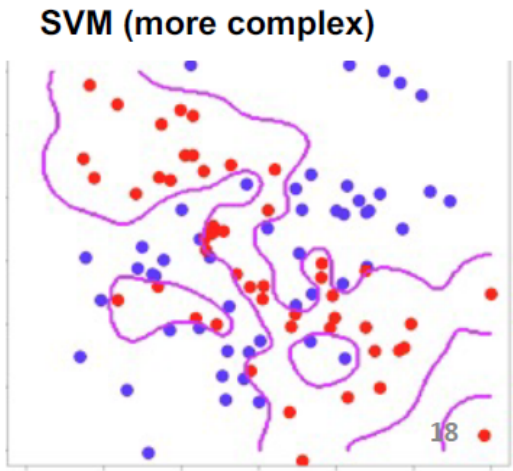
Nearest-neighbor Classifier

Naïve Bayes and Probabilistic Graphical Models

Artificial Neural Networks

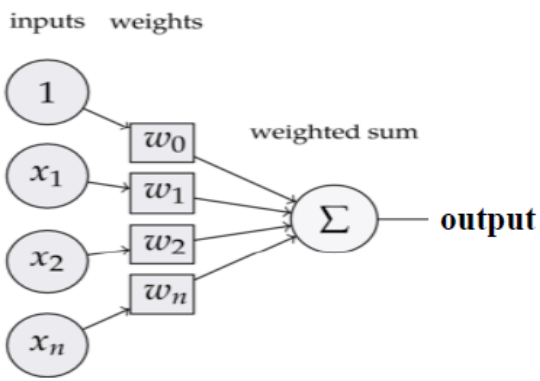


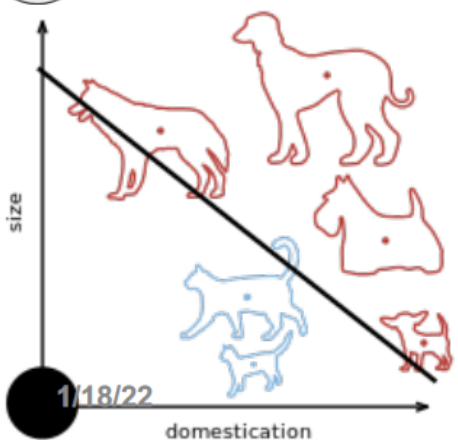




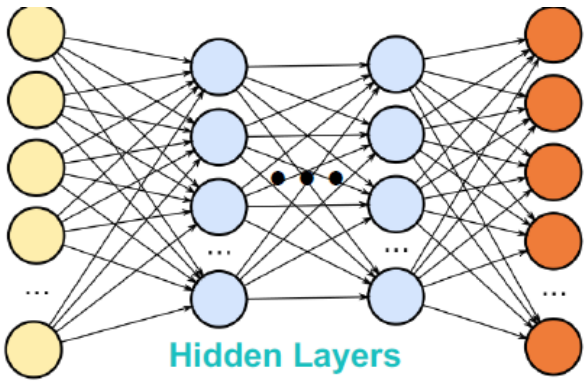
Example of Classification Model: Deep Learning

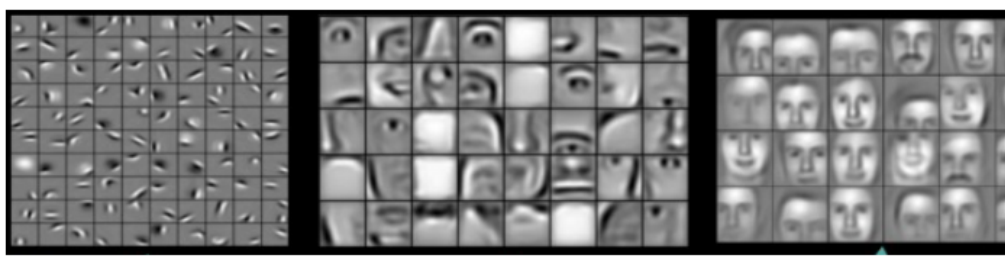
- Perceptron 1970's





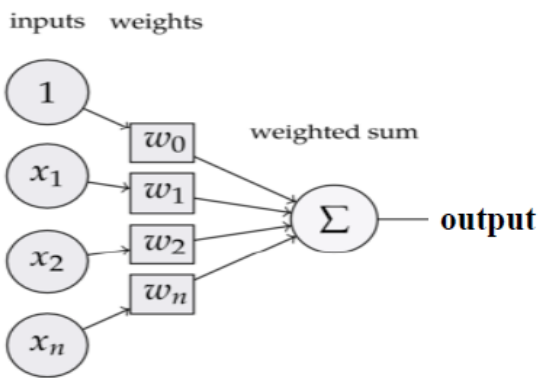
- Perceptron ~2010+

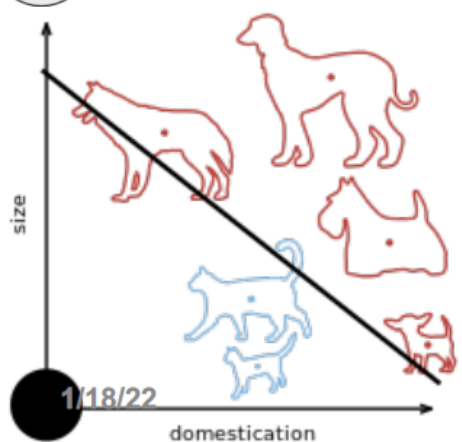




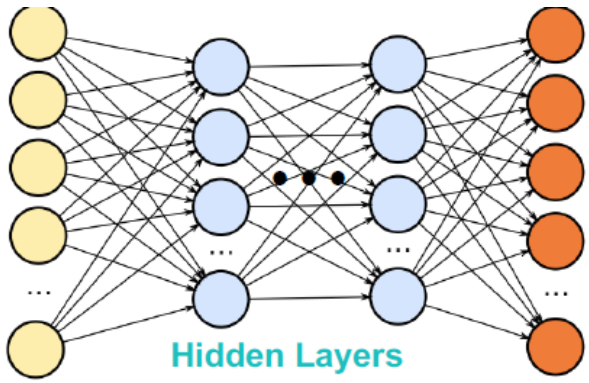
Example of Classification Model: Deep Learning

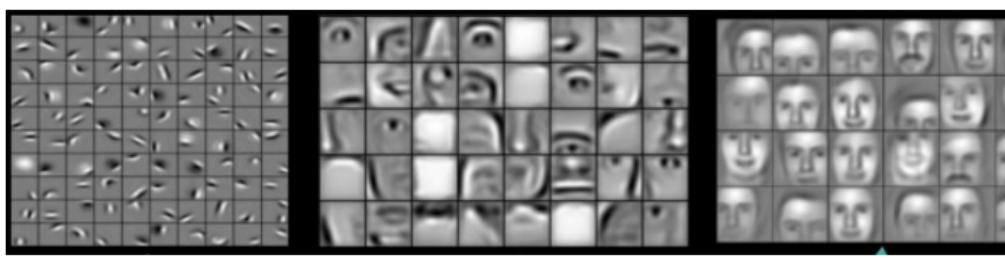
- Perceptron 1970's
- Single processing unit





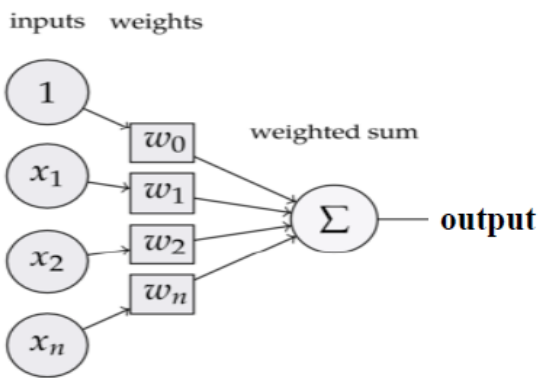
- Perceptron ~2010+
- Composition of a large number of processing units

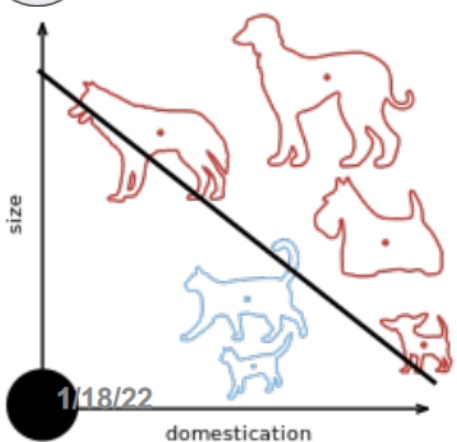




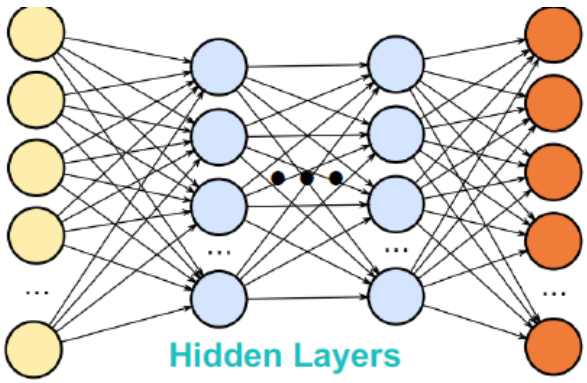
Example of Classification Model: Deep Learning

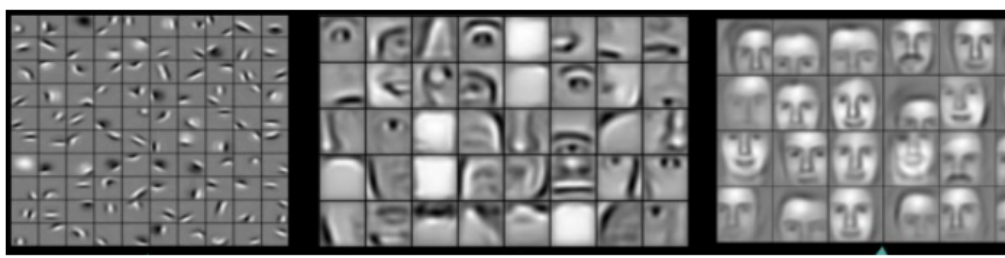
- Perceptron 1970's
- Single processing unit
- Can only learn linear decision





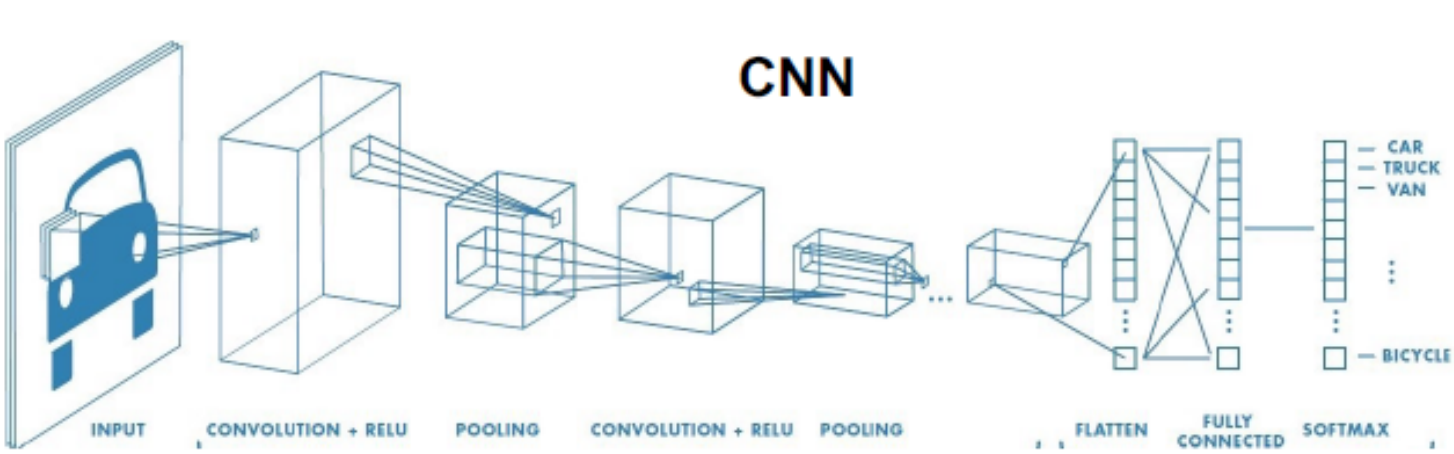
- Perceptron ~2010+
- Composition of a large number of processing units
- Can learn highly complex decision boundaries

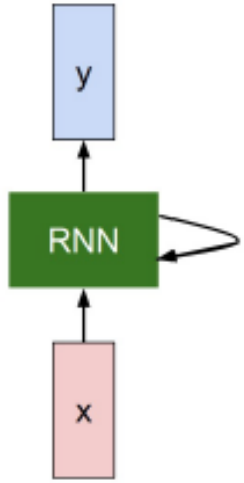




Deep Learning Topics

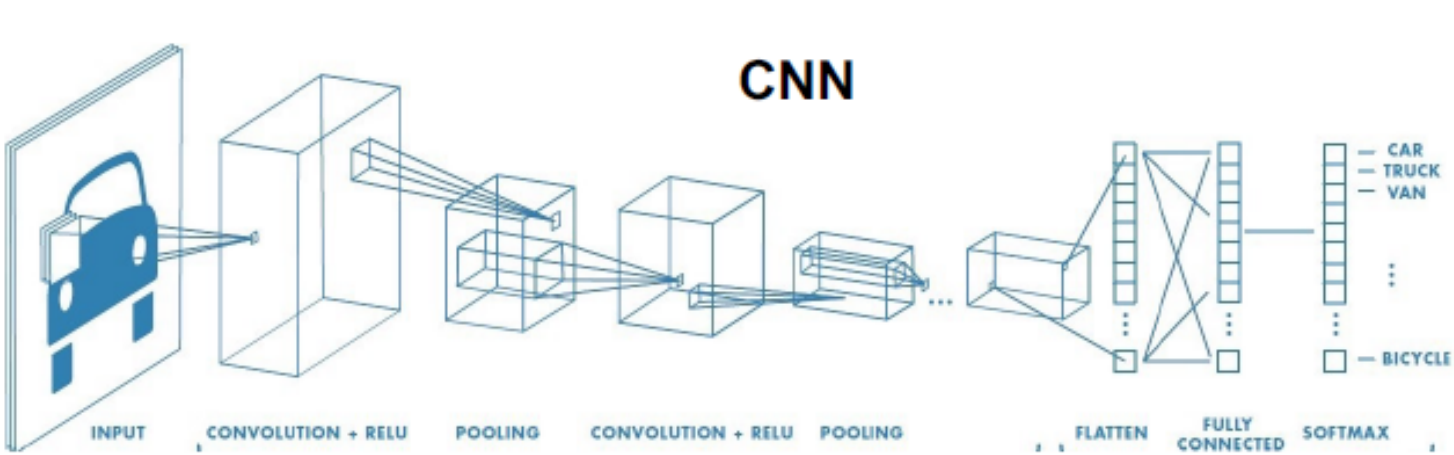
- Deep Learning architectures

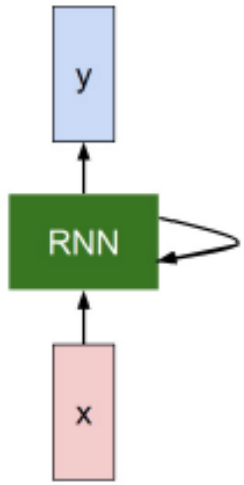




Deep Learning Topics

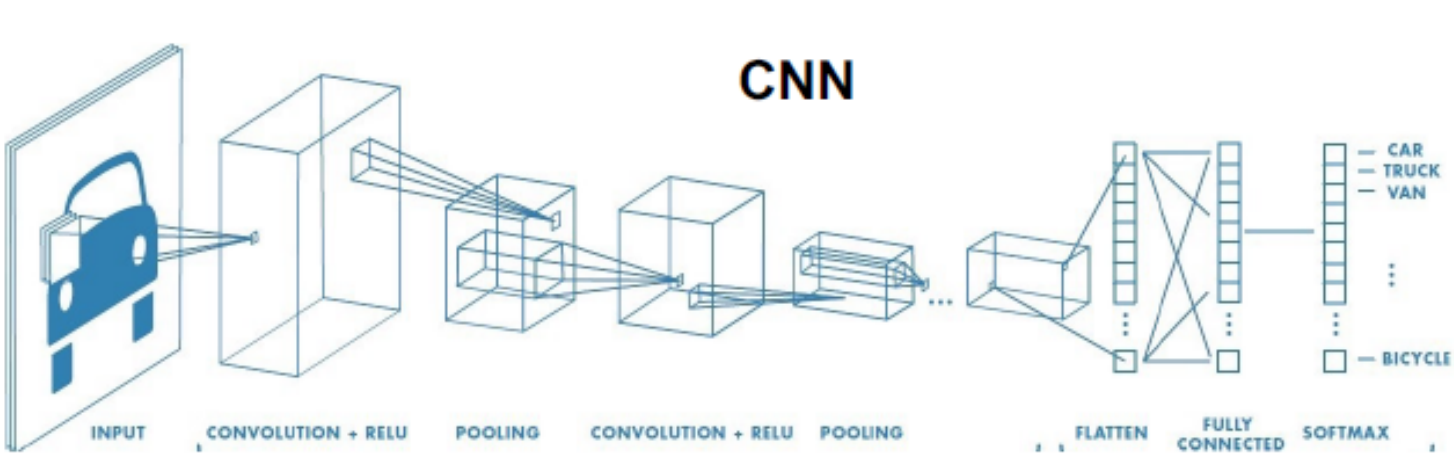
- Deep Learning architectures
  - Multilayer Perceptron (MLP)

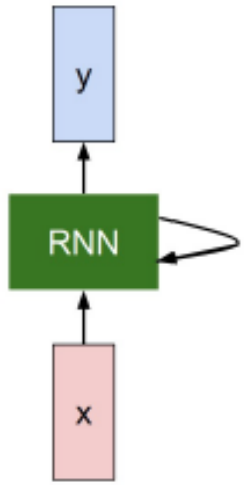




Deep Learning Topics

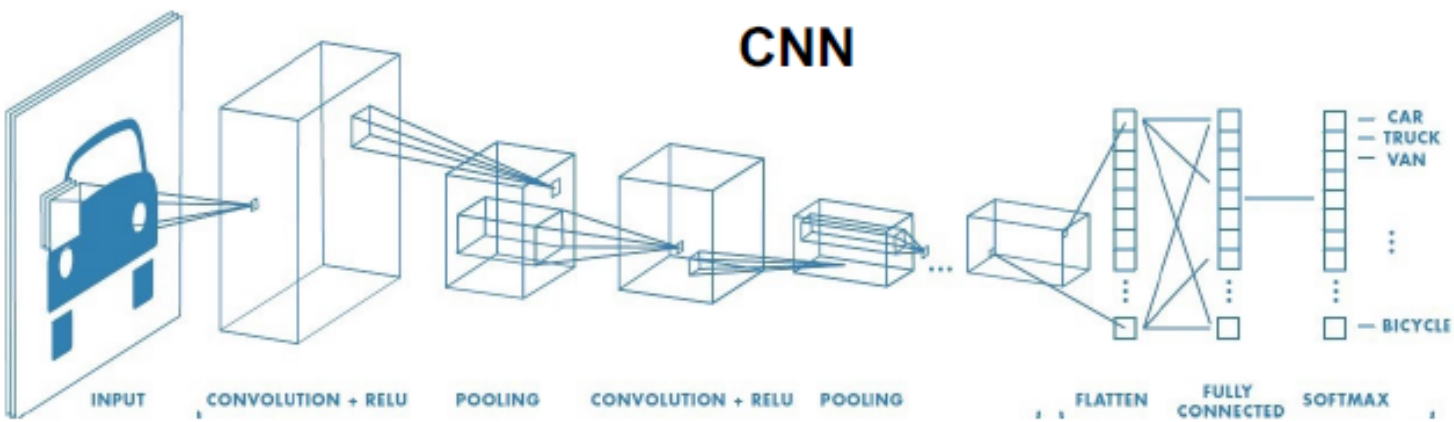
- Deep Learning architectures
  - Multilayer Perceptron (MLP)
  - Convolutional neural networks (CNNs)

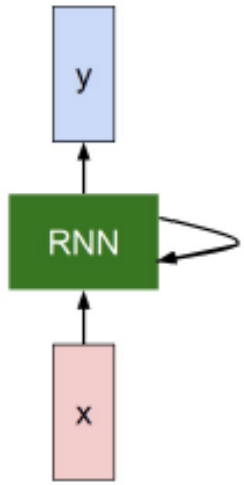




Deep Learning Topics

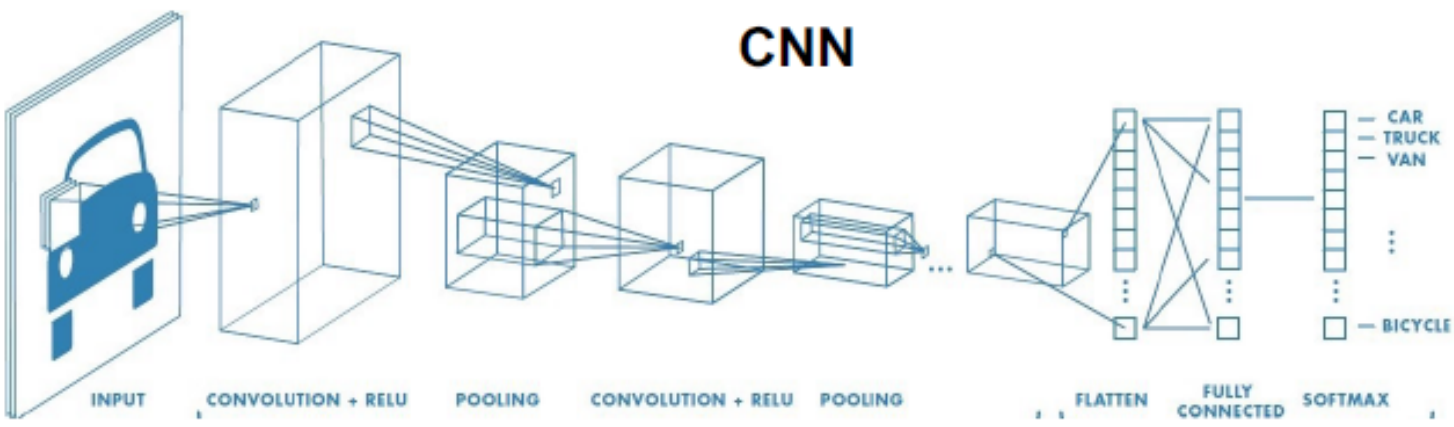
- Deep Learning architectures
  - Multilayer Perceptron (MLP)
  - Convolutional neural networks (CNNs)
  - Recurrent neural networks (RNNs)

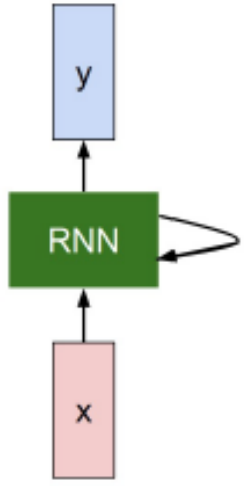




Deep Learning Topics

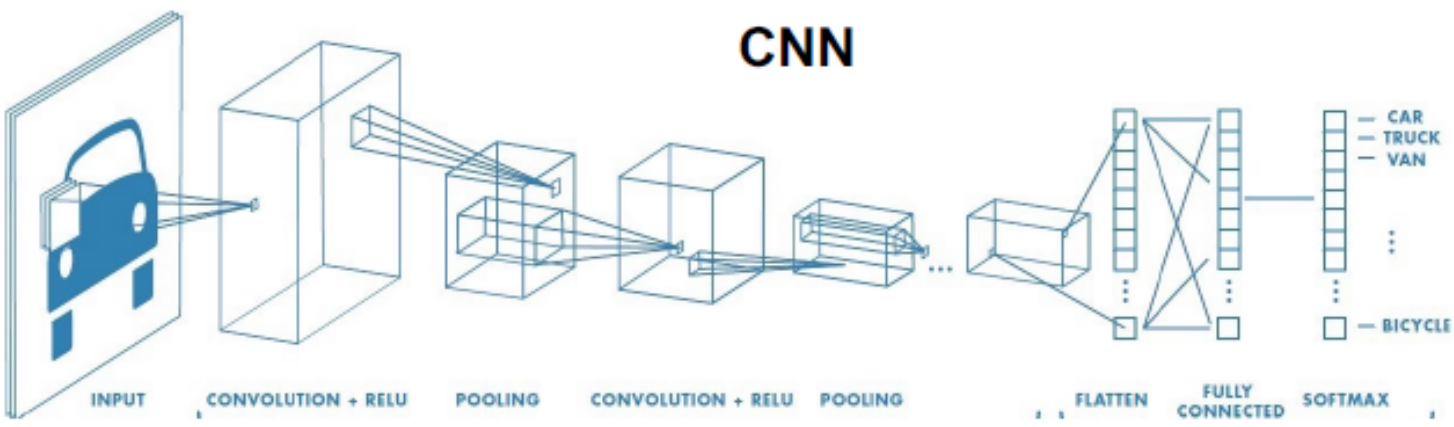
- Deep Learning architectures
  - Multilayer Perceptron (MLP)
  - Convolutional neural networks (CNNs)
  - Recurrent neural networks (RNNs)
  - Long short-term memory (LSTM)

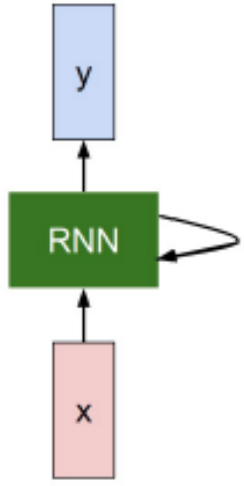




Deep Learning Topics

- Deep Learning architectures
  - Multilayer Perceptron (MLP)
  - Convolutional neural networks (CNNs)
  - Recurrent neural networks (RNNs)
  - Long short-term memory (LSTM)
  - Generative adversarial networks (GANs)





Additional Topics: Association Analysis

- Given a set of records each of which contain some number of items from a given collection

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Rules Discovered:

{Milk} --> {Coke}

{Diaper, Milk} --> {Beer}

Additional Topics: Association Analysis

- Given a set of records each of which contain some number of items from a given collection
  - Find patterns of co-occurrence of items

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Rules Discovered:

{Milk} --> {Coke}

{Diaper, Milk} --> {Beer}

Additional Topics: Association Analysis

- Given a set of records each of which contain some number of items from a given collection
  - Find patterns of co-occurrence of items
- Applications:

TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Rules Discovered:

{Milk} --> {Coke}

{Diaper, Milk} --> {Beer}

Additional Topics: Association Analysis

- Given a set of records each of which contain some number of items from a given collection
  - Find patterns of co-occurrence of items
- Applications:
  - Market-basket analysis: Rules are used for sales promotion, shelf management, and inventory management

TID	*Items*
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Rules Discovered:

{Milk} --> {Coke}

{Diaper, Milk} --> {Beer}

Additional Topics: Association Analysis

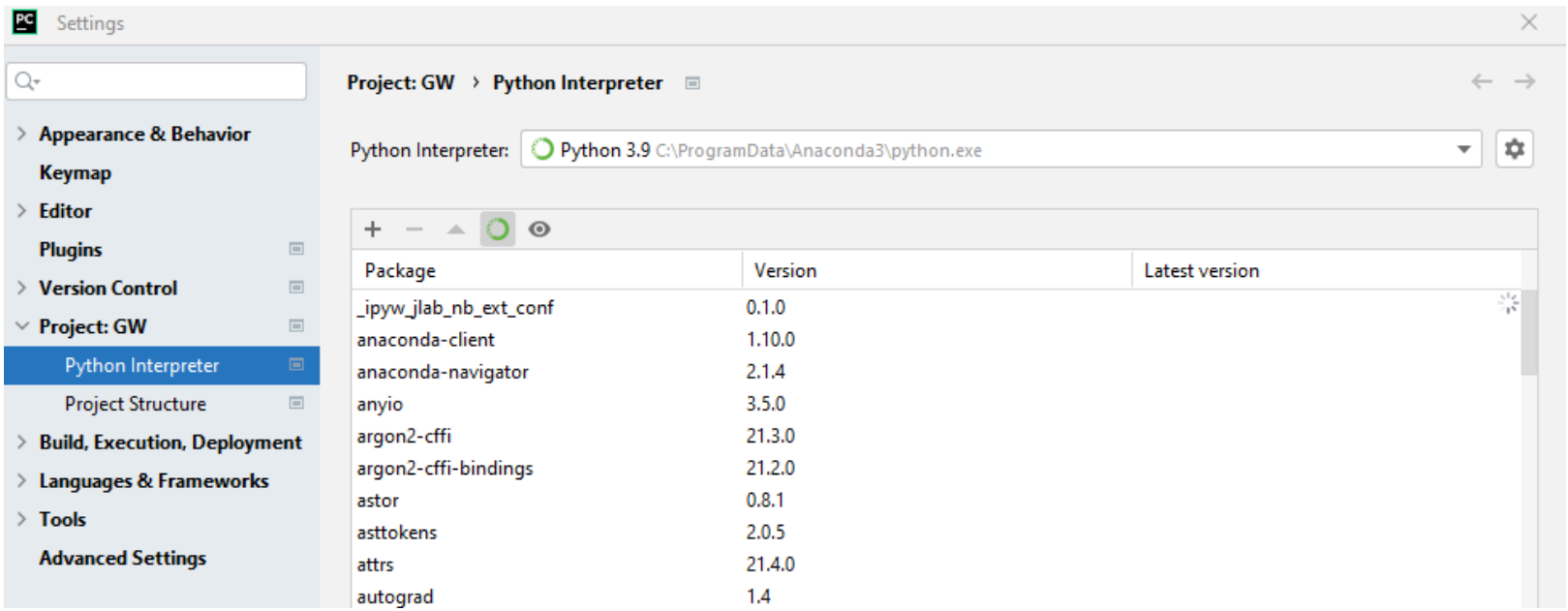
- Given a set of records each of which contain some number of items from a given collection
  - Find patterns of co-occurrence of items
- Applications:
  - Market-basket analysis: Rules are used for sales promotion, shelf management, and inventory management
  - Medical Informatics: Rules are used to find combination of patient symptoms and test results associated with certain diseases

TID	*Items*
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Rules Discovered:
{Milk} --> {Coke}
{Diaper, Milk} --> {Beer}

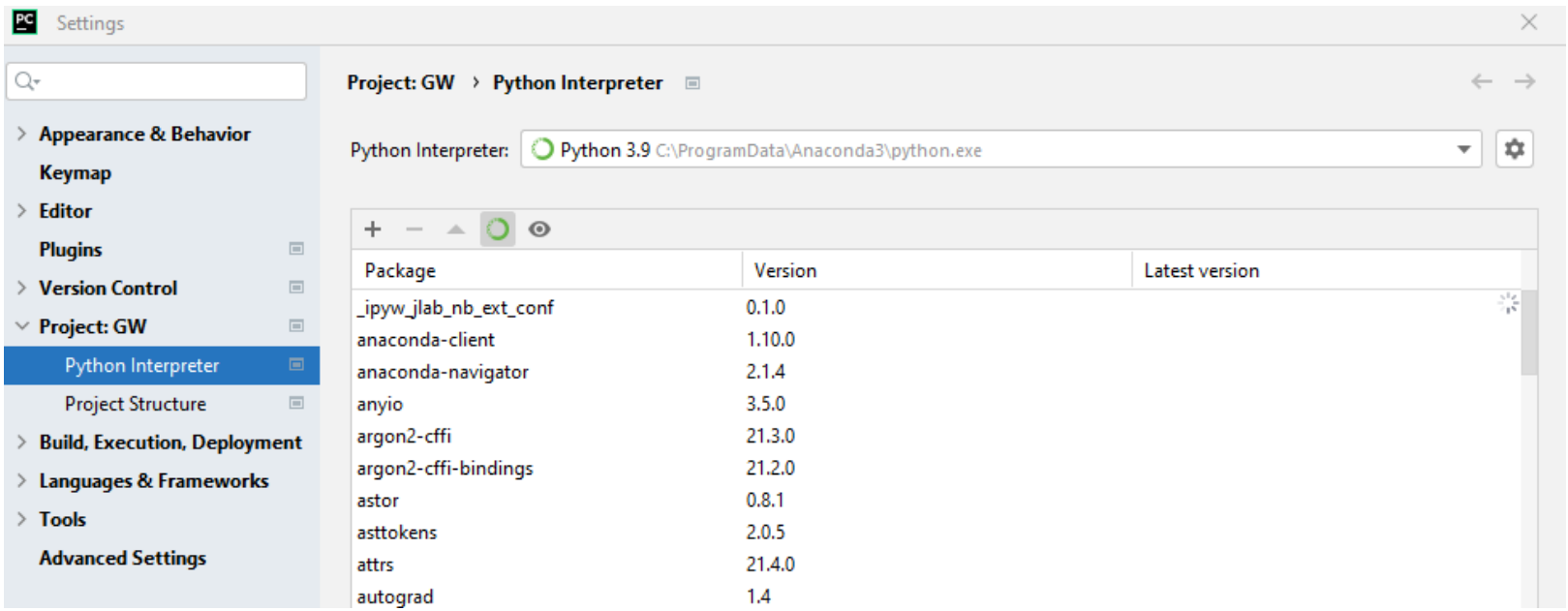
Required Software

- The main programming software for this course is python.



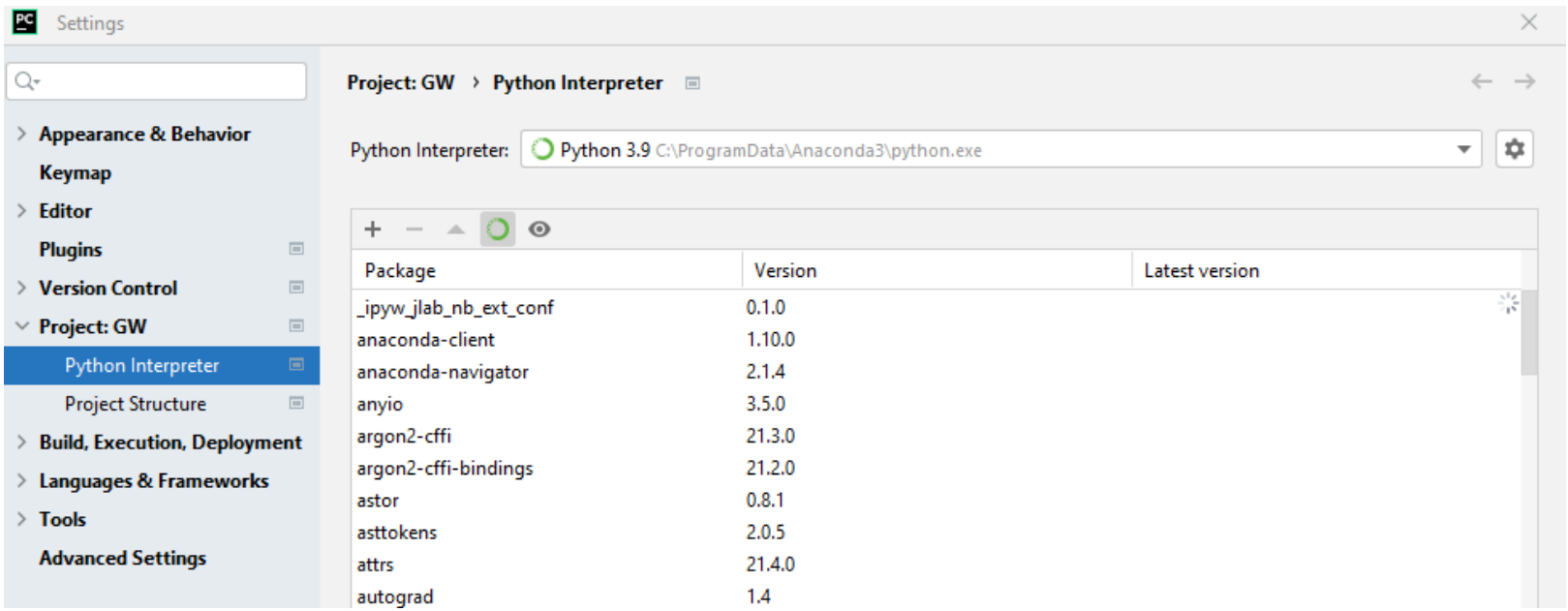
Required Software

- The main programming software for this course is python.
- The main IDE for this course is pycharm[Link].



Required Software

- The main programming software for this course is python.
- The main IDE for this course is pycharm[Link].
- You can access to the professional version of the pycharm, by creating an account using your vt.edu email.



Required Software

- The main programming software for this course is python.
- The main IDE for this course is pycharm[Link].
- You can access to the professional version of the pycharm, by creating an account using your vt.edu email.
- It is highly recommended to install Anaconda[Link] and set the pycharm interpreter as the pyhton 3.9+ using Anaconda.