Chp 1. Alpuydin. Data Mining. Ability to store vast amounts of data. Need to understand regularities in the data. Not perfect understanding. Good and useful approximations.

Examples:
- Financial: Credit / Fraud / Stock Market.
- Manufacturing: Optimization / Troubleshoot / Control.
- Medicine: Medical Diagnosis.
- Telecommunications: Network Optimization / Service.
- Science: Data Physics, Biology.
- Web: Search / Analysis.
Machine Learning in practice
programming computers to optimize a performance criterion using example data or past experience.

A model defined up to some parameters. Learning is the execution of a computer program to optimize these parameters using training data or past experience.

The model may be predictive to make predictions in future or descriptive to gain knowledge from data.

Machine Learning involves Statistics → build mathematical models with uncertainty. Make inferences from sample
Computer Science → efficient algorithms for optimization problems of learning, storing and process data. After learning → algorithms for inference, storage. Mathematics → optimization, geometric formulations.
(3) Examples of Machine Learning Applications

Learning Associations

Retail: Basket Analysis

Find associations between products bought by customers.

If people who buy X typically also buy Y - then client who buys X is a potential customer for Y.

Want an association rule

Conditional probability $P(Y|X)$.

E.g. $P(\text{chips} \mid \text{beer}) = 0.7$,

then 70% customers who buy beer will also buy chips.

More advanced - make distinctions between customers

$P(Y \mid X, D)$ - customer attributes

e.g. gender, age, marital status.

Bookseller - products are books or authors

Web portal - links to webpages, what links will user click, download pages in advance.
(4) Classification

Credit Scoring.

Bank loans money at interest. What risk is associated with loan? What probability that customer will fail/default to pay back all/part of the money.

Credit Scoring — bank calculates the risk given the amount of credit and information about customer. Attributes:

- income, savings, collateral, profession, age, financial history,

Bank has records of past loans including defaults.

Bank wants to infer a general rule codifying the association between customer attributes and his risk.

Classification problem — two classes:

(a) low-risk, (b) high-risk.

Information about customer's attributes are input to a classifier whose task is to assign the customer to one of these two classes.
(5) Classification (cont)

\[
\Theta_2 \Theta \text{ are data instances.}
\]

"+" classified as low-risk

"-" classified as high-risk.

Example: two attributes (for simplicity)

Savings, Income

Example classification rule:

IF income > \( \Theta_1 \), AND Savings > \( \Theta_2 \)

THEN low-risk

ELSE high-risk.

This is an example of a discriminant function that separates the data into two classes.

Main application is prediction. If future is similar to the past, then we can make prediction for novel instances.

In some cases, instead of 0/1 decision, we may want to calculate prob. \( P(Y|X) \) (i.e. learn an association).
Pattern Recognition Examples:

- Optical Character Recognition (OCR) recognizes character codes from images.
- Variability in writing styles and redundancy of language—successive characters are not independent but are constrained by the words of the language.

Medical Diagnosis

Inputs are relevant information about the patient and the classes are the illnesses.

- Input - Age, gender, medical history, symptoms
- Some tests may not have been applied and these inputs are missing.
- Tests are expensive, take time and are only worth applying them to gain valuable information. Wrong decision is very bad—classifier must take this into account.

Speech recognition.
(7) Knowledge Extraction.
learning a rule from data also gives knowledge extraction.

The rule is a simple model that explains the data — looking at this model gives an explanation for the process underlying the data.

This knowledge can be used — e.g. to advertise to low-risk customers for bank loans.

Learning also performs compression, since we get an explanation that is simpler than the data. It requires less memory to store and less computation to process.

(If you know the law of addition, you don't need to remember the sum of all possible pairs of numbers.)

Outlier detection — find instances which do not obey the rule and are exceptions.

→ e.g. to detect anomalies requiring attention (e.g. fraud).
(8) **Regression**

System to predict the price of a used car

- Inputs — car attributes — brand, year, engine capacity, mileage, and so on.

- Output — price of car.

Regression — predict $y$ from $x$ using training data.

Fit a function — e.g. $y = ax + b$

Regression (and classification) are supervised learning problems

Models $y = g(x \theta)$ & parameters

$y \in \{0, 1\}$ classification $g(.)$ discriminant function

Machine learning — optimizes for parameters $\theta$ to minimize a performance measure.

Other application (most of statistics?)

- Navigation by mobile robots, autonomous car.
- Machine to roast coffee
  - Inputs — temperature, time, coffee bean
  - Output — consumer satisfaction.
(9) Two other to ICS:
this course will put less emphasis on these.

(A) Unsupervised Learning → no supervisor aim is to find regularities in the data.

*Clustering* → find groups / clusters of

e.g. inputs are attributes of customers.

- provides natural grouping of customers (e.g. to give $P(I|X, D)$)

image compression → efficient ways to store images

image “vocabulary”

bioinformatics → DNA / amino acids are series of bases $A, G, C, T$.

alignment problem → match difficult sequences
difficult problem → ambiguity

clustering learn motifs which are sequences of amino acids that regularly occur.→ correspond to structural, functional elements.→ less ambiguous to match.
Reinforcement Learning.

Output of system is a sequence of actions.

Want to determine the policy—the sequence of actions to reach the goal.

Game playing—backgammon—Gerald Tesman
Robot navigation.

Overall, the importance of Machine Learning will increase. More and more data is becoming available. Machines are becoming better at analysing it than humans.

History. Artificial intelligence—tried to make computers think like humans. Neural networks—basis of neural networks lies in statistics. (What do neurons do?) What results is a merge between C.S. (logic) and statistics.