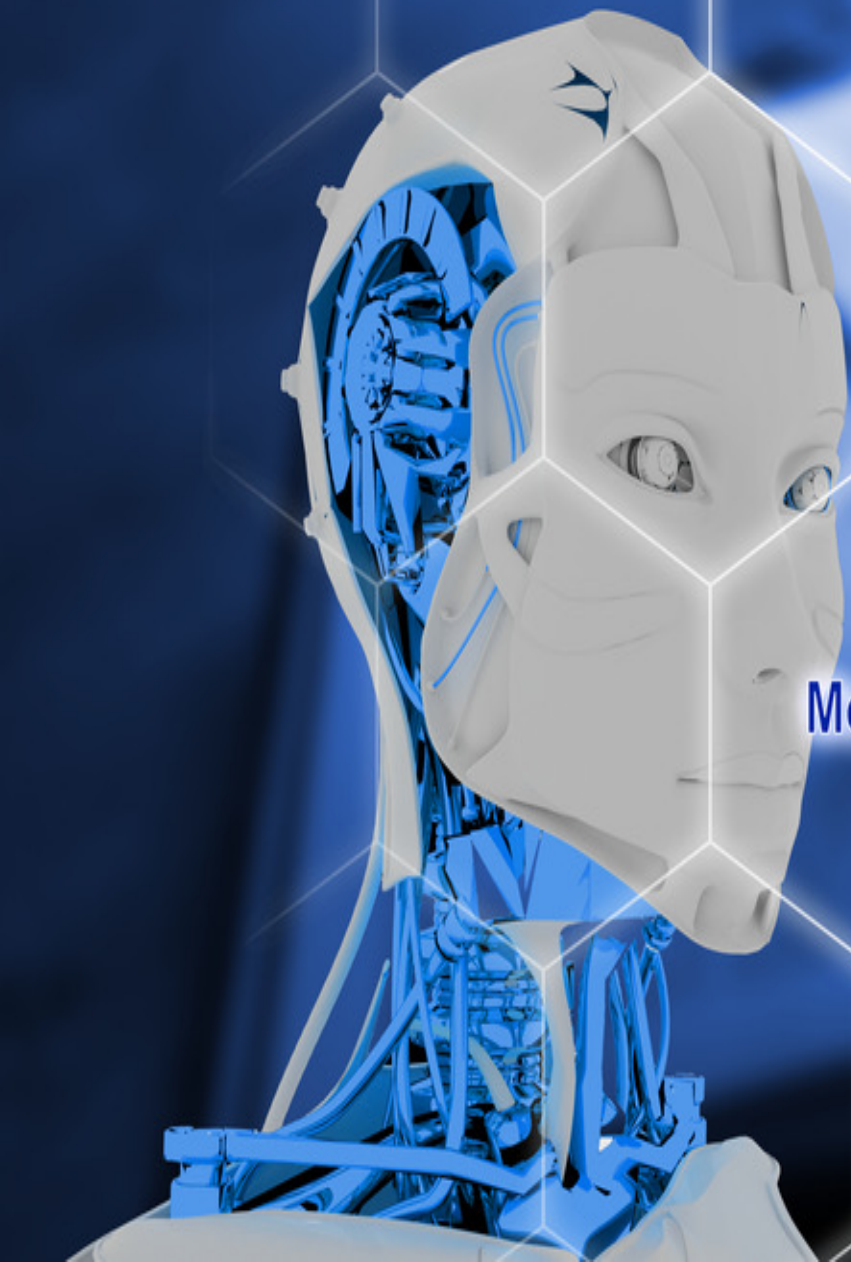
Dr. Korris Chung
Dept of Computing



Big data

Artificial intelligence

COMP4432

Machine learning

Neural network

Theory

Model

Algorithms

Data mining

Science

Examples

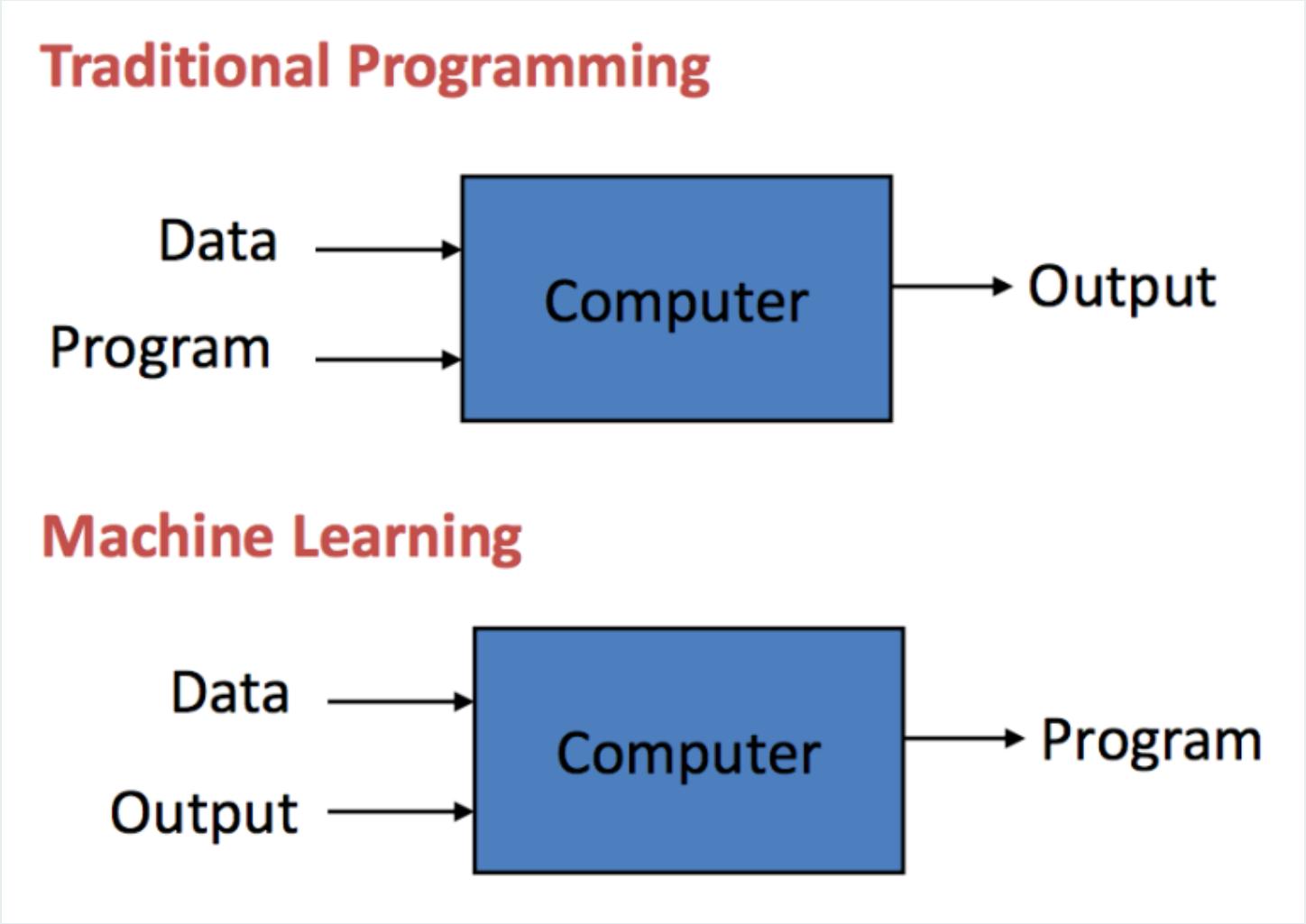
Roadmap

- Basic Concepts
  - Data vs Feature vs Model
  - Theoretical/Mathematical aspect
- Machine Learning - Why? What? and Where?
- Related Disciplines
- ML Models
  - Supervised Learning
  - Unsupervised Learning
- Issues & Resources
- Take-home messages!

Basic Concepts

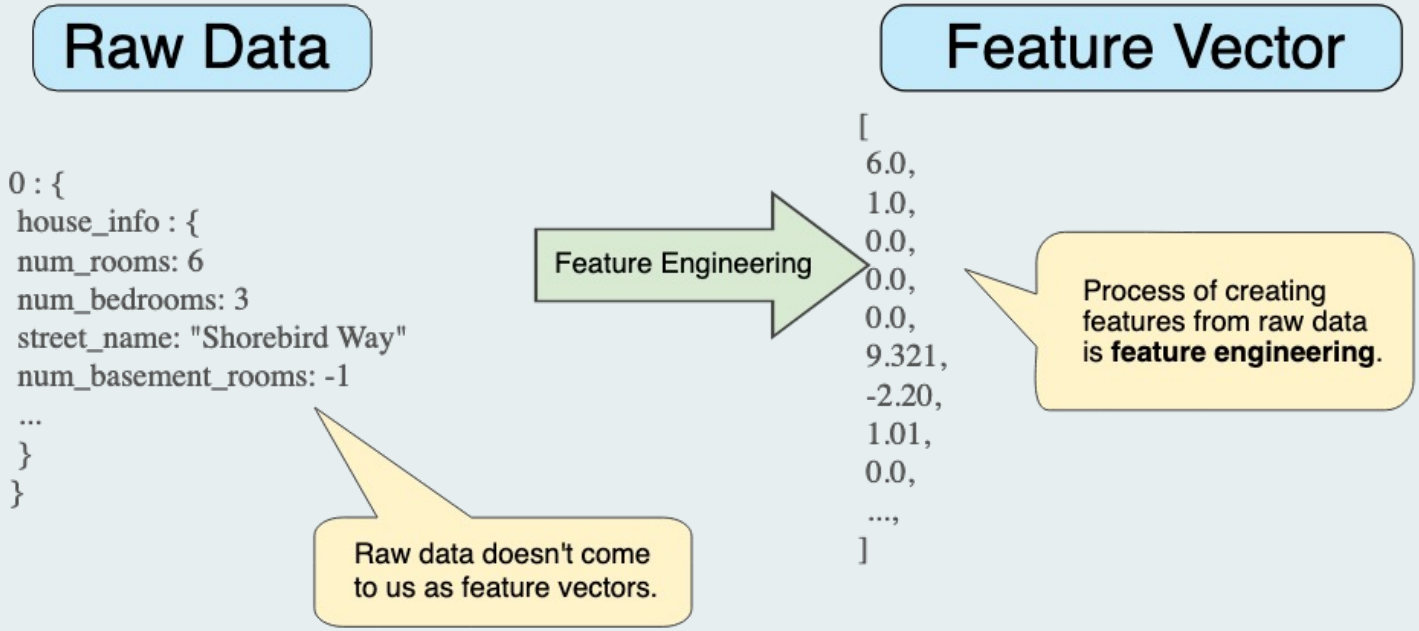
(to be learned) in this course

Basic Concepts



Really Important Views/Concepts

- Data vs Feature

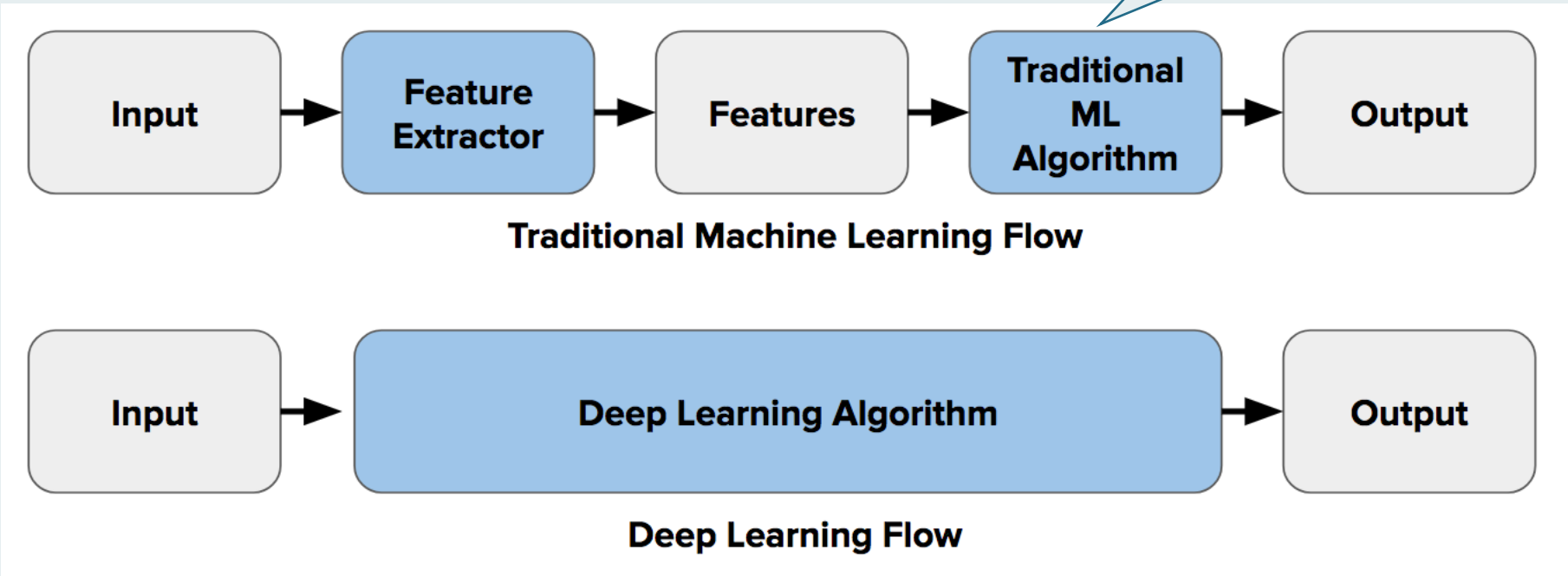


Could be a major data analytics work --- Feature engineering (FE)!
If your model doesn't work, it may not be due to the model itself (e.g. simple model, bad training). It could be due to ineffective FE!

Really Important Views/Concepts

- Data vs Feature vs Model

ML concerns with the model itself, assuming appropriate feature input and desired output.



Deep learning model could step in to carry out feature learning (not FE) and "machine learning"!!!

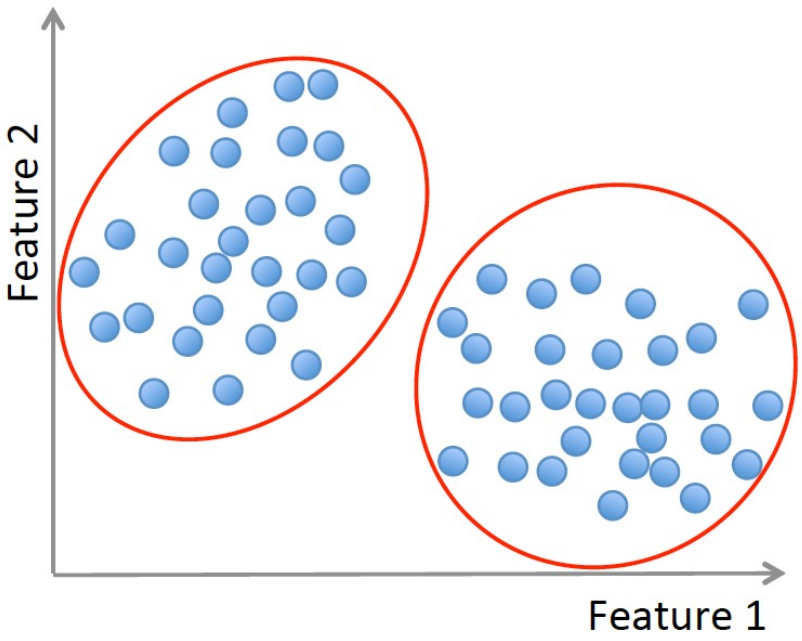
Really Important Views/Concepts

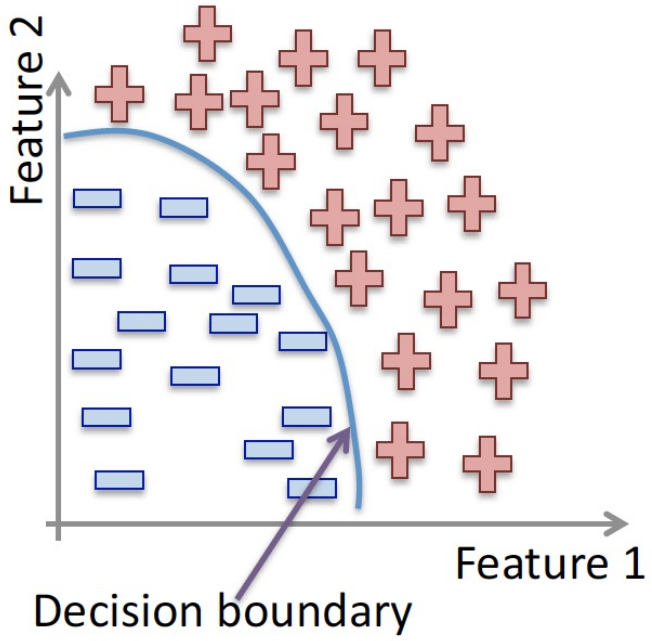
If you can comprehend these concepts and manipulate them sophistically, you could be very successful in any ML scientist/developer/engineer job and/or Data Scientist/Analytics job!

Probably, you can drop this course!

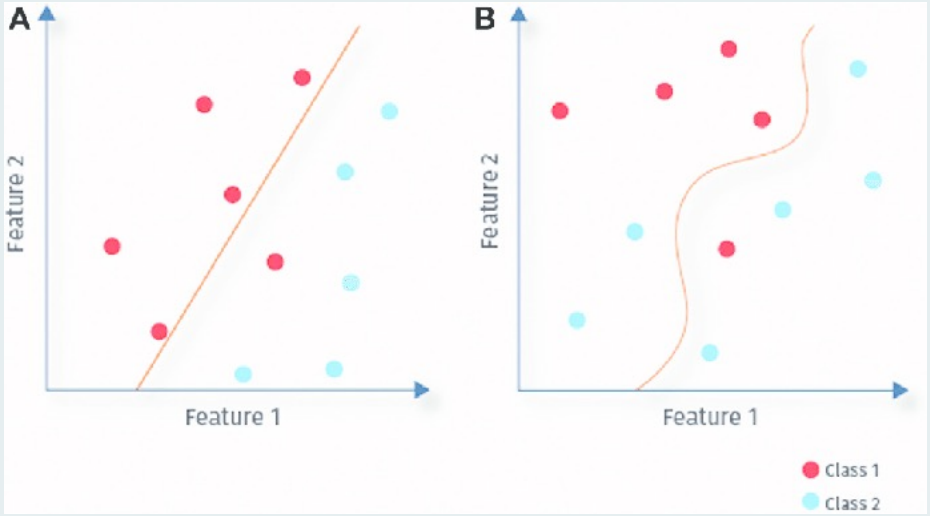
Let’s go a bit deeper!

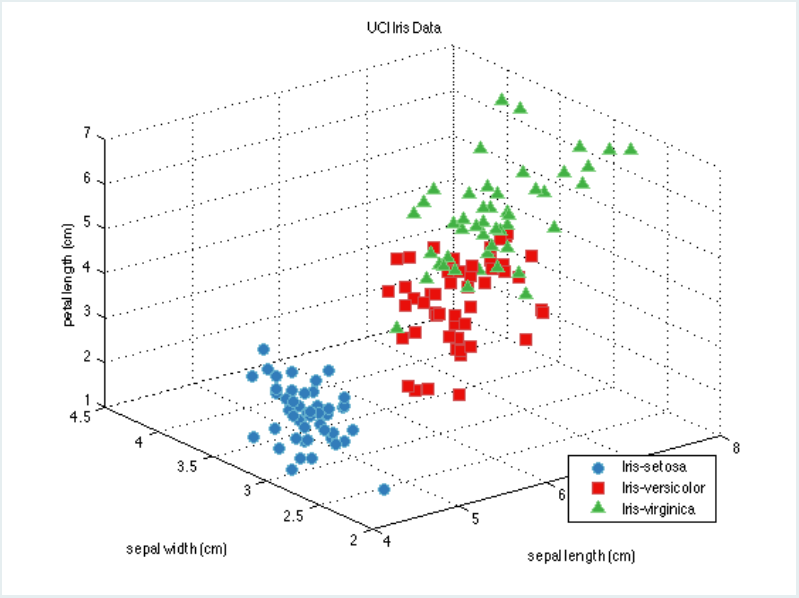
Feature space and decision boundary

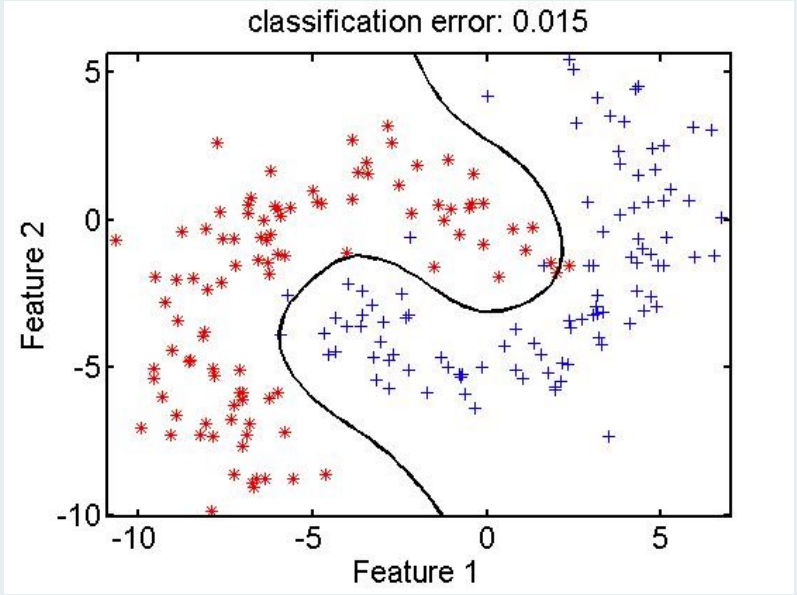




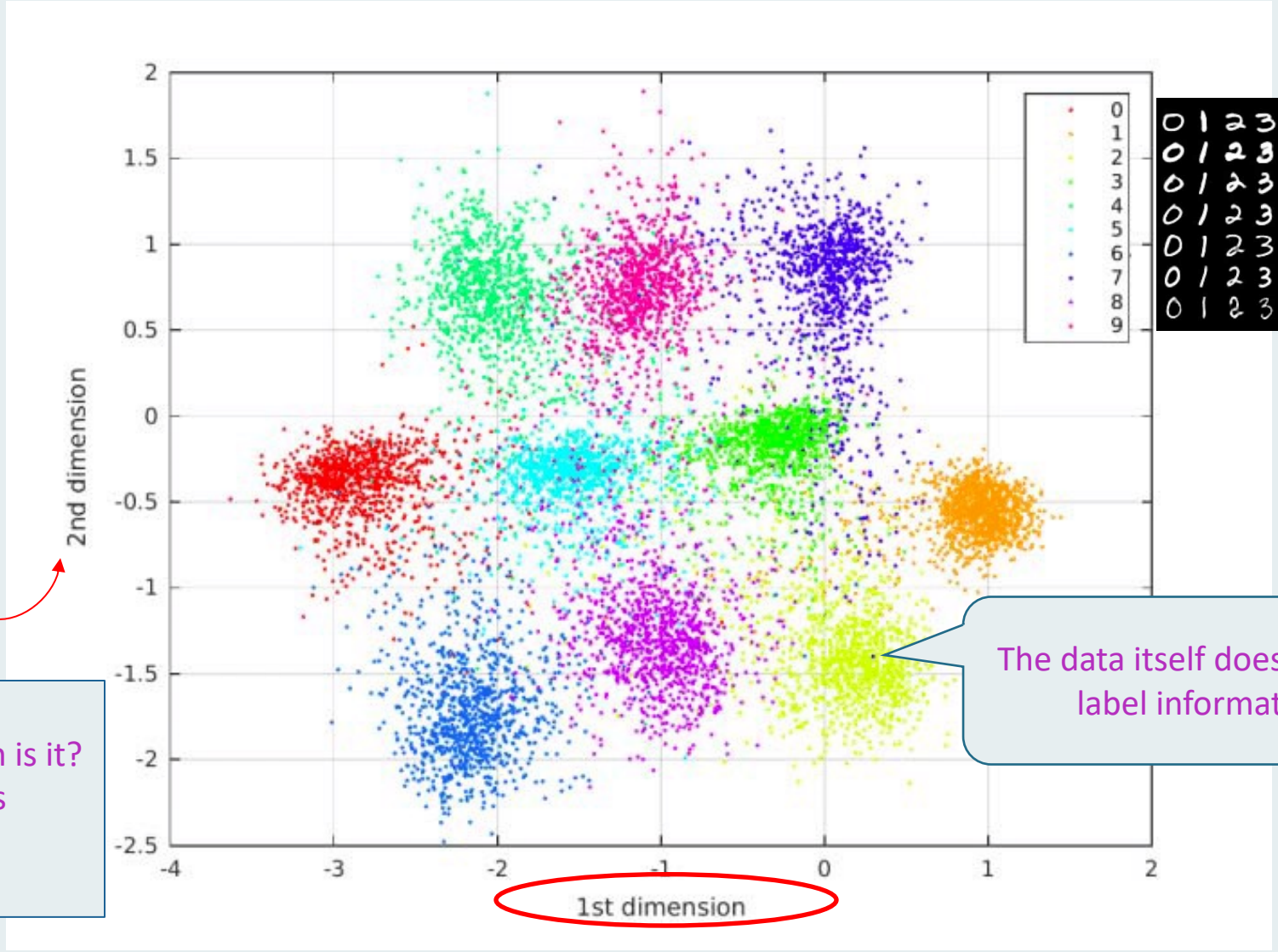
Supervised Learning (Classification) Space

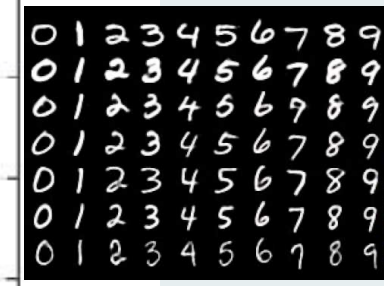






Unsupervised Learning (Clustering) Space



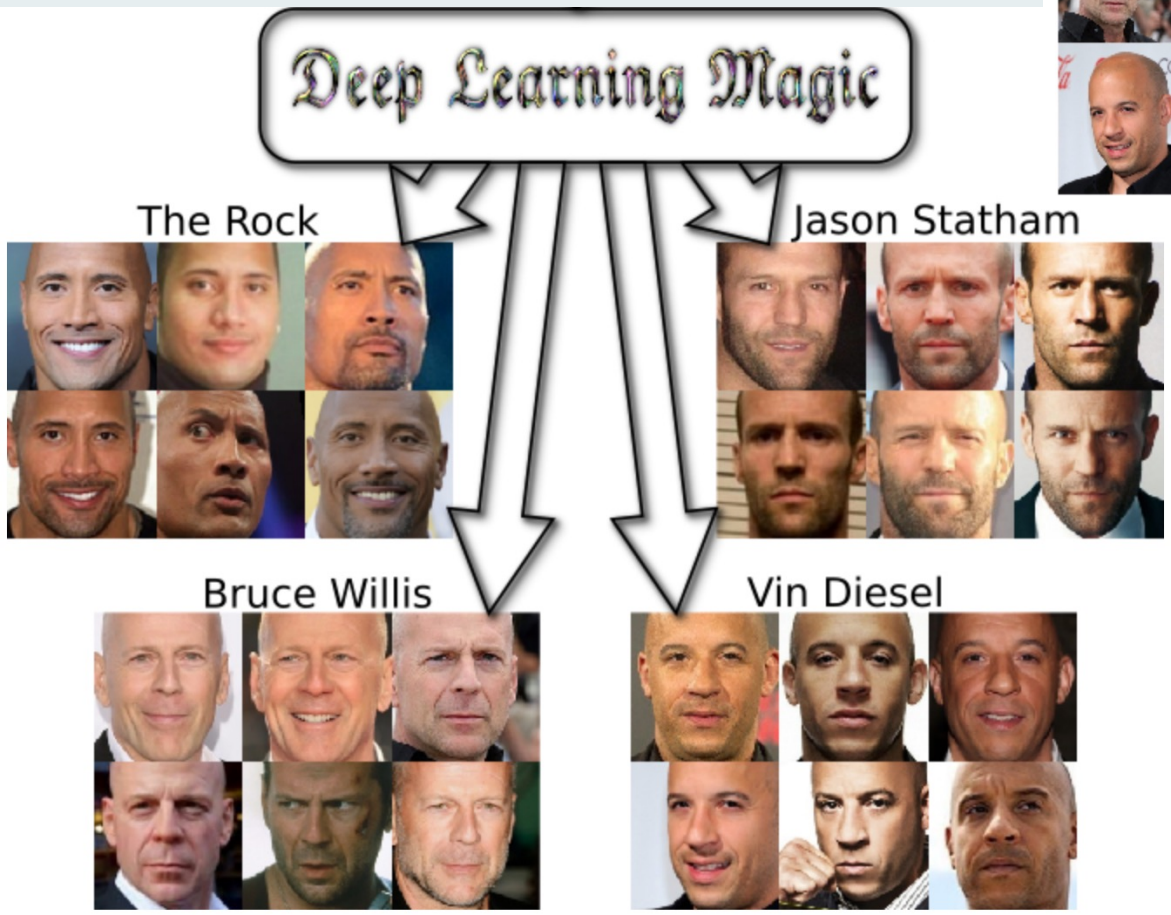


What dimension is it? What is its physical meaning?

The data itself does not have label information!

Example: Face clusters

Example:
Face Classification



After all, how are data samples represented?

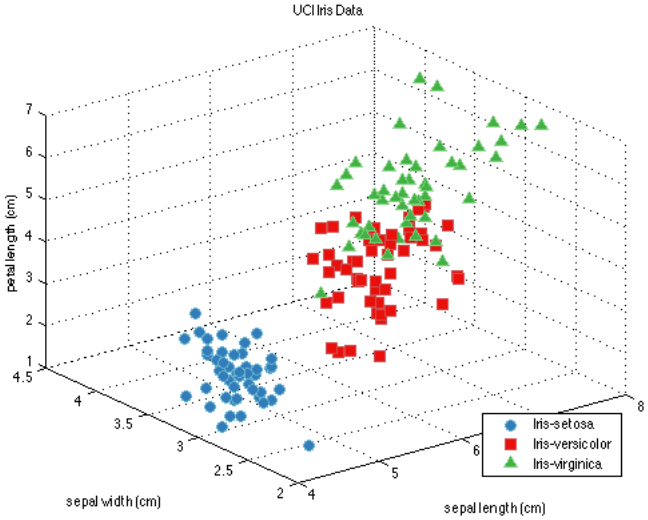
Iris dataset: sepal length and width, petal length

IRIS dataset

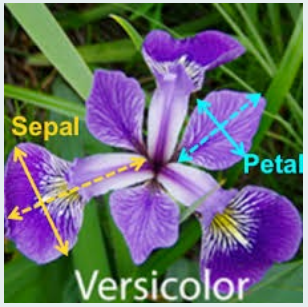
Iris Versicolor

Iris Setosa

Iris Virginica

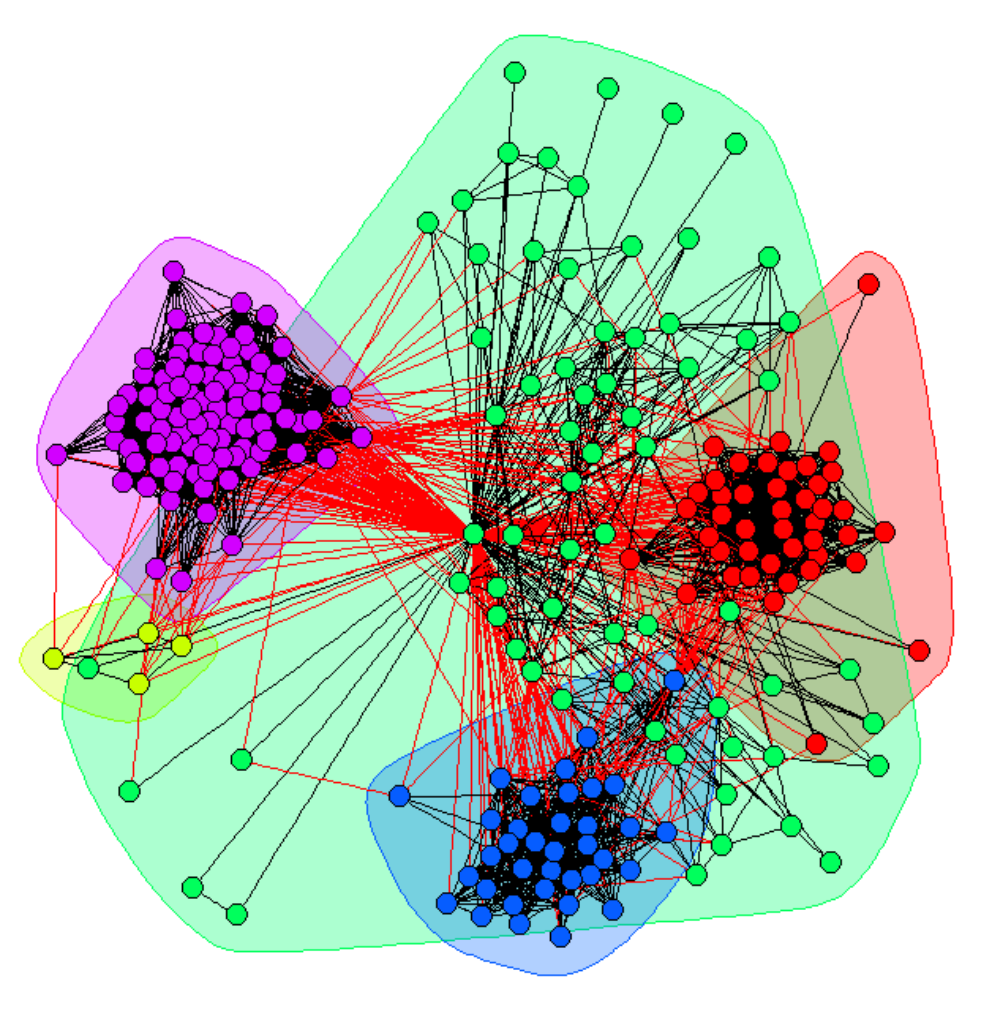


Feature engineering result ->



Face geometry?

Social Network Geometry?

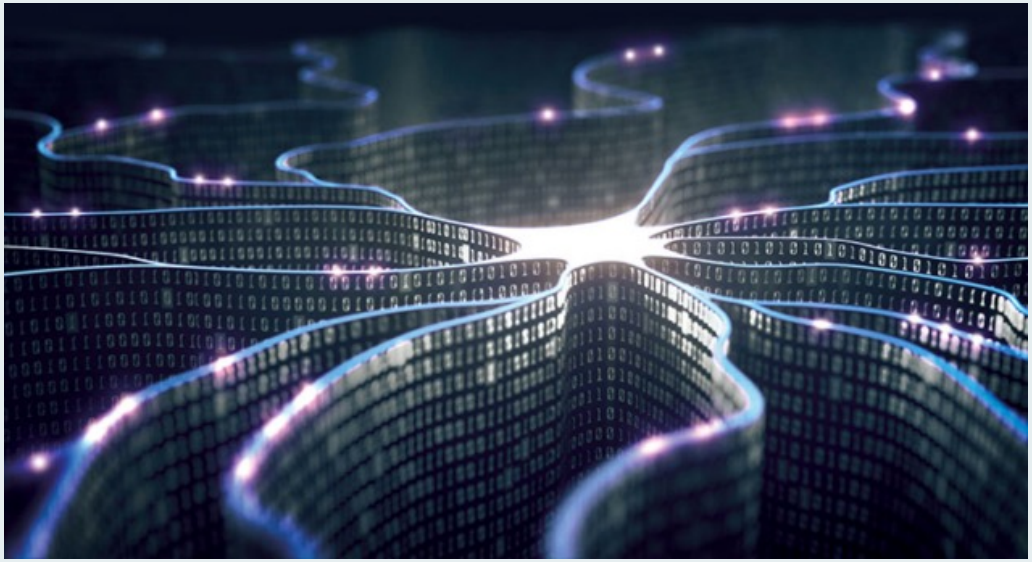


How are social network data represented?

How should stock data be represented?

It requires careful, domain knowledge driven feature engineering work, e.g. technical analysis like open-high-close-low (OHCL).

After all, can data representation be learned?



Yes, through Deep Learning
(Representation Learning)

Yet more advanced applications

So, what are the data, feature and output in this application?

Yet more advanced applications…

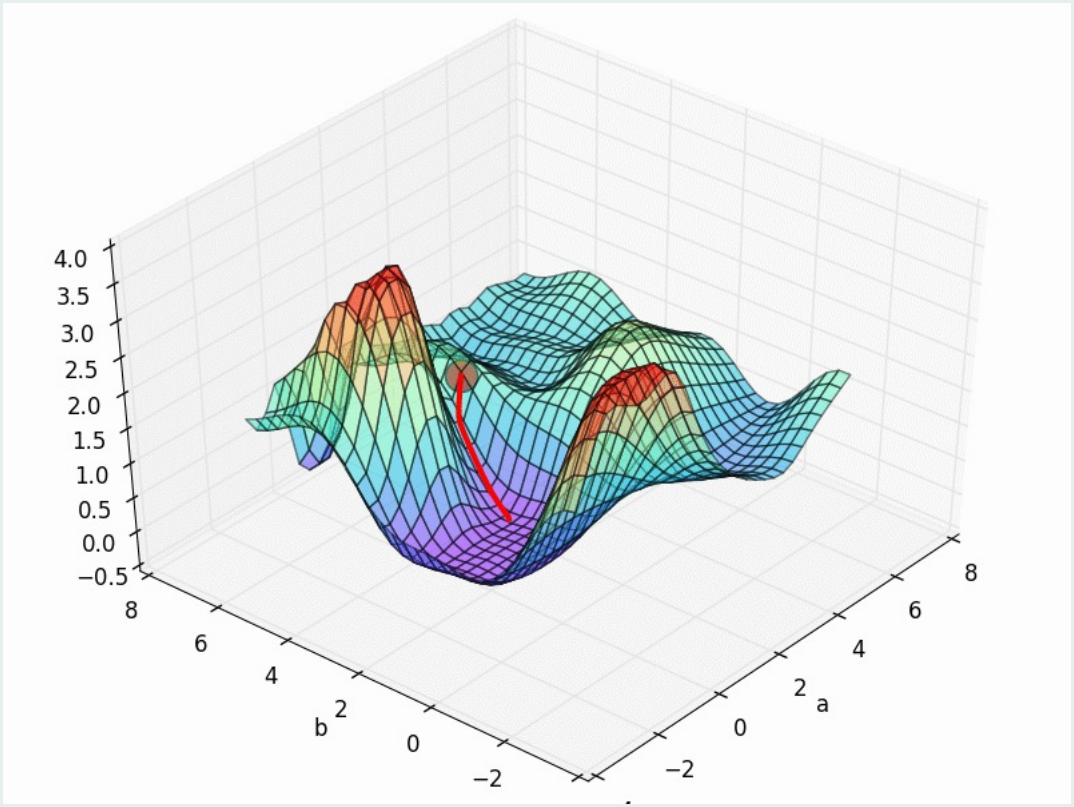
It’s just so powerful…

Image Completion

ML Concept - Theoretical/Mathematical Aspect

(to be learned) in this course

Stochastic Gradient Descent (SGD)



Cost Function

$$J(\Theta_0, \Theta_1) = \frac{1}{2m}\sum_{i=1}^{m}[h_\Theta(x_i) - y_i]^2$$

Predicted Value — True Value

Gradient Descent

$$\Theta_j = \Theta_j - \alpha\frac{\partial}{\partial\Theta_j}J(\Theta_0, \Theta_1)$$

Learning Rate

Now,

$$\frac{\partial}{\partial\Theta}J_\Theta = \frac{\partial}{\partial\Theta}\frac{1}{2m}\sum_{i=1}^{m}[h_\Theta(x_i) - y]^2$$

$$= \frac{1}{m}\sum_{i=1}^{m}(h_\Theta(x_i) - y)\frac{\partial}{\partial\Theta_j}(\Theta x_i - y)$$

$$= \frac{1}{m}(h_\Theta(x_i) - y)x_i$$

Therefore,

$$\Theta_j := \Theta_j - \frac{\alpha}{m}\sum_{i=1}^{m}[(h_\Theta(x_i) - y)x_i]$$

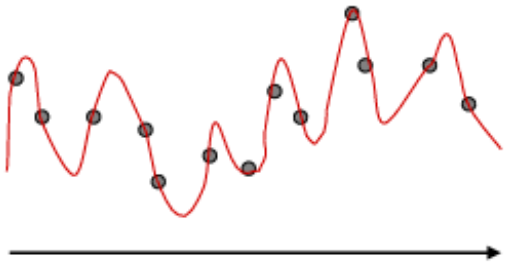
Regularization

– The minimization

$$\min_f |Y_i - f(X_i)|^2$$

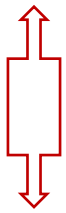
may be attained with zero errors.

But the function may not be unique.



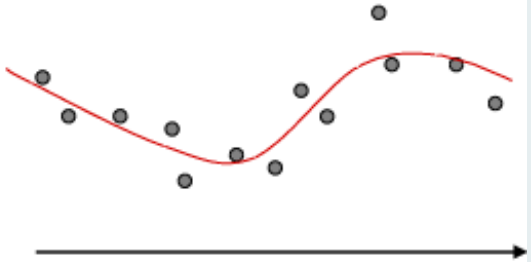
– Regularization

$$\min_{f \in H} \sum_{i=1}^{n} |Y_i - f(X_i)|^2 + \lambda \|f\|_H^2$$



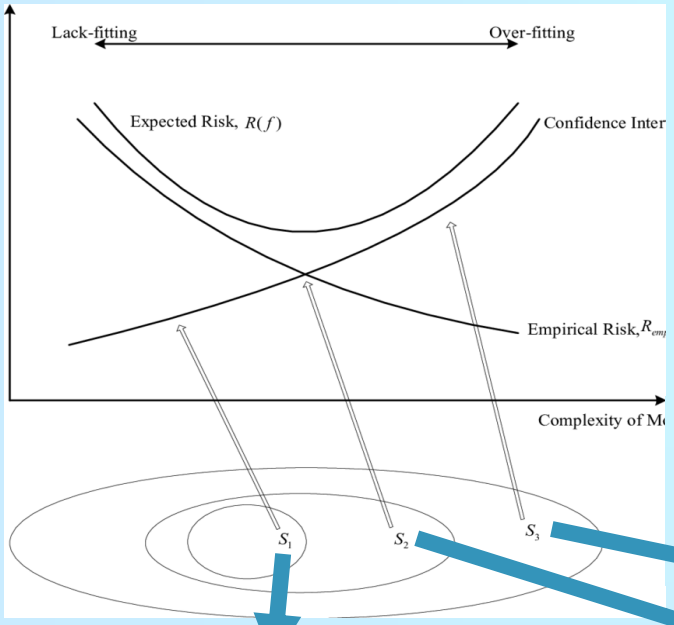
Dataset is the same but fitting is "improved", i.e., generalize better to unseen data!

- Regularization with smoothness penalty is preferred for uniqueness and smoothness.



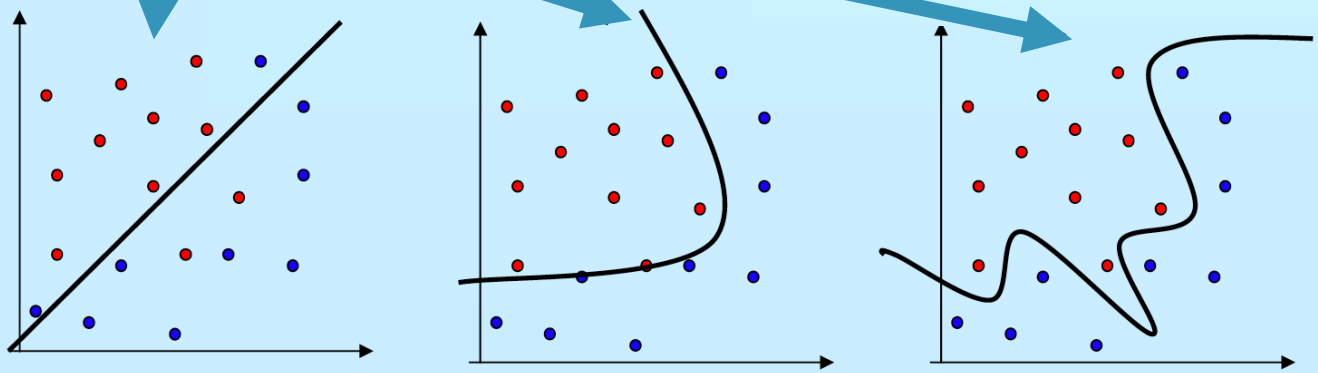
II-26

Bias-Variance Trade-off



- Define a set of learning functions, {S}
- Order it in terms of complexity, $\{S_1, \ldots, S_N\}$
- Select the optimal S^*

$$F = \{f(x, \alpha),\ \alpha \in \Lambda\}$$



Why?

What?

Where?

Why “Learn”?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to “learn” to calculate payroll (which is deterministic, according to a formulae)
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition, painting style)
  - Solution changes in time (routing on a computer network, stock market)
  - Solution needs to be adapted to particular cases (user biometrics, Chinese chatbot)

What We Talk About When We Talk About "Learning"

- Learning general models from data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Build a model that is *a good and useful approximation* to the data.

What is Machine Learning?

Definition Stuff:

"How do we create computer programs that improve with experience?"

Tom Mitchell

http://videolectures.net/mlas06_mitchell_itm/

"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. Machine Learning 1997.

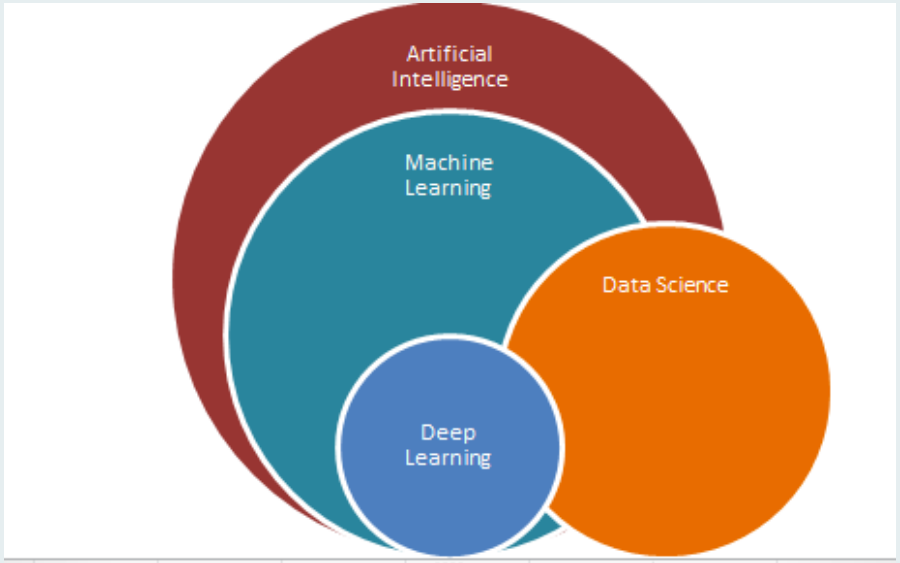
What is Machine Learning?

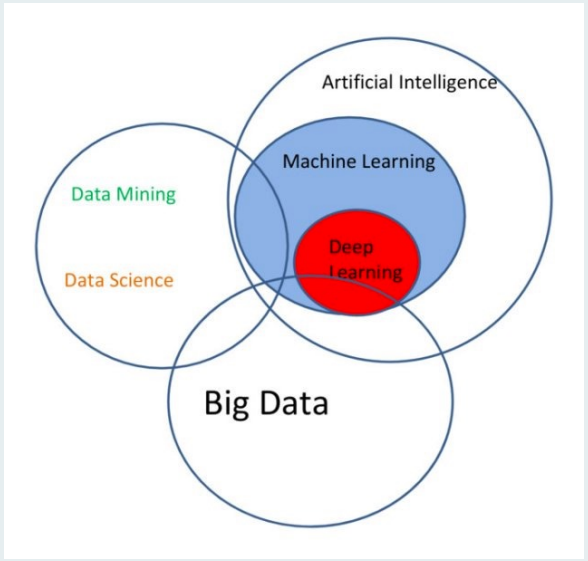
- Machine Learning
  - Study of algorithms that improve their performance at some task with experience
- Optimize a performance criterion using example data or past experience.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem, e.g. via SGD
  - Representing and evaluating the model for inference

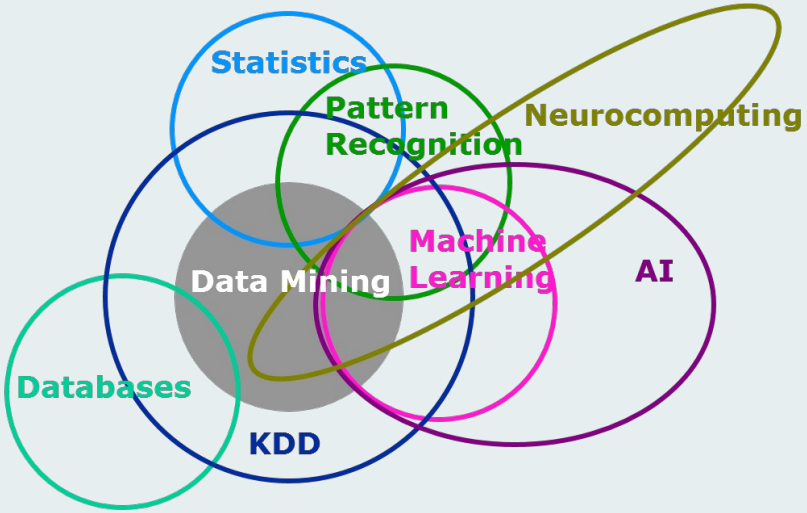
Why Study Machine Learning?
- Developing Better Computing Systems

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

Where is Machine Learning?

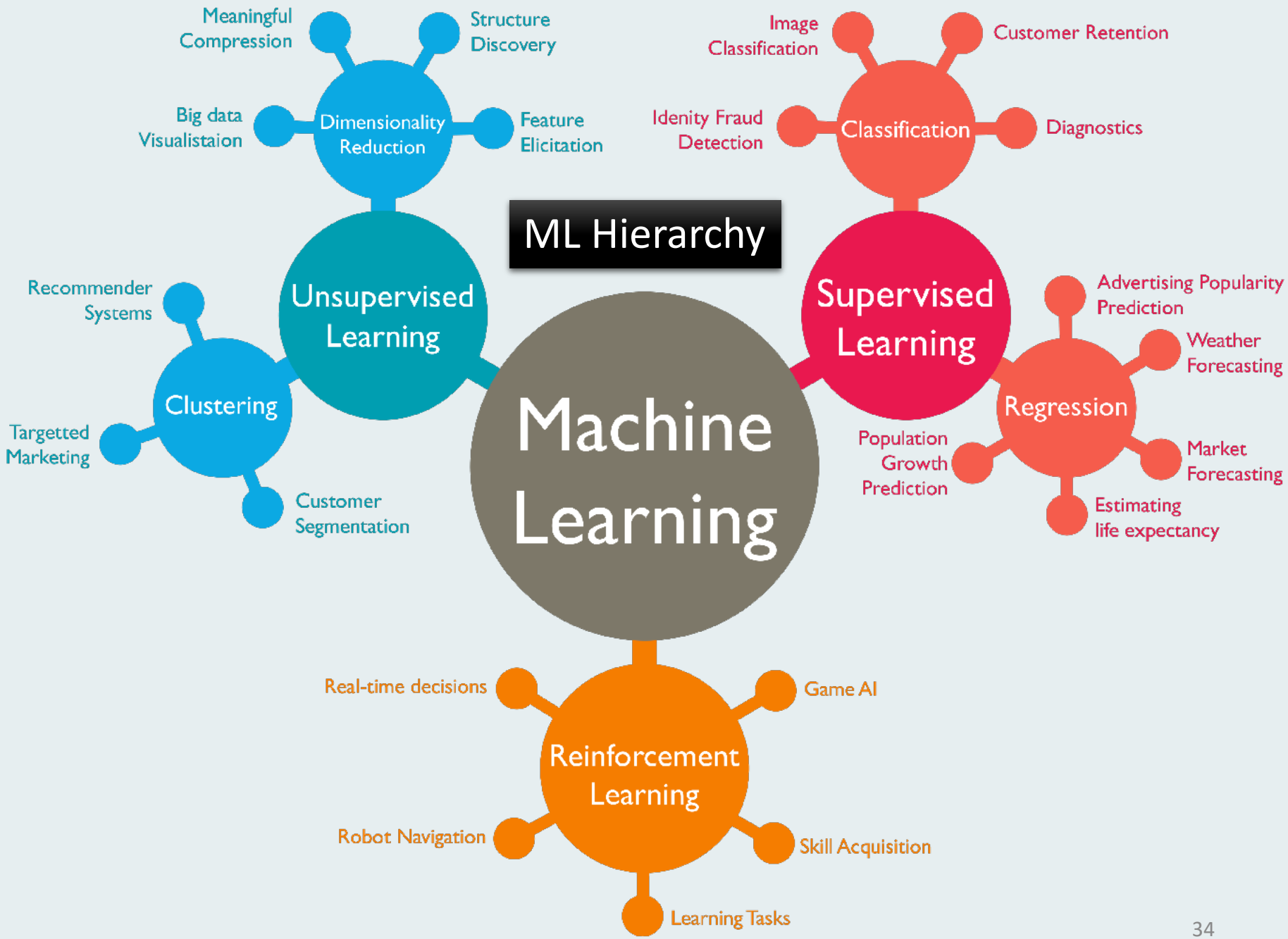






Related Disciplines

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


ML Hierarchy
Machine Learning
Unsupervised Learning
Supervised Learning
Reinforcement Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating life expectancy
Population Growth Prediction
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Growth of Machine Learning

- Machine learning is preferred approach to
  - Speech recognition, Natural language processing
  - Computer vision
  - Medical outcome analysis
  - Robot control
  - Computational biology
- This trend is accelerating
  - Improved machine learning algorithms
  - Improved data capture, networking, faster computers
  - Software too complex to write by hand
  - New sensors / IO devices
  - Internet of Things (IoT) and 5G Network
  - Demand for self-customization to user, environment
  - It turns out to be difficult to extract knowledge from human experts→*failure of expert systems in the 1980's.*

Models

- Association Analysis (in Data Mining)
- Supervised Learning
  - Classification
  - Regression/Prediction
- Unsupervised Learning
- Reinforcement Learning (in AI)
- Also, semi-supervised learning (in ML)

Learning Associations (DM stuff)

Famous quote: 60% of the customers buy diaper also buy beer

- Market Basket Analysis (MBA):

  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

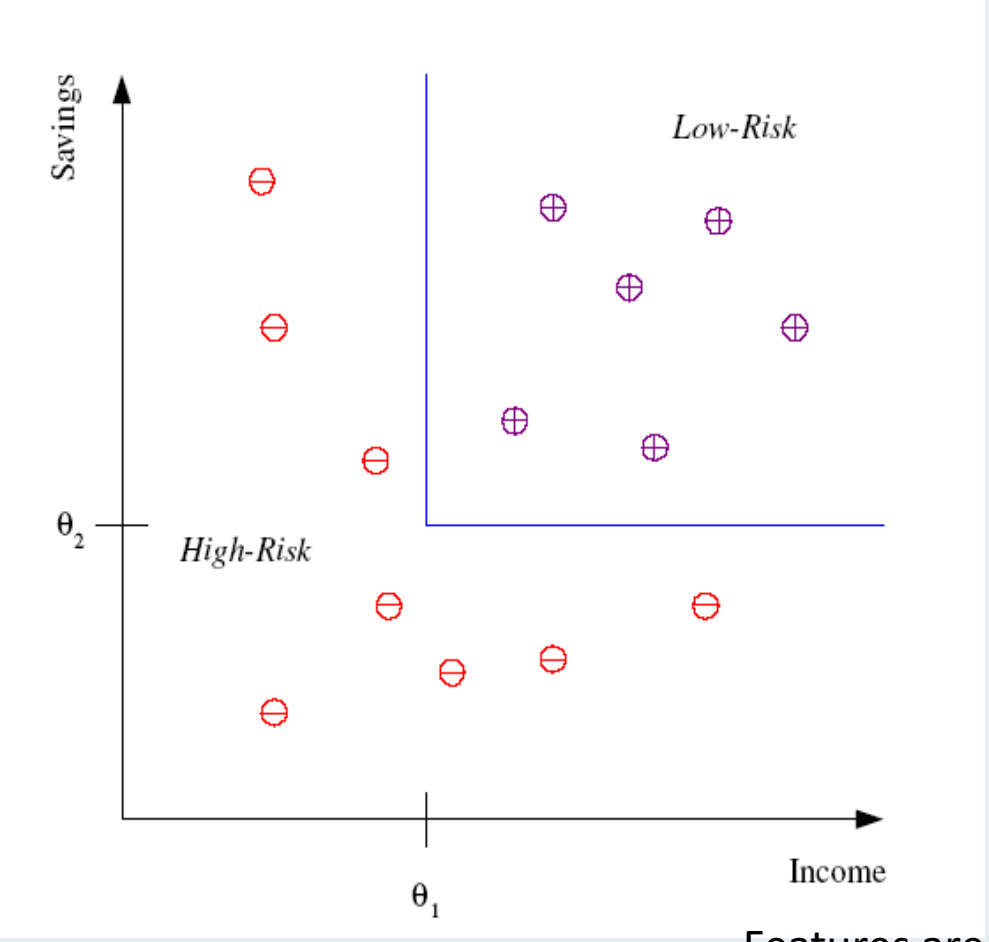
  Example: $P(\text{Diaper} \mid \text{Beer}) = 100\%$; $P(\text{Beer} \mid \text{Diaper}) = 75\%$

Market-Basket transactions

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

Classification

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*



ML/DM Model

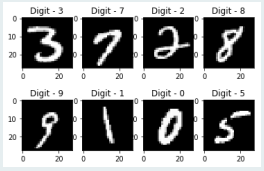
Discriminant:

IF *income* $> \theta_1$ AND *savings* $> \theta_2$
THEN low-risk ELSE high-risk

Features are "income and savings"! Why not "age"?

Classification: Applications

- Aka Pattern Recognition (PR)
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; e.g. visual (lip image) and acoustic for speech
- Medical diagnosis: From symptoms to illnesses
- Web Advertizing: Predict if a user clicks on an ad on the Internet.

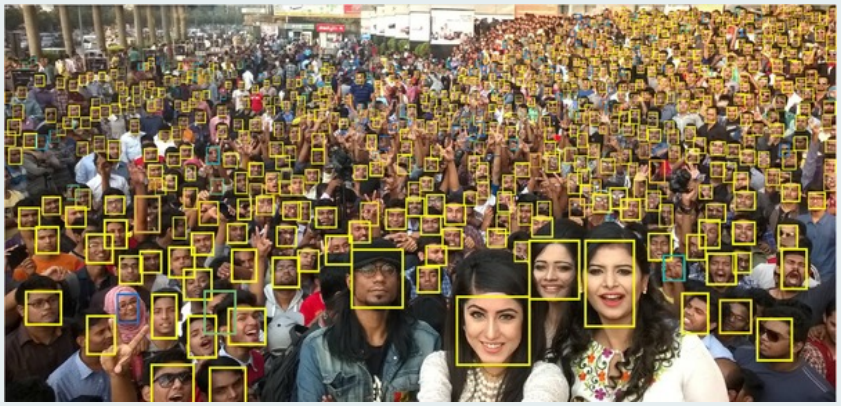


Face Recognition

Training examples of a person

Test images

AT&T Laboratories, Cambridge UK
http://www.uk.research.att.com/facedatabase.html



Prediction: Regression

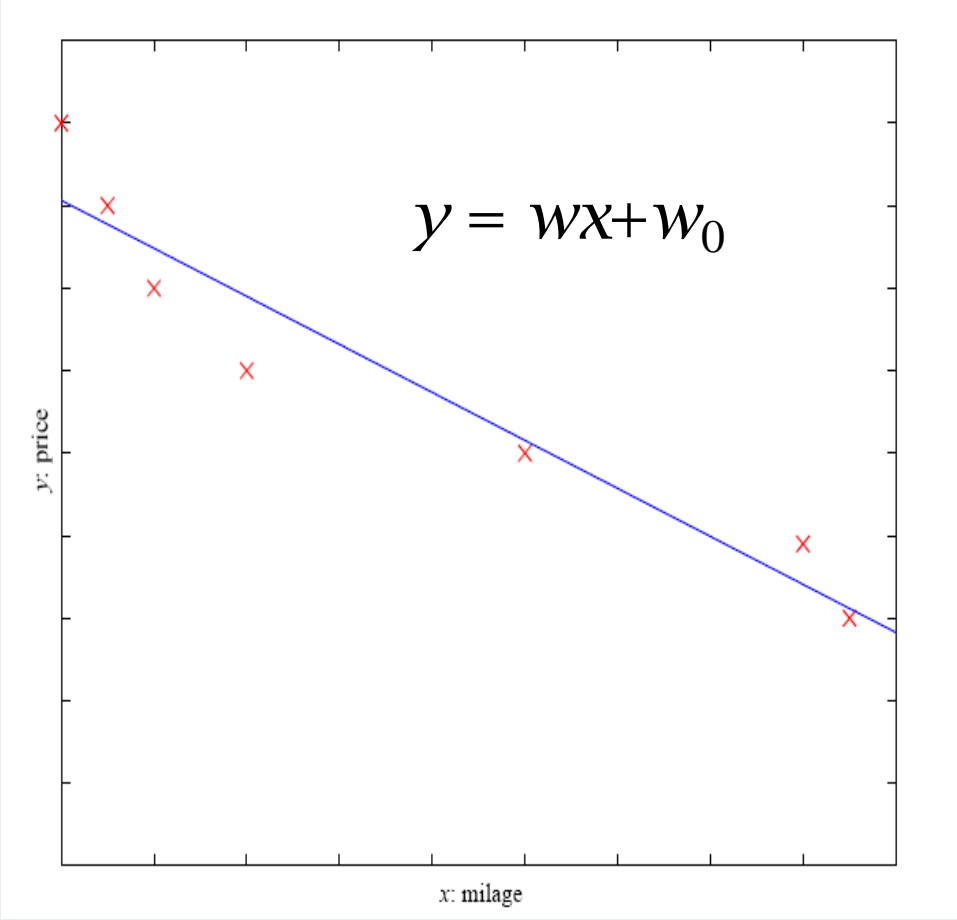
- Example: Price of a used car
- x : car attributes

  y : price

  $y = g(x \mid \theta)$

  $g(\ )$ model,

  θ learnable parameters (w, w_0)



Supervised Learning: Classification

- Example: Cancer diagnosis

Patient ID	# of Tumors	Avg Area	Avg Density	Diagnosis
1	5	20	118	Malignant
2	3	15	130	Benign
3	7	10	52	Benign
4	2	30	100	Malignant

Training Set

- Use this training set to learn how to classify patients where diagnosis is not known:

Patient ID	# of Tumors	Avg Area	Avg Density	Diagnosis
101	4	16	95	?
102	9	22	125	?
103	1	14	80	?

Test Set

Input Data Classification

- The input data is often easily obtained, whereas the classification or label is not.

Notationally,

Supervised Learning: Classification

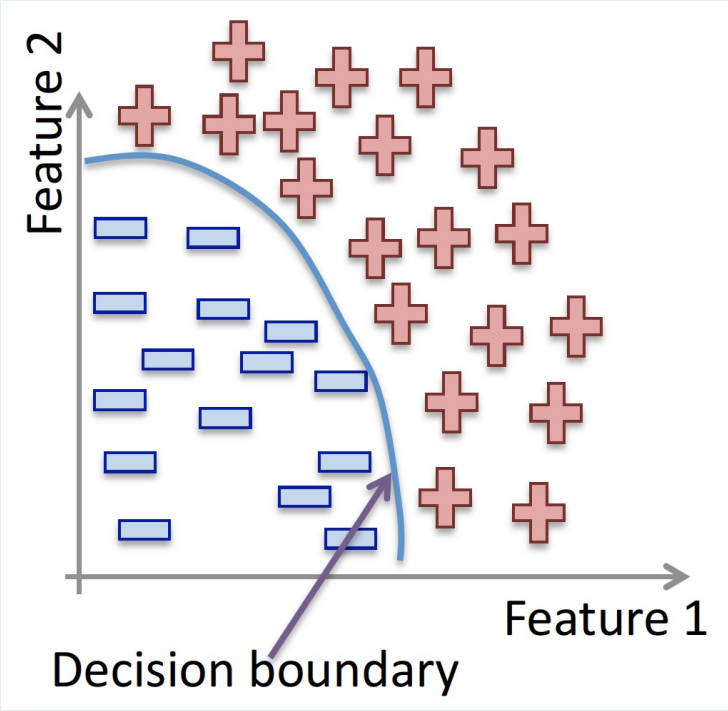
Given: Training data: $(x_1, y_1), \ldots, (x_n, y_n)$ / $x_i \in \mathbb{R}^d$ and y_i is the label.

example $x_1 \rightarrow$	x_{11}	x_{12}	$\ldots$	x_{1d}	$y_1 \leftarrow$ label
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
example $x_i \rightarrow$	x_{i1}	x_{i2}	$\ldots$	x_{id}	$y_i \leftarrow$ label
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
example $x_n \rightarrow$	x_{n1}	x_{n2}	$\ldots$	x_{nd}	$y_n \leftarrow$ label

Goal: Learn a model from labeled data; Use training set + some learning method to produce a predictive model; Use this predictive model to classify new data.

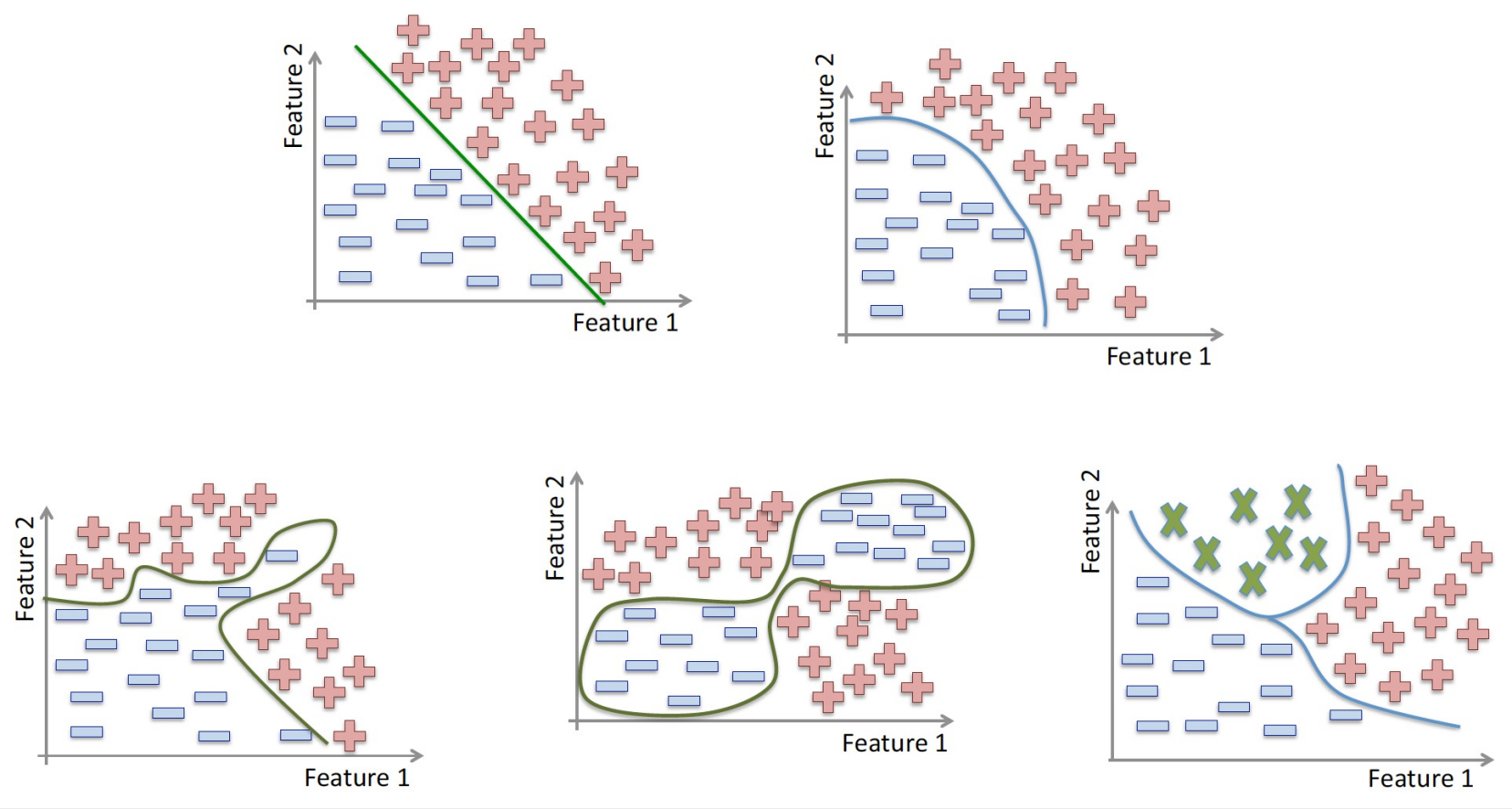
Pictorially,
Supervised Learning: Classification

What's happening in the feature space?



- Methods: Support Vector Machines (SVM), neural networks (NN), decision trees, K-nearest neighbors, naive Bayes, convolutional NN (CNN), etc.

More sophisticated classification



Unsupervised Learning

- Learning “what normally happens”
- No output or label
- Clustering: Grouping similar instances
- Other applications: Summarization, Association Analysis
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

Notationally,

Unsupervised Learning: Clustering

Training data: "examples" x.

$$x_1, \ldots, x_n, \; x_i \in X \subset \mathbb{R}^n$$

- **Clustering/segmentation:**

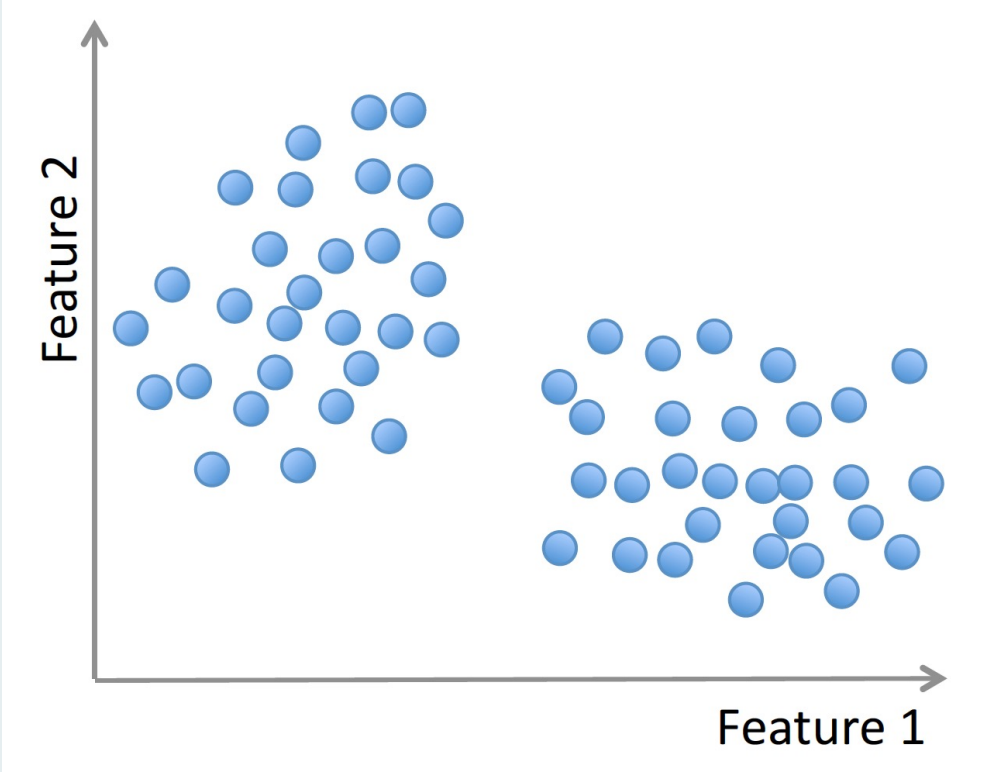
$$f : \mathbb{R}^d \longrightarrow \{C_1, \ldots C_k\} \text{ (set of clusters).}$$

Example: Find clusters in the population, fruits, species.

Pictorially,

Unsupervised Learning: Clustering

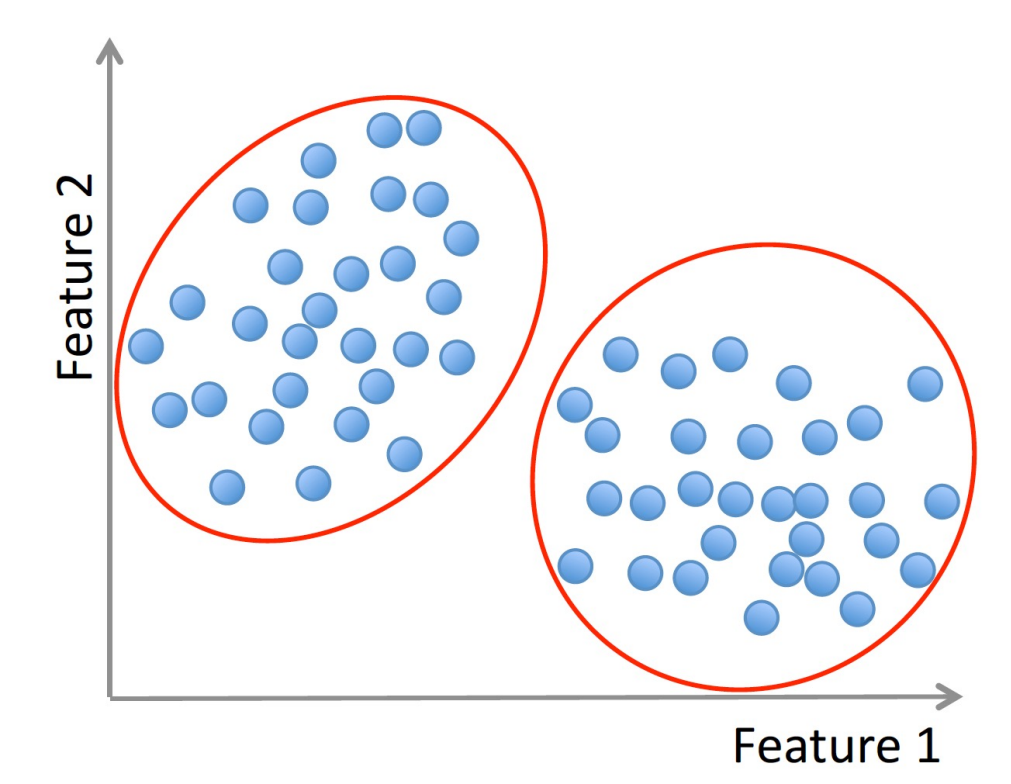
What's happening in the feature space?



Pictorially,

Unsupervised Learning: Clustering

What's happening in the feature space?



- Methods: K-means, gaussian mixtures, hierarchical clustering, spectral clustering, etc.

Reinforcement Learning

- Topics:
  - Policies: what actions should an agent take in a particular situation
  - Utility estimation: how good is a state (→used by policy)
- No supervised output but delayed reward
- Credit assignment problem (what was responsible for the outcome?)
- Applications:
  - Game playing
  - Robot in a maze
  - AWS DeepRacer
  - Multiple agents, partial observability, ...

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
- Integrating learning with systems
- What are the theoretical limits of learnability?
- Continuous (life-long) learning
- Transfer learning
- Few-shot learning
- Interpretable ML (or explainable AI)
- Many others…

Measuring Performance

- Generalization accuracy
- Solution correctness
- Solution quality (length, efficiency)
- Speed of performance (scalability)

Scaling issues in ML

- Number of
  - Inputs
  - Outputs (e.g. Extreme Classification (with lots of labels))
  - Batch vs real-time
  - Training vs testing

Resources: Datasets

- UCI ML Repository: https://archive.ics.uci.edu/
- Kaggle: https://www.kaggle.com/datasets
- Tianchi (天池): https://tianchi.aliyun.com/dataset/

and many others…

Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- …

Resources: Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Computational Learning
- International Joint Conference on Artificial Intelligence (IJCAI)
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)
- IEEE Int. Conf. on Data Mining (ICDM)
- Etc.

Take-home Messages

About ML and Data Analytics (DA)

- DA concerns more about data+feature
- ML concerns more about model

About this subject

- Introductory course that covers a range of fundamental machine learning techniques and some popular deep learning models.
- It requires lots of abstract thinking.
- A bit more theoretical/mathematical oriented than Data Mining or Big Data Analytics like subjects.
- ML development platform to be self-learnt and practiced in competition/project.
- **It's going to be fun and hard work.**

Acknowledgement

- Slides of
  - E. Alpaydin, Introduction to Machine Learning. 2nd Ed. MIT Press, 2010.
  - C.F. Eick, U of Houston.
- Photos from Internet