Machine Learning Review

Resana Machine Learning Workshop

prepared by Hosein Hasani

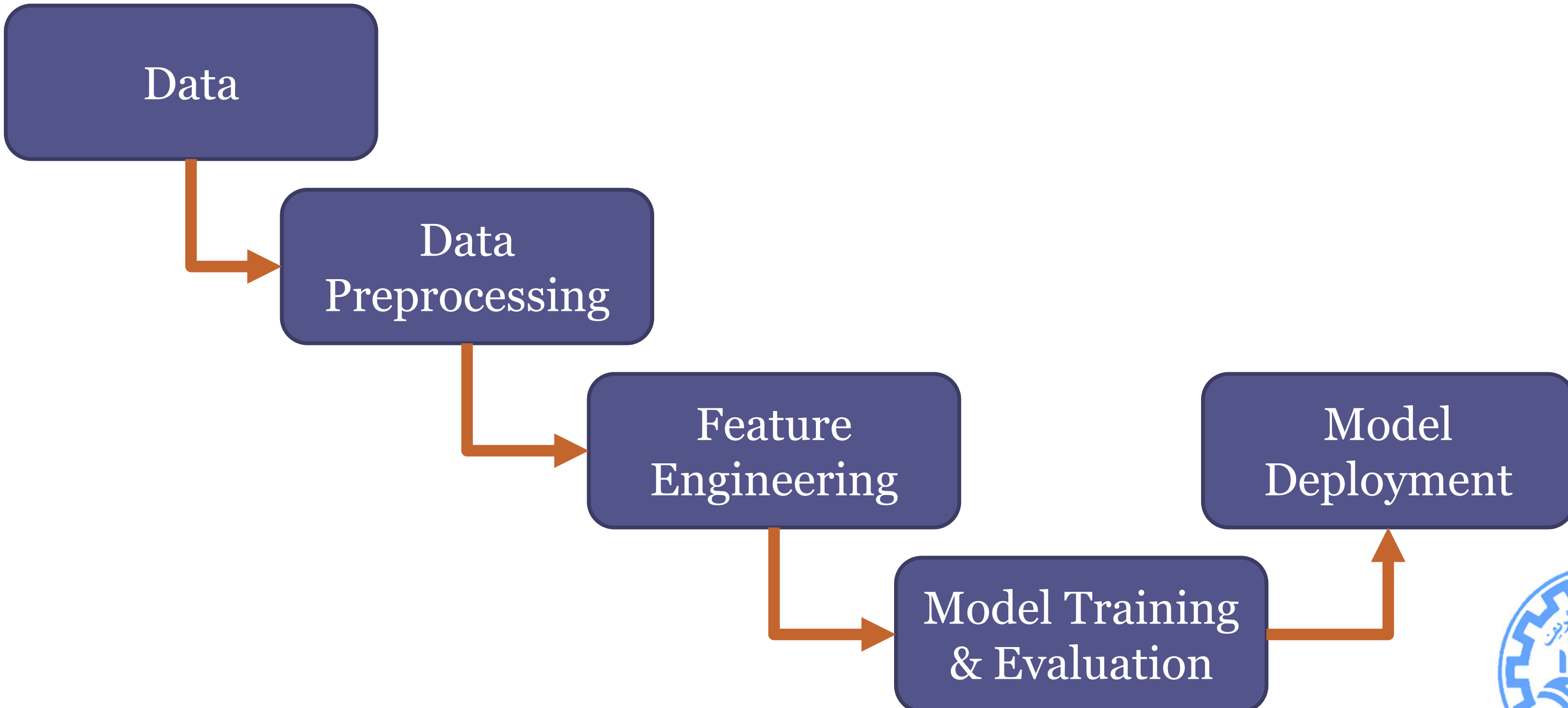
Summer 1399

Outline

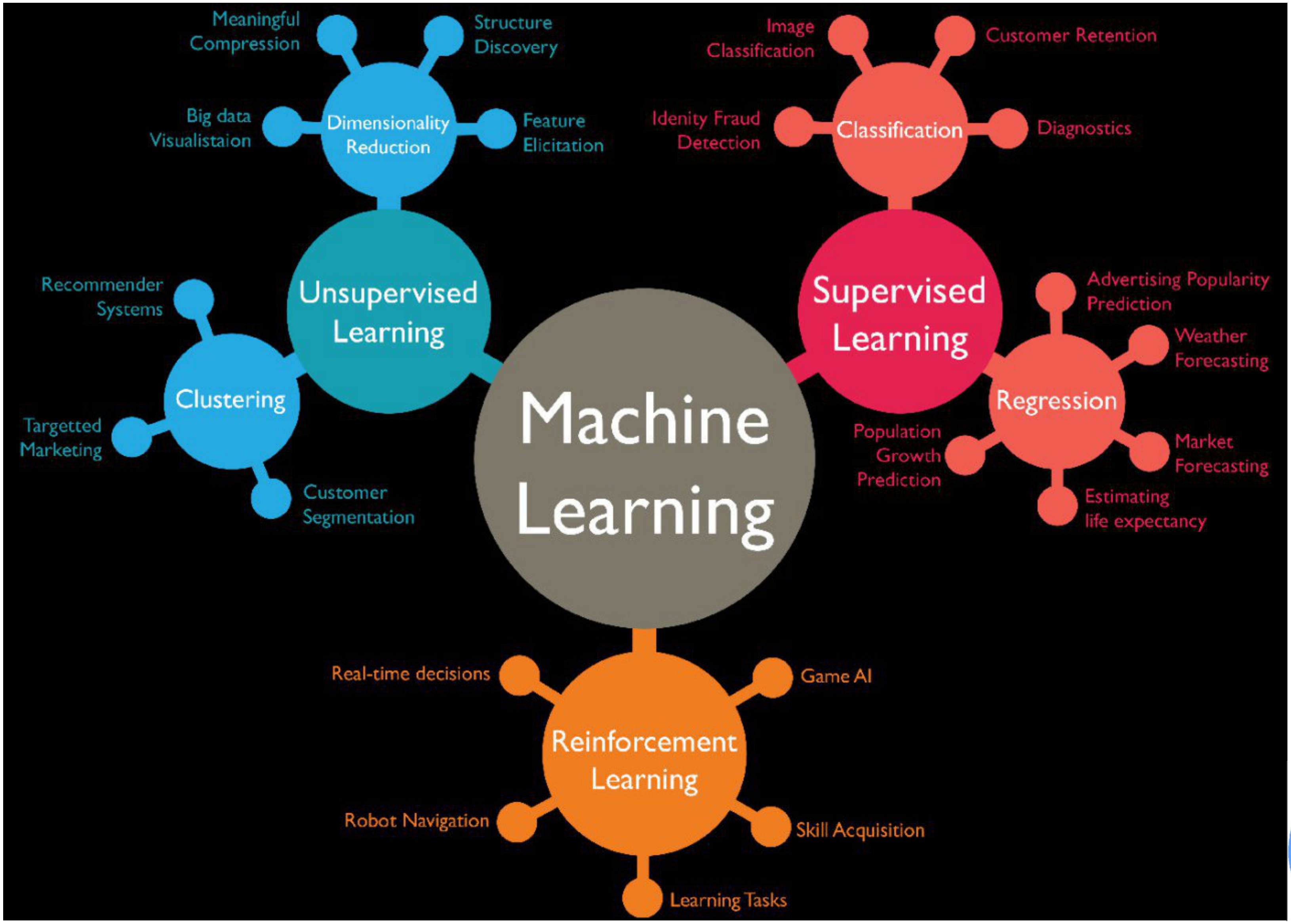
- Common ML Problems:
  - Supervised Learning:
    - Classification
    - Regression
  - Unsupervised Learning :
    - Clustering
    - Dimensionality Reduction
- Generalization and Over-fitting Problem
- Model Selection and Validation

Machine Learning

- Using statistical models and algorithms to perform a specific task by learning data patterns, without being explicitly programmed.
- Typical steps:



Common ML Problems



Supervised Learning vs. Unsupervised Learning

- Supervised learning

  Given: Training set

  Labeled set of N input-output pairs $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$

  Goal: Learning a mapping from x to y

- Unsupervised learning

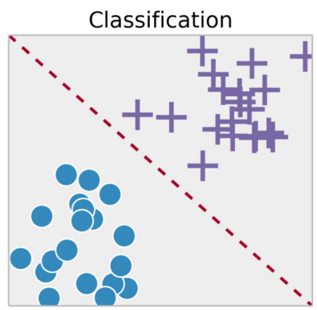
  Given: Training set

  $$D = \{(x^{(i)})\}_{i=1}^{N}$$

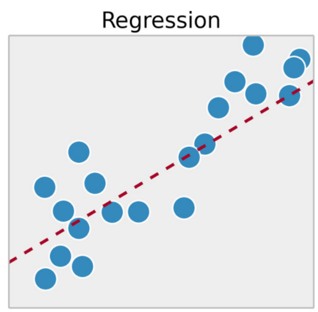
  Goal: Revealing structure in the observed data and finding groups or intrinsic structures in the data

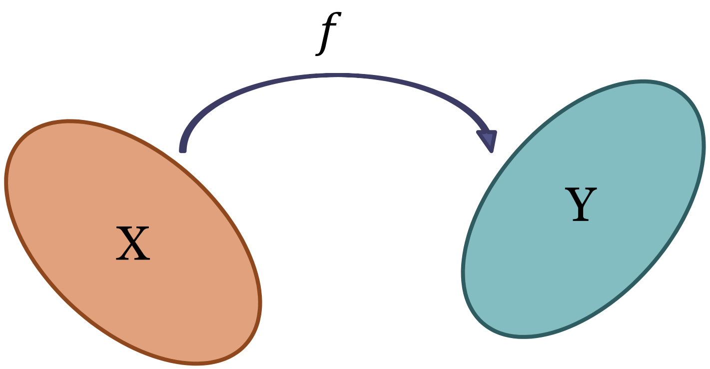
Supervised Learning: Regression vs. Classification

- Classification: predict a discrete target variable e.g. $y \in \{1, 2, \ldots, C\}$



- Regression: predict a continuous target variable e.g. $y \in [0, 1]$





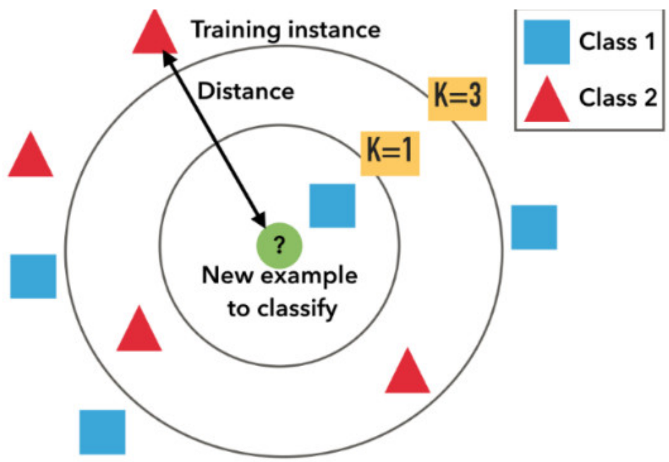
Classification

- A function $f : R^n \rightarrow \{1, \ldots, k\}$ specifies which of k categories an input vector x belongs to.

Classification

- A function $f : R^n \rightarrow \{1, \ldots, k\}$ specifies which of k categories an input vector x belongs to.

- Case Study: KNN (K Nearest Neighbors)
  - Stores all training cases and classify new cases based on similarity measure (like Euclidean distance)



Regression

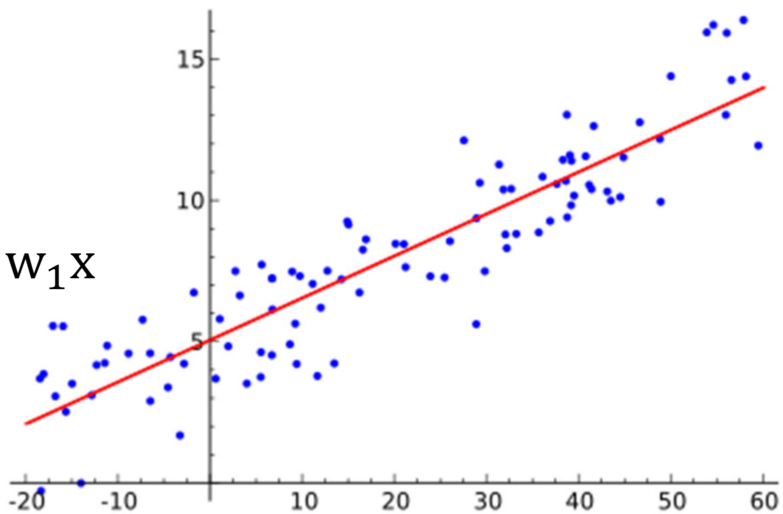
- A function $f : R^n \rightarrow R$ that maps an input vector x to a continues value y.

Regression

- A function $f : R^n \rightarrow R$ that maps an input vector x to a continues value y.

- Case study: Linear Regression

$$f(x; w) = \mathrm{w}_0 + \mathrm{w}_1 \mathrm{x}$$



$w = [w_0, w_1]$: Parameters that be estimated during optimization

Regression

- Other example: Face Landmark Detection

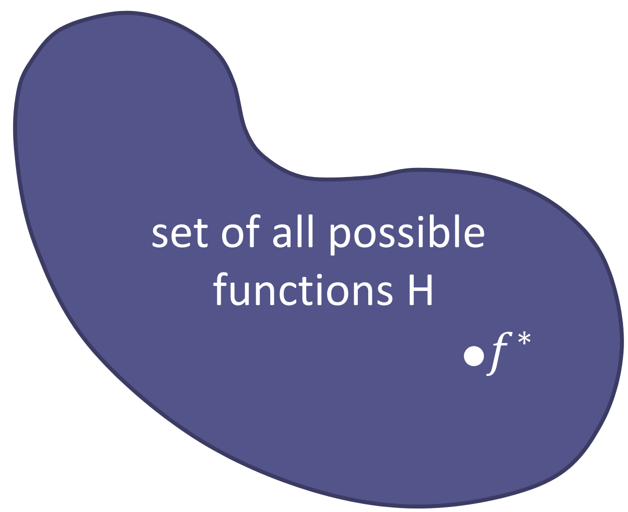
Hypothesis Class and Inductive Bias

- The aim of supervised learning is to find f^* (best solution) from **hypothesis space** (e. g. the set of all possible functions)

- Example:

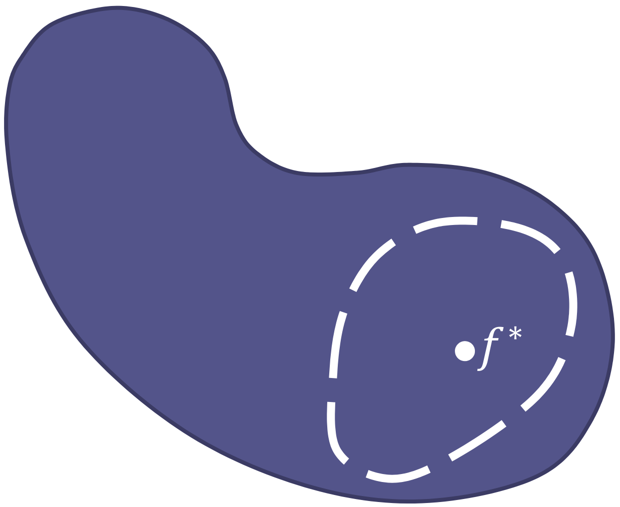
In linear regression the hypothesis space include all possible functions with the form of

$$f(x; w_0, w_1) = w_0 + w_1 x$$



Hypothesis Class and Inductive Bias

- **Inductive bias** is the set of assumptions that a learner uses to predict outputs of given inputs.
- Some times we use our knowledge about the nature of data to **restrict** the hypothesis space.



Unsupervised Learning

Unsupervised learning

Given: Training set

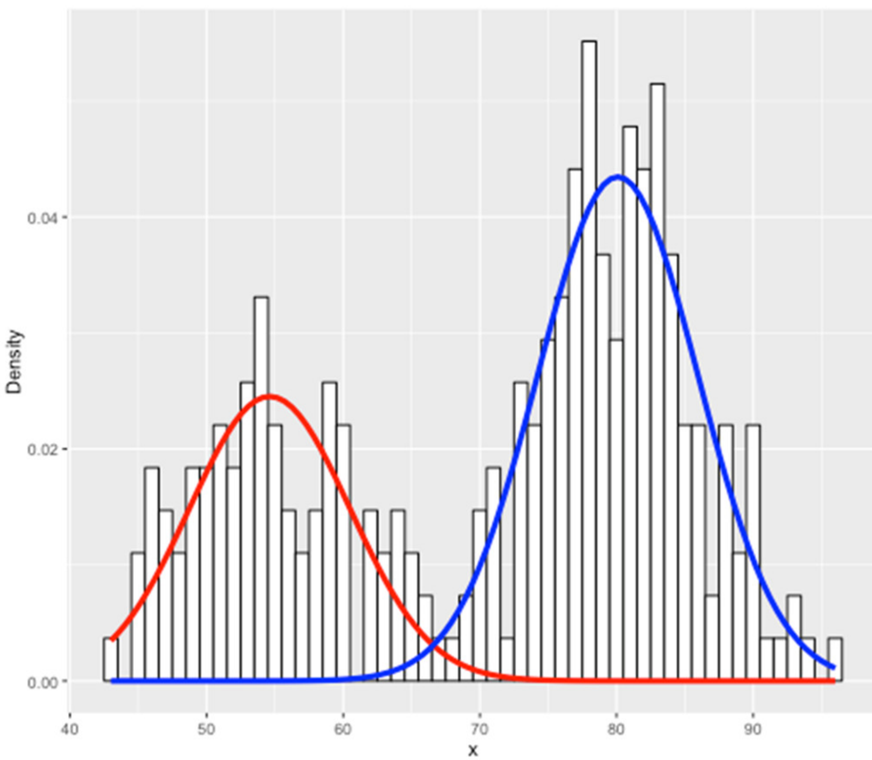
$$D = \{(x^{(i)})\}_{i=1}^{N}$$

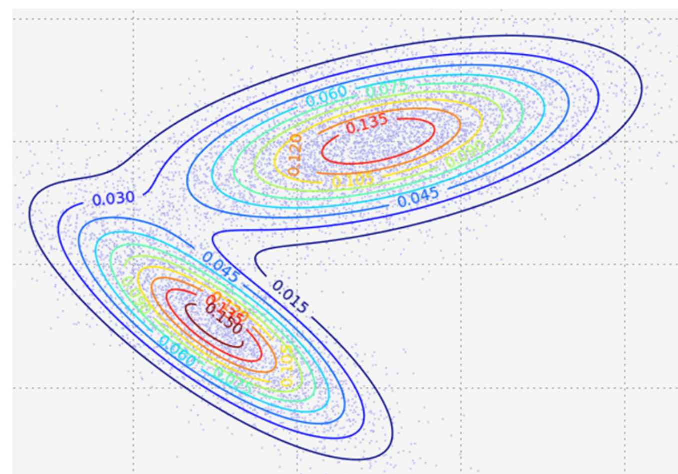
Goal: Revealing structure in the observed data and finding groups or intrinsic structures in the data

- Related Algorithms:
  - Clustering
  - Dimensionality Reduction
  - Generative Models

Clustering

- A technique to assign each point into a specific group.





Gaussian Mixture Model

Clustering: Case Study

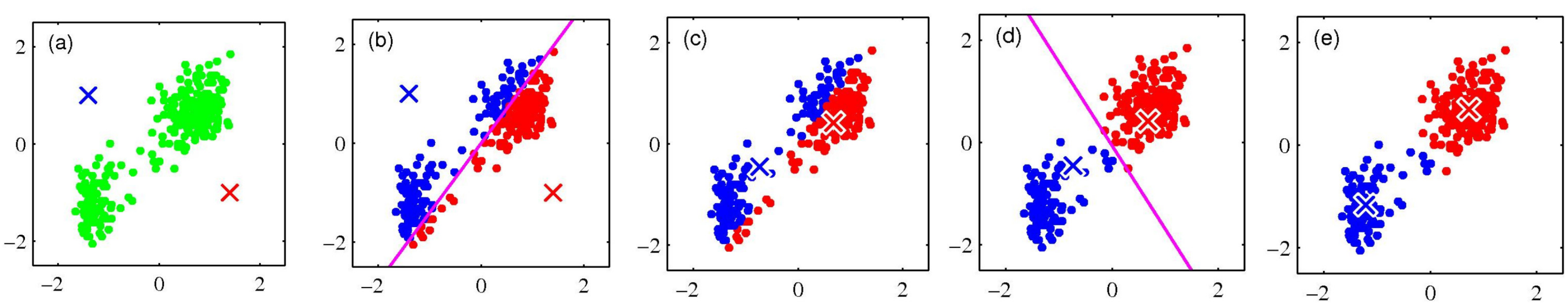
- K-means:

1. Choose number of clusters K.
2. Pick K random points as cluster centers (centroid)
3. Compute the distance between data points and all centroids.
4. Assign each data point to the closest centroid
5. Compute the centroids for the clusters (average of the data points that belong to each cluster)
6. Iterate steps 3-5 until convergence (no change to the centroids).

Clustering: Case Study

- K-means:

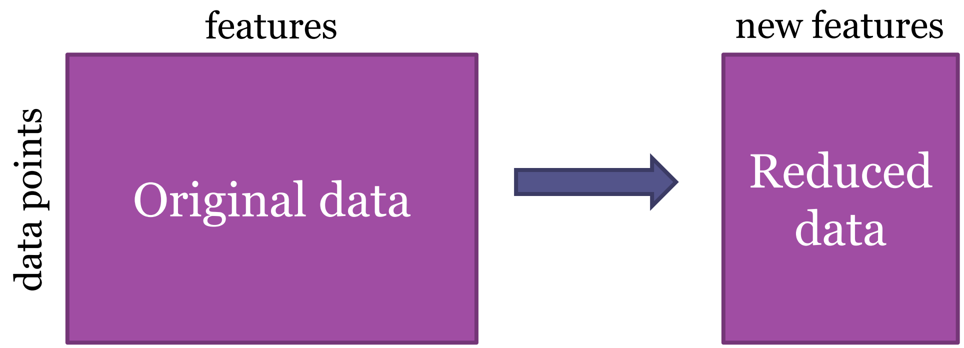
1. Choose number of clusters K.
2. Pick K random points as cluster centers (centroid)
3. Compute the distance between data points and all centroids.
4. Assign each data point to the closest centroid
5. Compute the centroids for the clusters (average of the data points that belong to each cluster)
6. Iterate steps 3-5 until convergence (no change to the centroids).



(2) (3-4) (5) (3-4) (6)

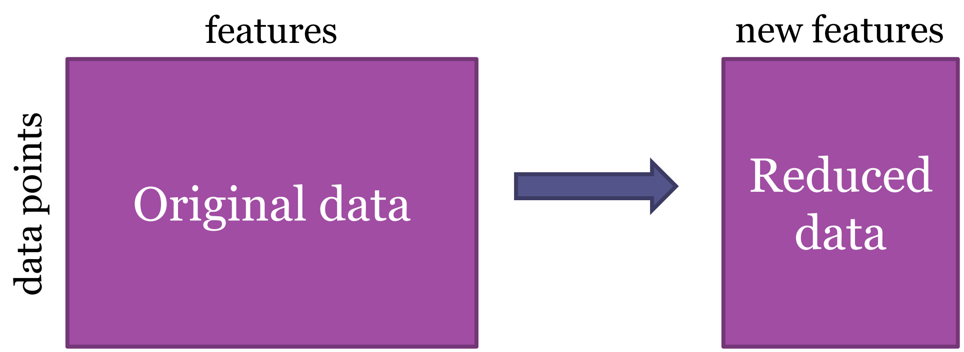
Dimensionality Reduction

- A technique to find a lower-dimensional representation of data features that preserves some of its properties.
- Motivations: Computation, Visualization, Feature extraction.

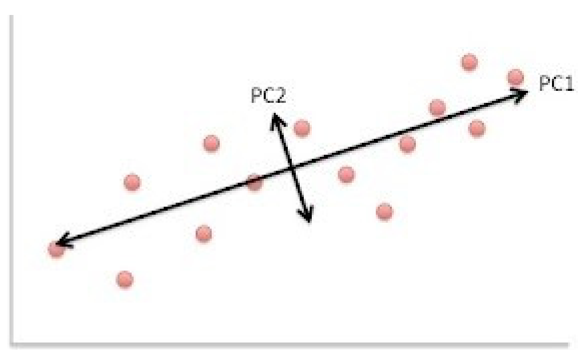


Dimensionality Reduction

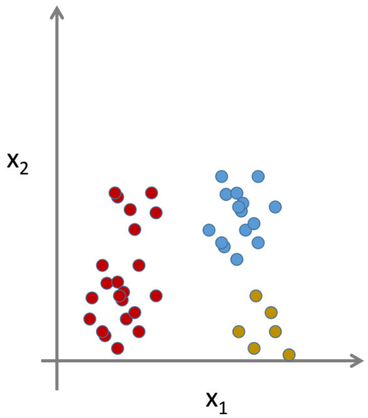
- A technique to find a lower-dimensional representation of data features that preserves some of its properties.
- Motivations: Computation, Visualization, Feature extraction.

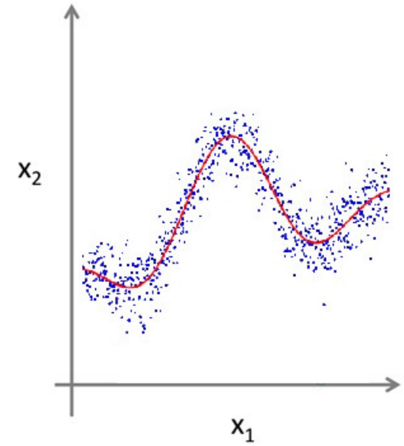


- Case study: PCA (Principal Component Analysis)

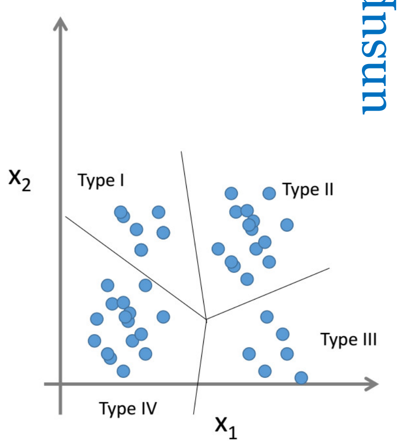


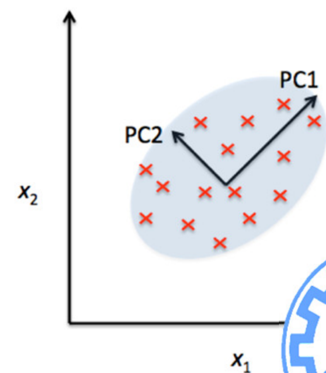
Comparison





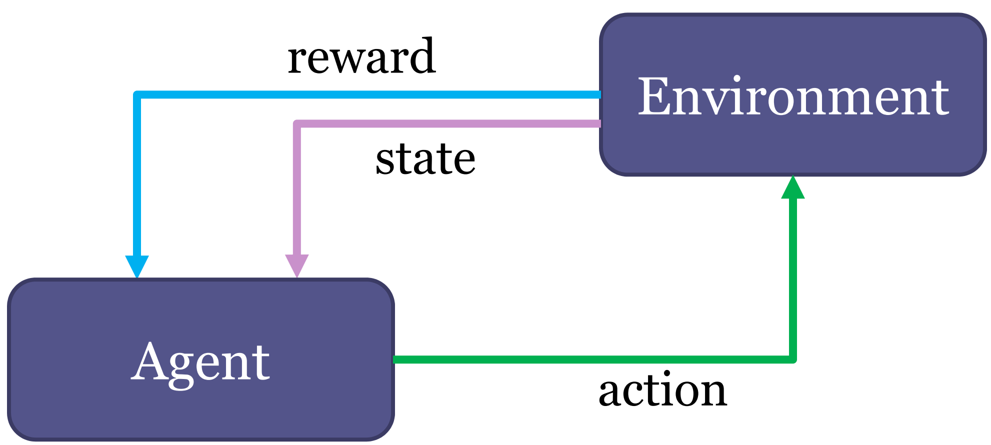
	discrete	Continues
supervised	**classification**	**regression**
unsupervised	**clustering**	**dimensionality reduction**





Reinforcement Learning

- An agent learns appropriate actions (policy) based on environment feedbacks to maximize reward.

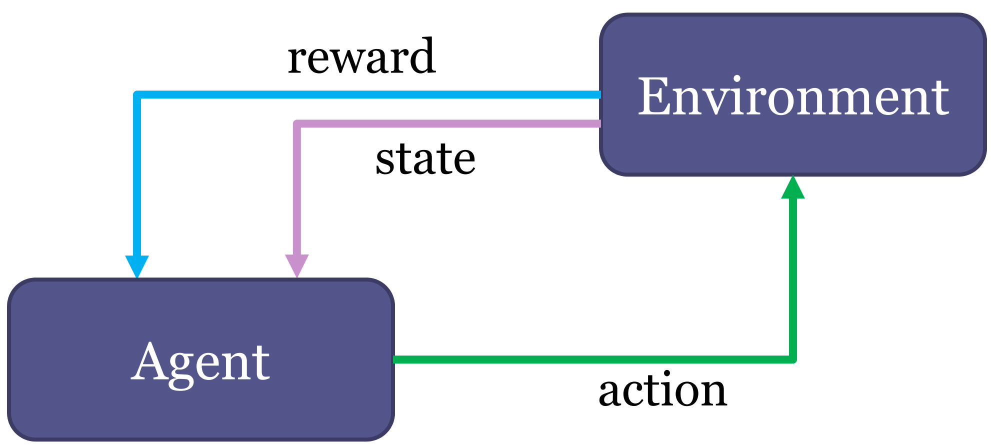


Reinforcement Learning

- An agent learns appropriate actions (policy) based on environment feedbacks to maximize reward.
- Provides only an indication as to whether an action is correct or not
  - Data in supervised learning:

    (input, label)
  - Data in reinforcement learning:

    (input, some output, a grade of reward for this output)



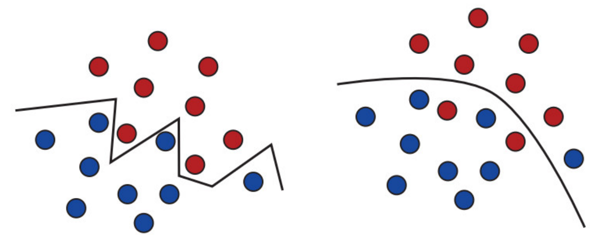
Generalization

- We don’t intend to memorize data but need to figure out the pattern.

- A core objective of learning is to generalize from the experience .

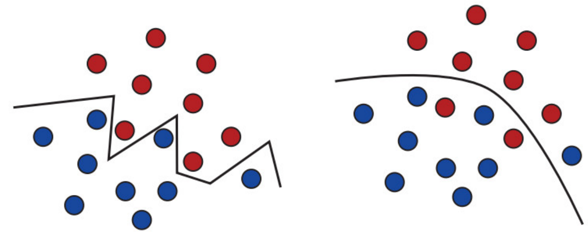
Generalization: ability of a learning algorithm to perform accurately on new, unseen examples after having experienced.

Overfitting



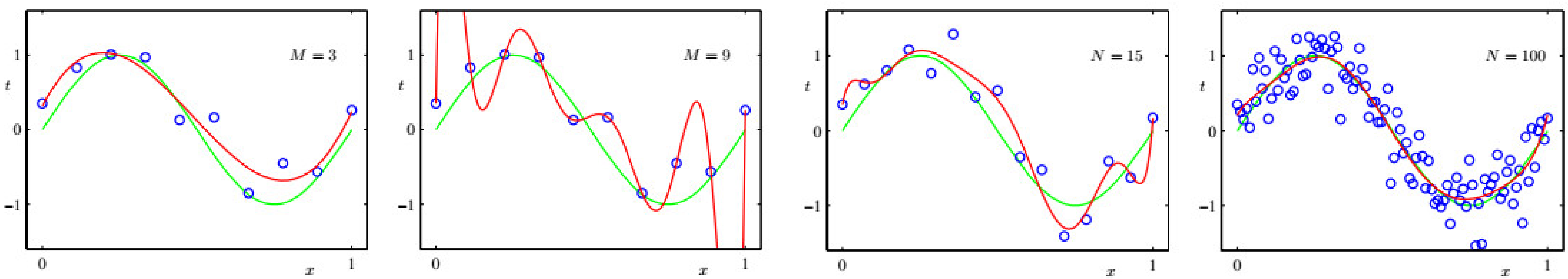
- Overfitting happens when a model learns the detail and noise in the training data but it loses the performance on new (test) data.
- Bad Generalization: Model fails to generalize unseen samples.

Overfitting



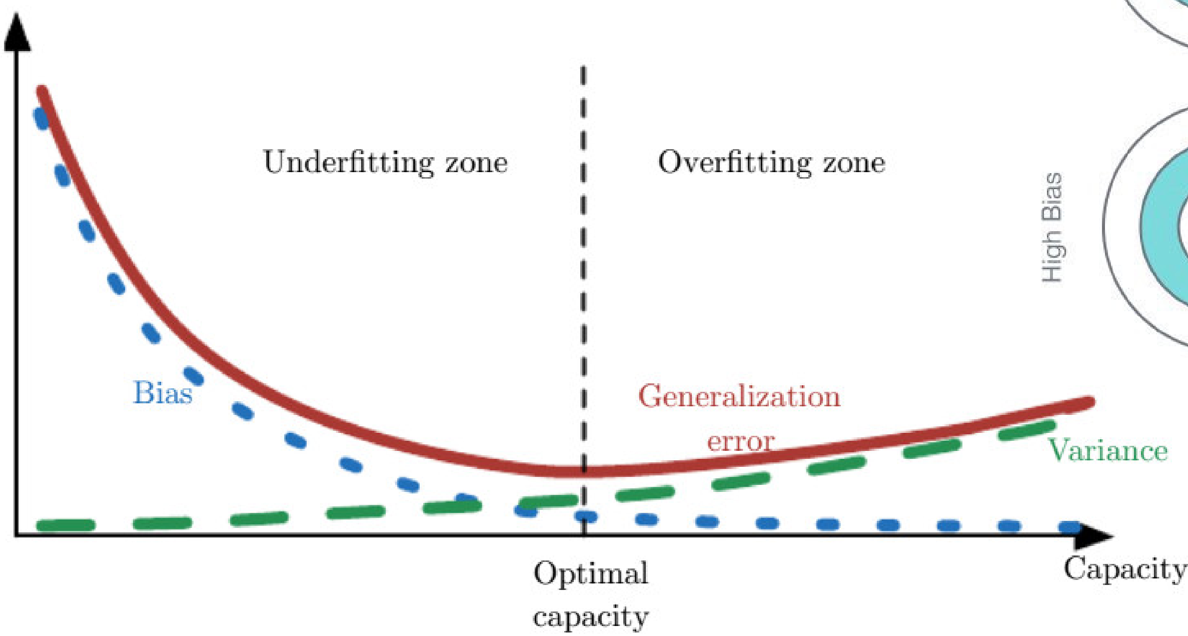
- Overfitting happens when a model learns the detail and noise in the training data but it loses the performance on new (test) data.
- Bad Generalization: Model fails to generalize unseen samples.
- Main reasons: a) Model complexity, b) Small data size

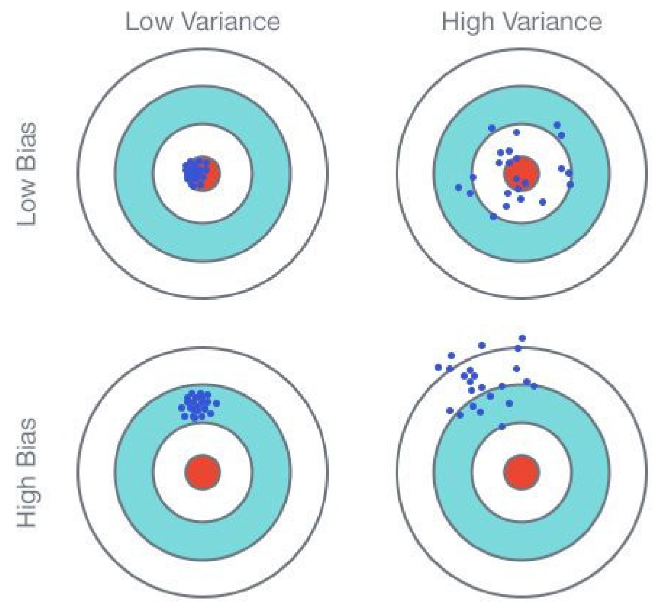
a b



[Bishop]

Trading off Bias and Variance

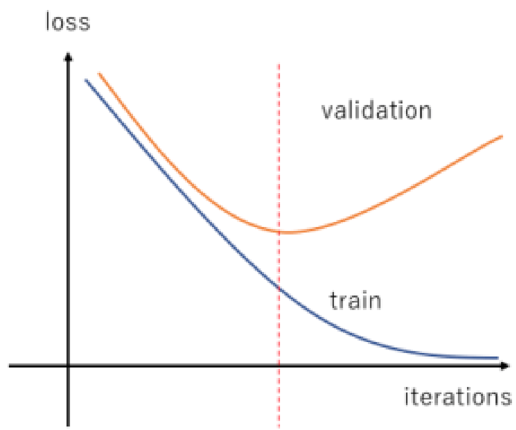




[Goodfellow, et al.]

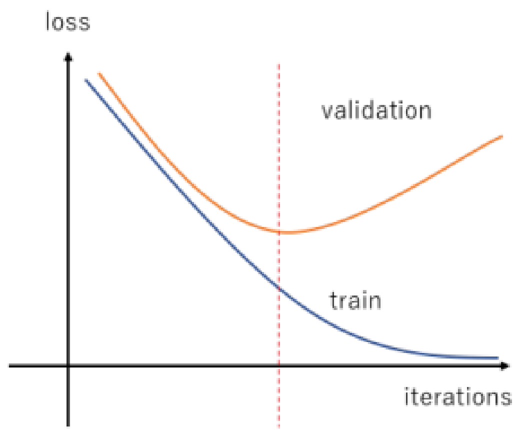
Model Selection and Validation

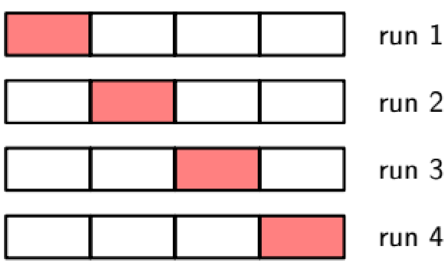
- Avoiding over-fitting: Determine a suitable model complexity based on validation error.



Model Selection and Validation

- Avoiding over-fitting: Determine a suitable model complexity based on validation error.





Cross-validation

Simple hold-out method

Related Resources

- Deep Learning (Goodfellow, et al.), chapter 5.
- Pattern Recognition and Machine Learning (Christopher Bishop), chapter 1,3,4.
- Foundations of Machine Learning (Mehryar Mohri), chapter 12.
- Understanding machine learning: From theory to algorithms (Shai Shalev-Shwartz), chapter 22.