



Machine
Learning

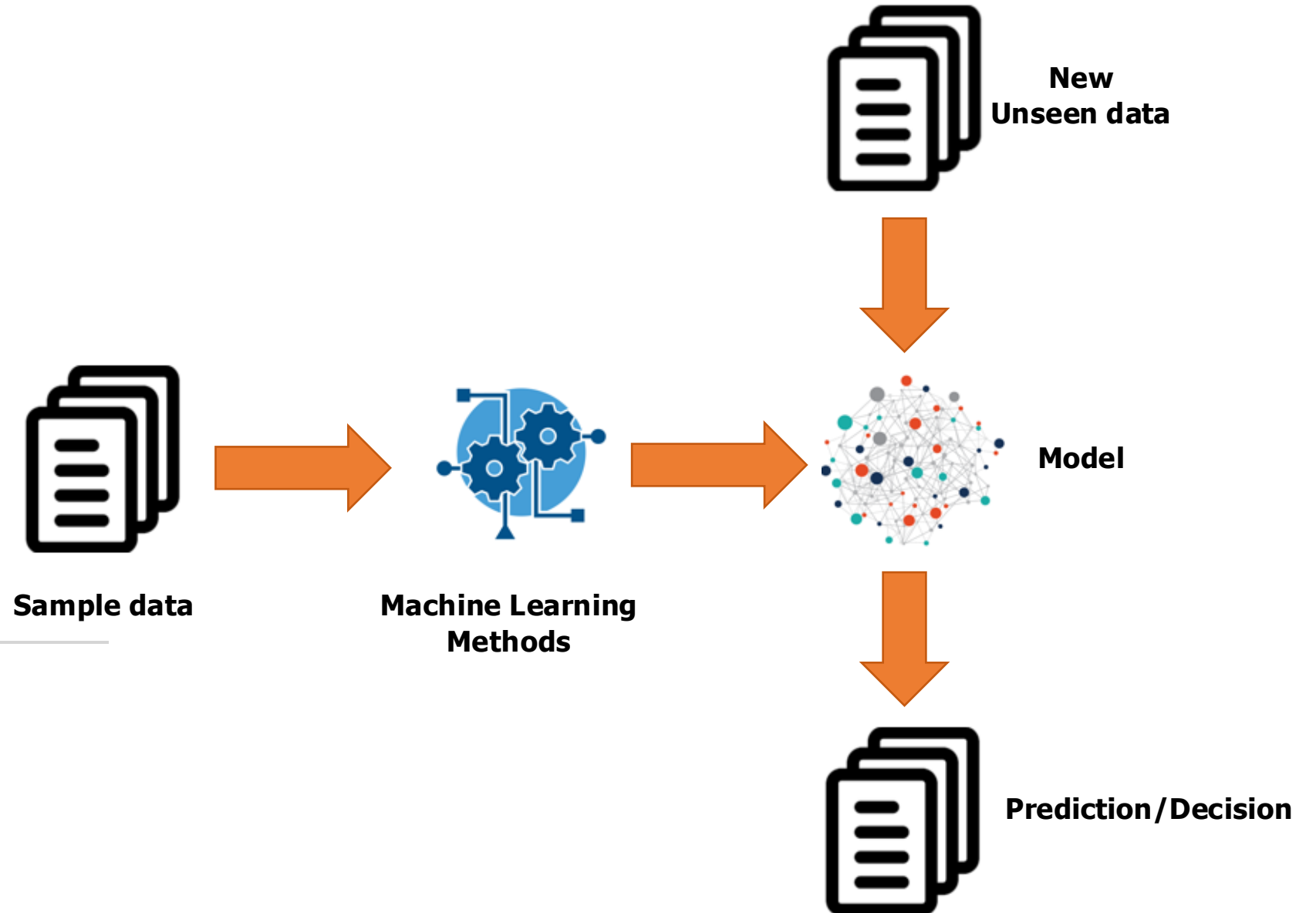
MML

-

1

- What is Machine Learning
- Learning Types
- Data Preparation
- Cross Validation
- Overfitting

What is Machine Learning ?

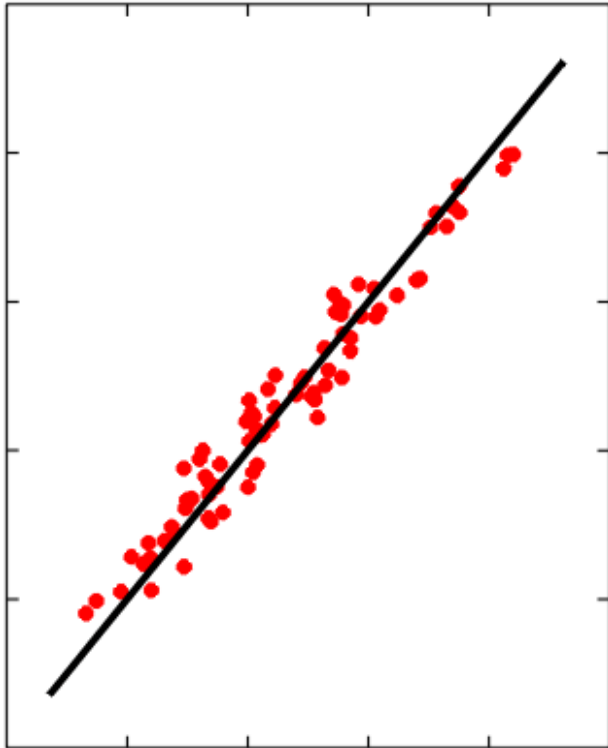


Learning Types

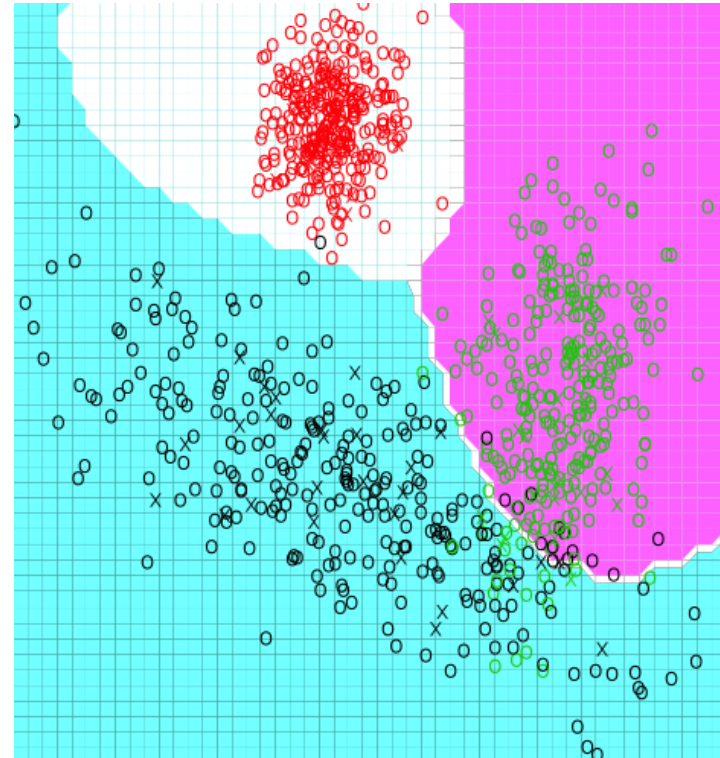
- **Supervised Learning**
 - Regression
 - Classification
- **Unsupervised Learning**
 - Clustering
 - Dimensionality Reduction
- **Reinforcement Learning**
 - Reward and Penalty

Supervised Learning

Regression



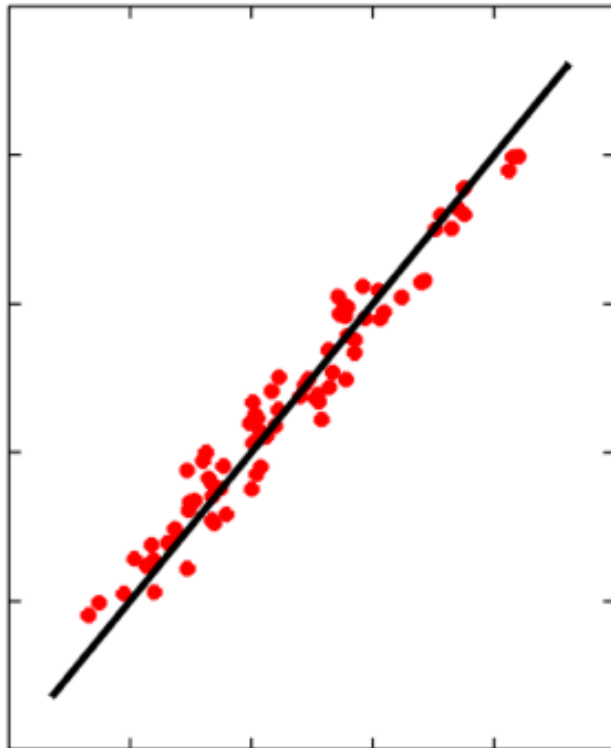
Classification



Supervised Learning

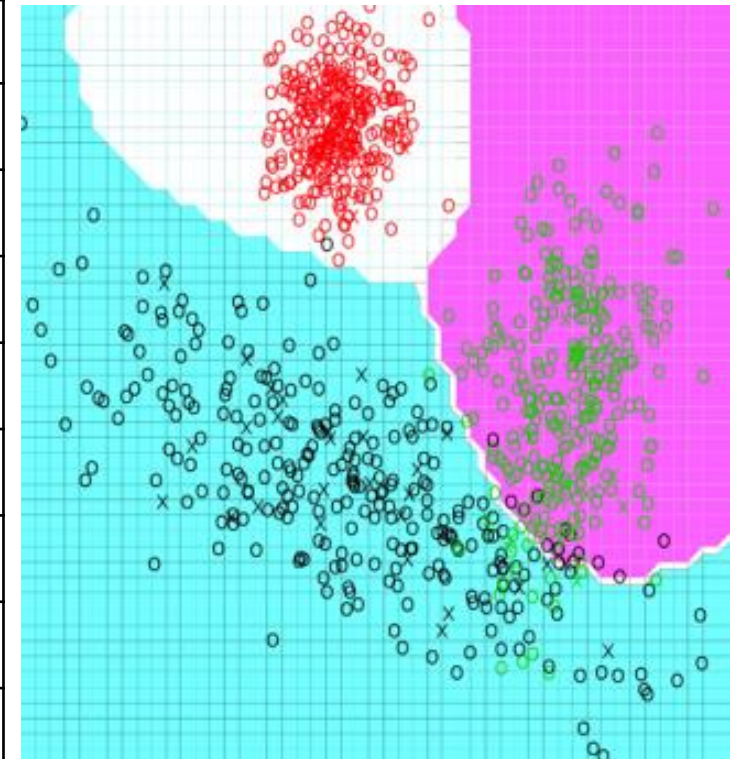
Regression

| X | D |
|----|----|
| 3 | 9 |
| 5 | 15 |
| 8 | 24 |
| 10 | 30 |
| 19 | 57 |
| 24 | 72 |
| 27 | 81 |
| 31 | 93 |



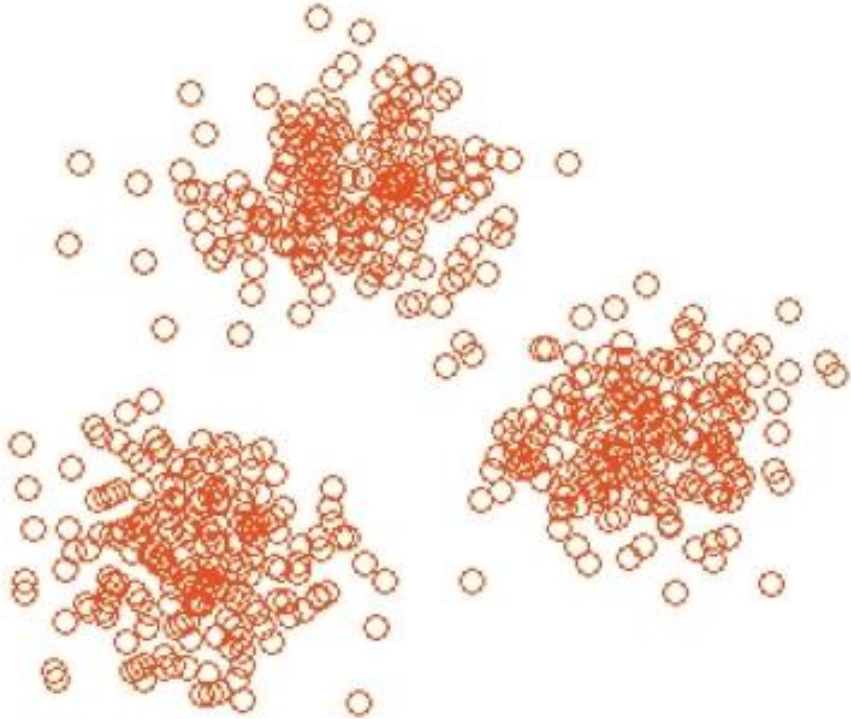
Classification

| X_1 | X_2 | D |
|-------|-------|---|
| 10 | 6 | A |
| 12 | 5 | A |
| 20 | 9 | B |
| 24 | 11 | B |
| 30 | 19 | B |
| 13 | 15 | C |
| 16 | 17 | C |
| 18 | 19 | C |

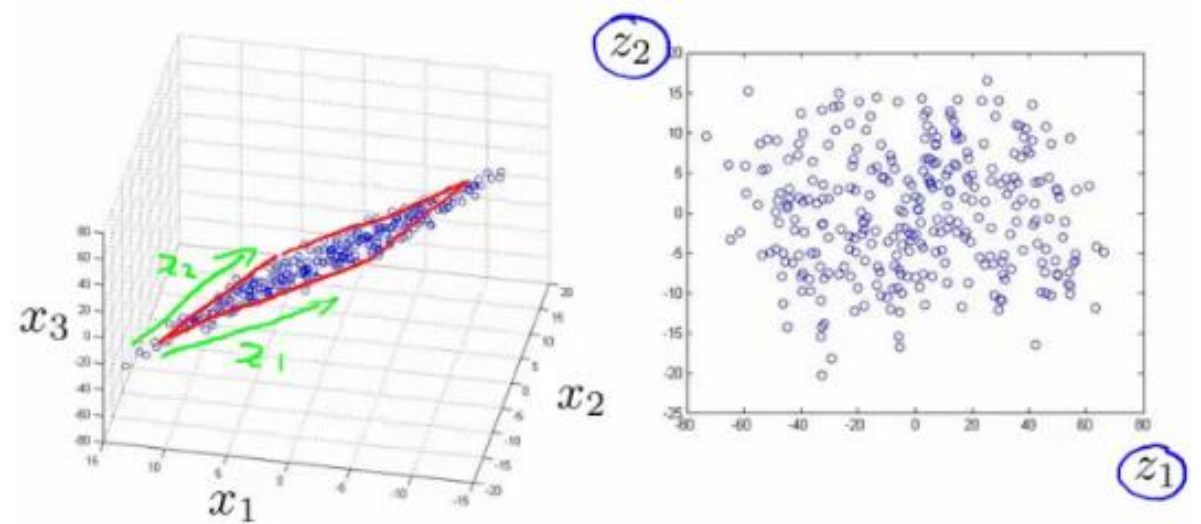


Unsupervised Learning

Clustering

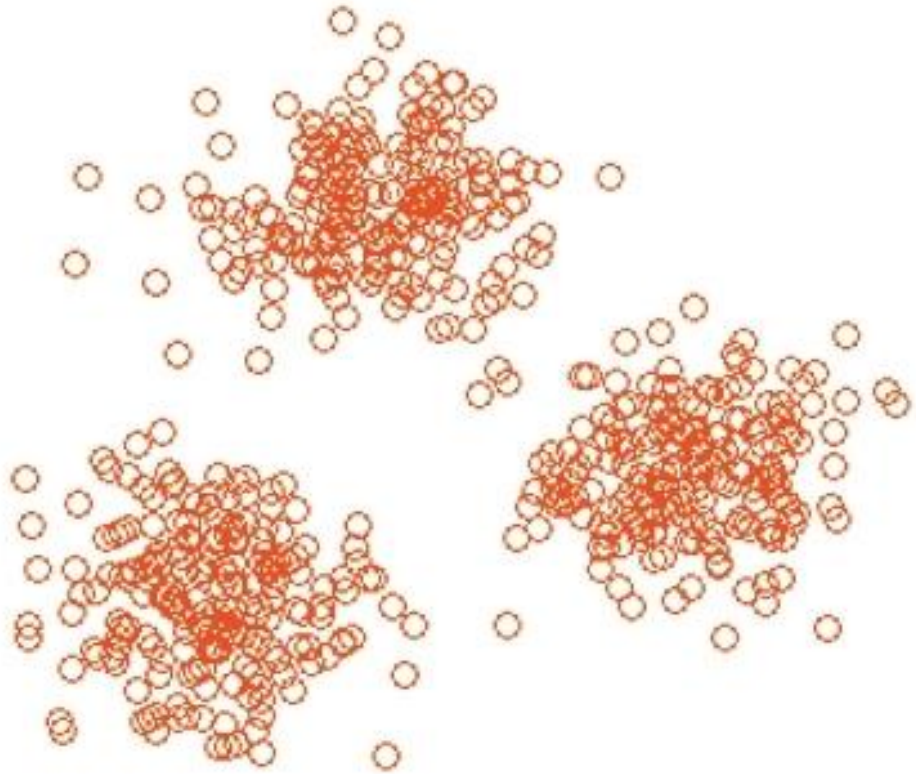


Dimensionality Reduction



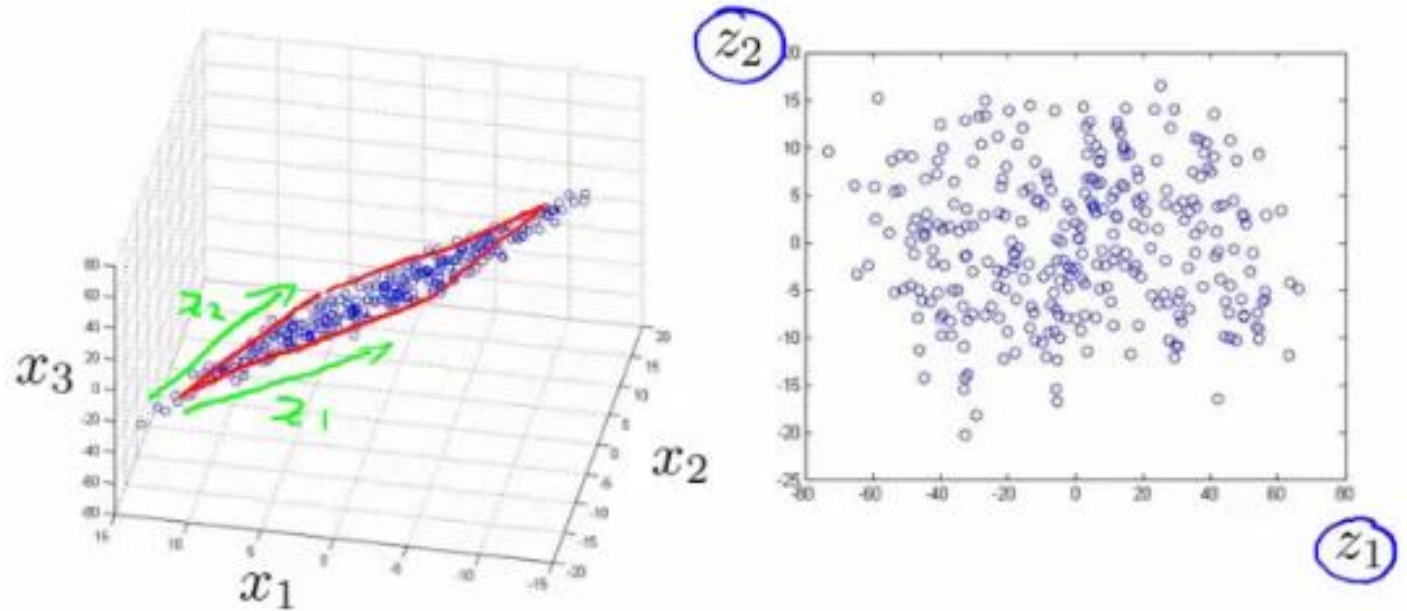
Unsupervised Learning

Clustering dealing with unlabeled data is the process of organizing similar objects into the same groups and dissimilar objects into different groups.



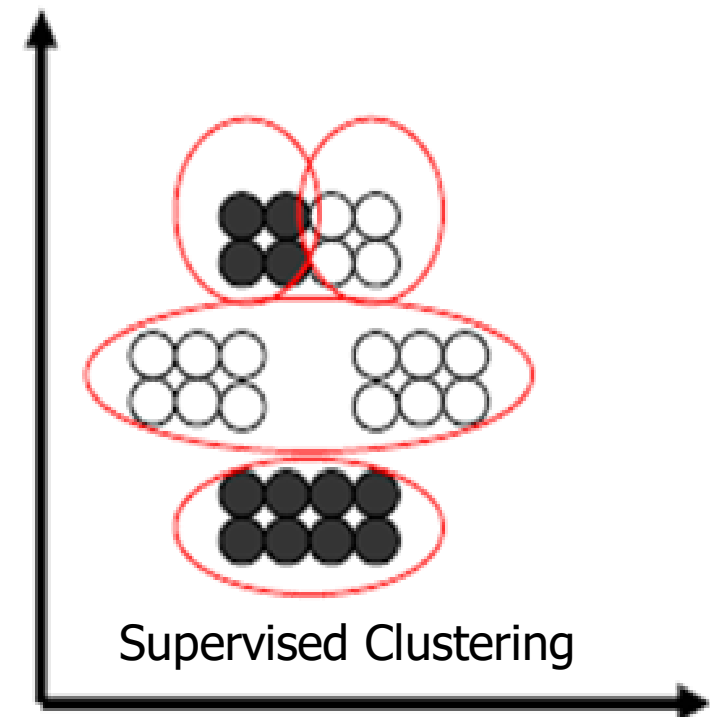
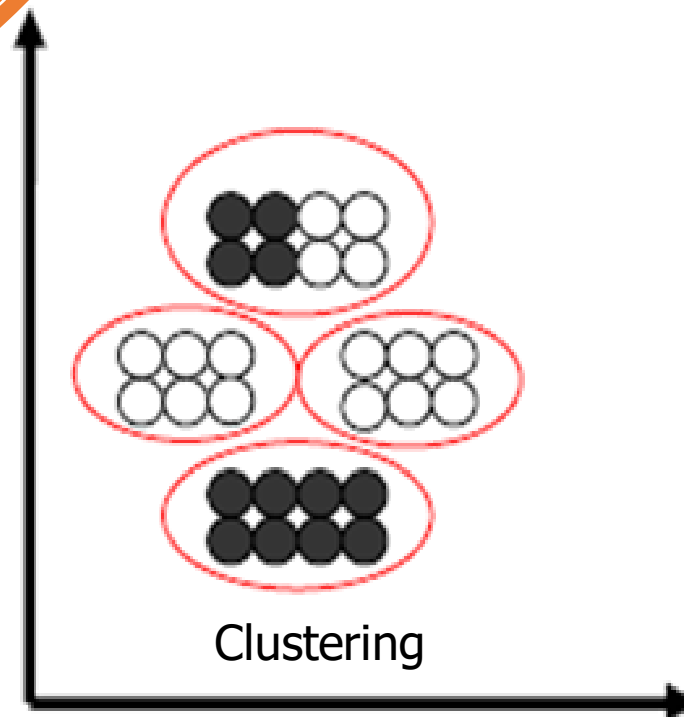
Unsupervised Learning

Dimensionality reduction is a process used in machine learning and data analysis to reduce the number of input variables in a dataset.



Hybrid Methods

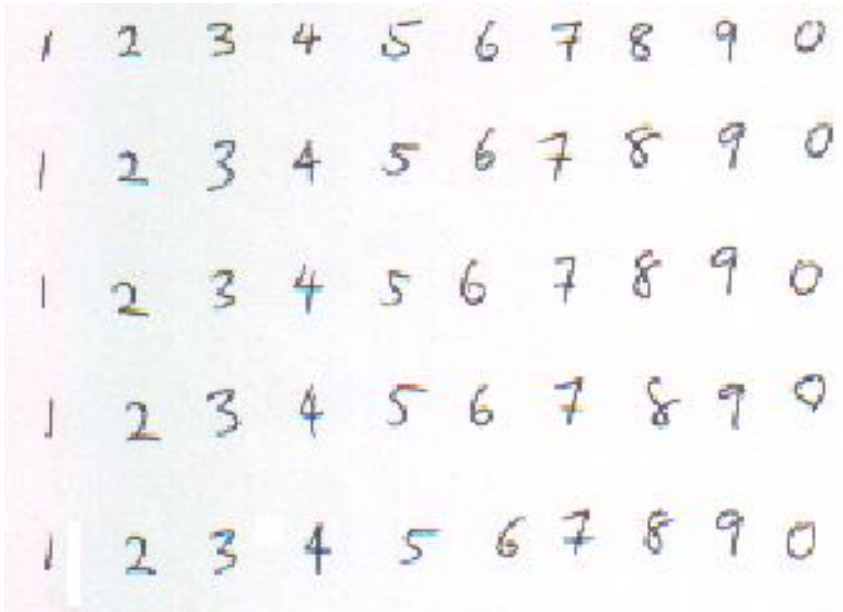
Supervised clustering is a hybrid approach that integrates elements of both supervised learning and traditional clustering.



Reinforcement Learning

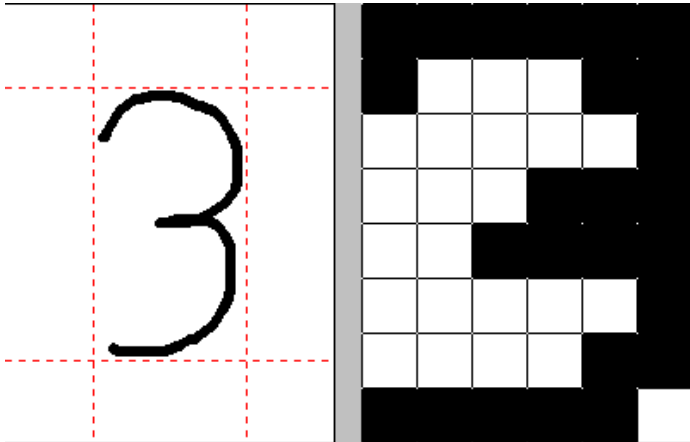
- Act by trial-and-error
- Interacting environment
- Feedback(Reward/Penalty)
- Improve decisions by maximizing rewards





Data Preparation

- Improve Data Quality
- Feature Selection and Extraction
- Data Scaling and Normalization



Data Preparation

- Imbalanced datasets
- Data Augmentation
- Data splitting

A 5x10 grid of handwritten digits from 1 to 0. A horizontal red dashed line is drawn between the second and third rows, representing a split by row. This method ensures that each row in the training set contains all classes of digits, while each row in the test set also contains all classes, maintaining the class distribution across both sets.

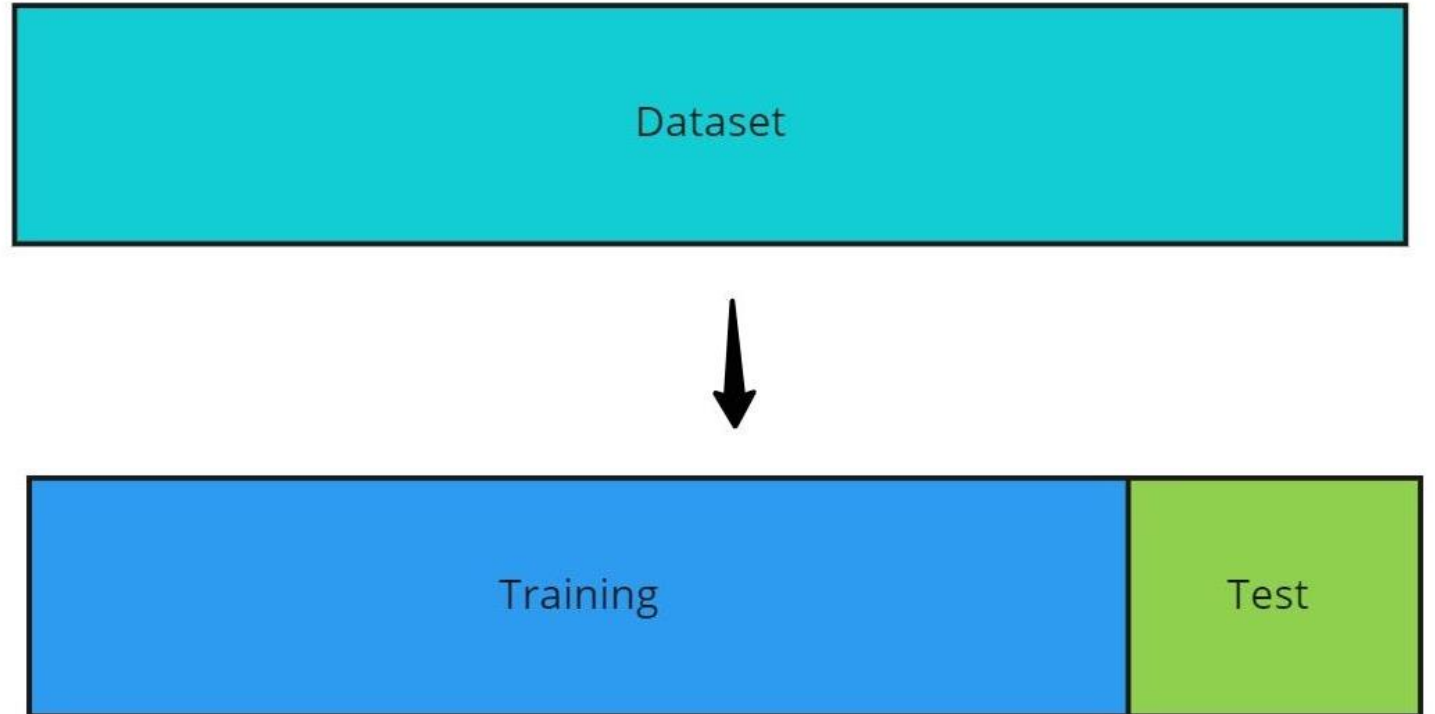
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |

A 5x10 grid of handwritten digits from 1 to 0. A vertical red dashed line is drawn between the fifth and sixth columns, representing a split by column. This method results in the training set containing only digits 1-5 and the test set containing only digits 6-0, which is an incorrect split as it does not maintain the class distribution across both sets.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |

Cross Validation

The system is trained using the training set, and its success is measured on the test set.

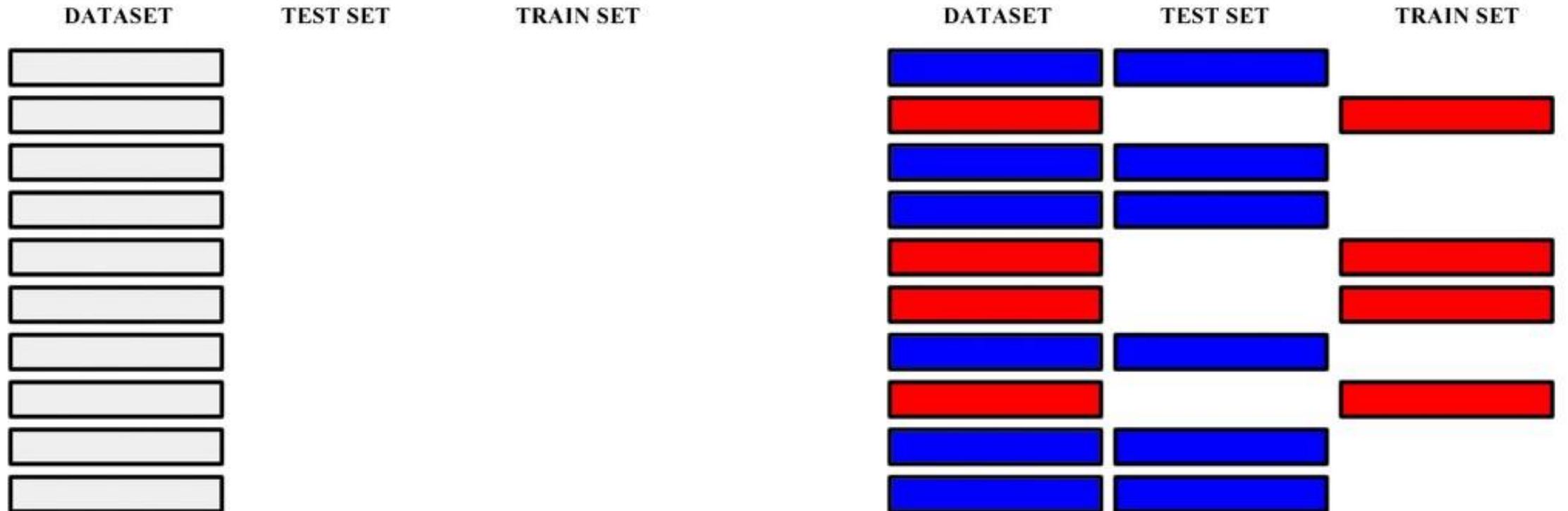


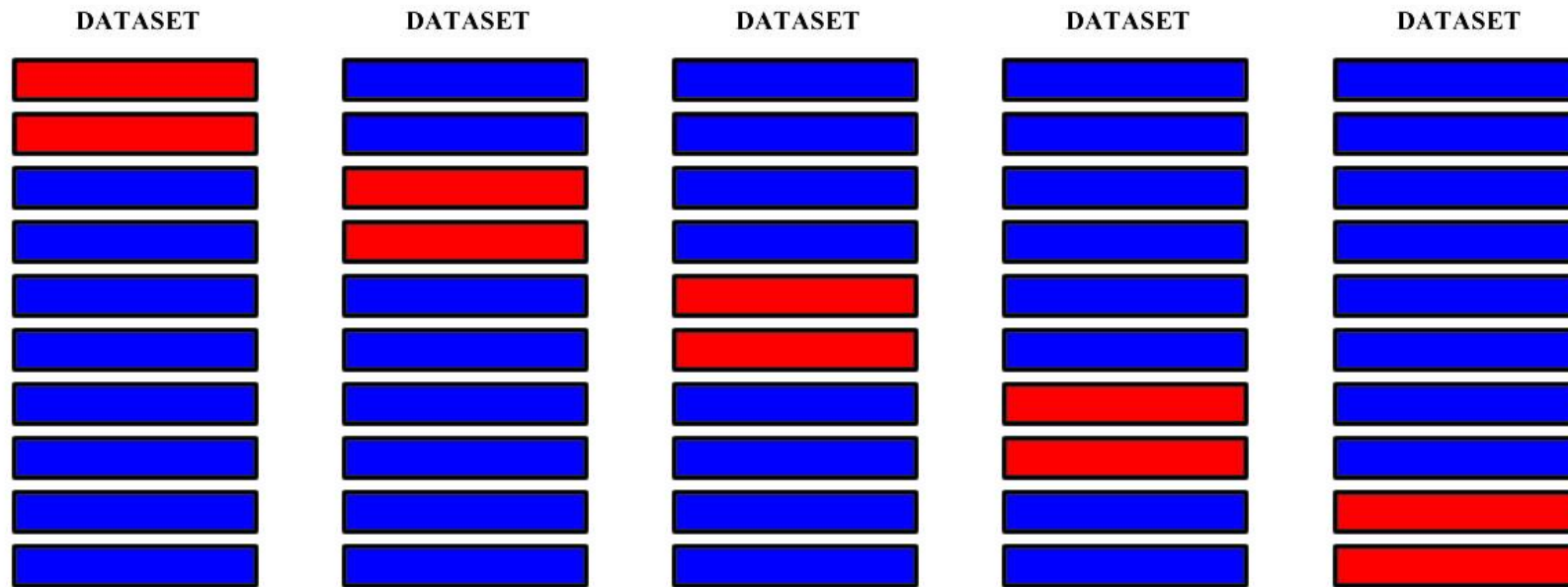
Cross Validation

Some methods of cross-validation are:

- Random sampling
- K-fold
- Leave one out

Random Sampling





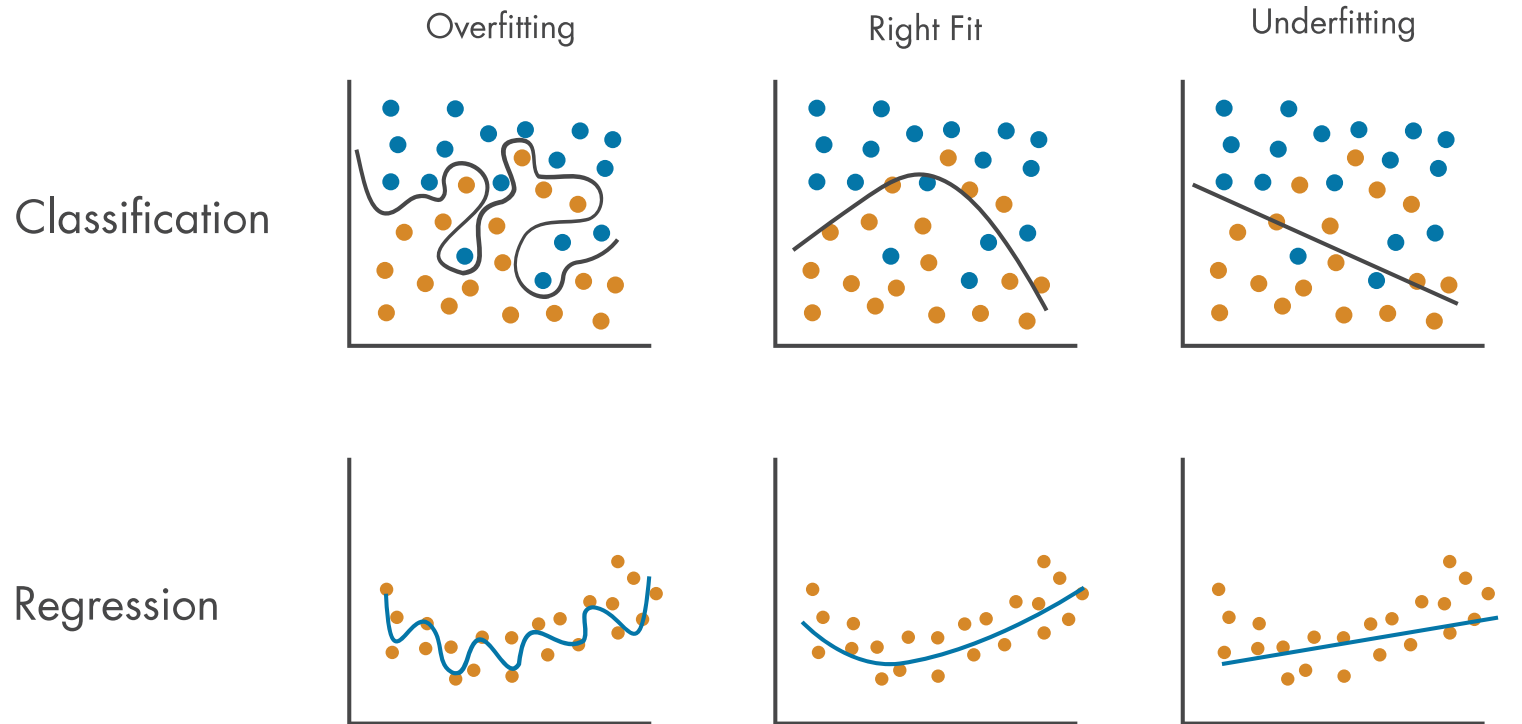
K-Fold

Red folds show test-set, Blue ones are train-set.

Leave One Out

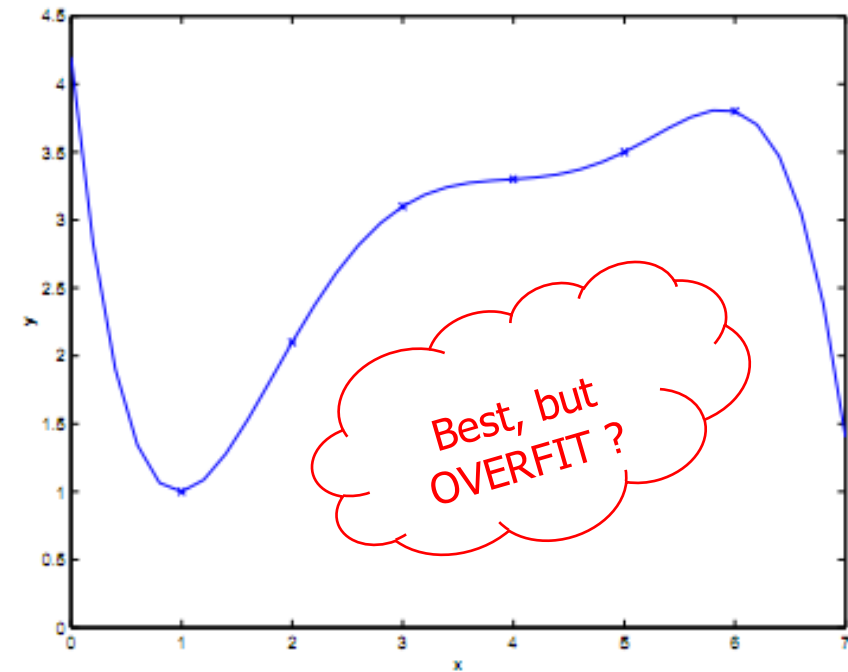
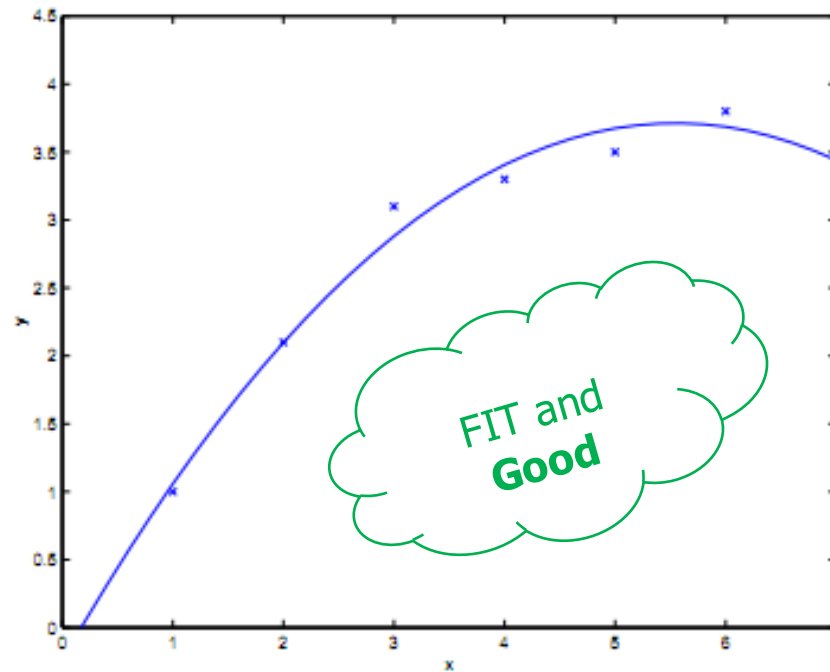
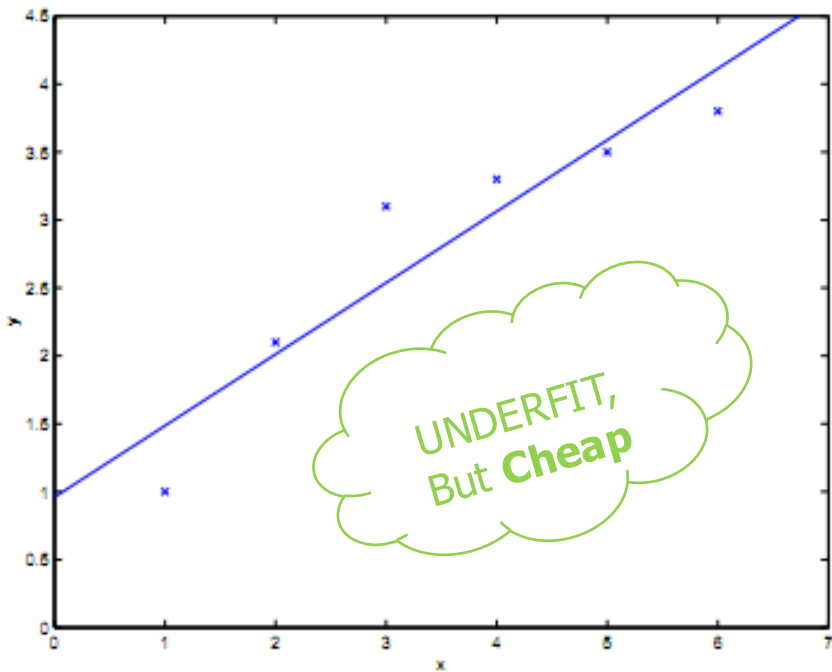
- It is a special version of K-fold ($K = N$).
- For a dataset with N samples, if we choose number of folds (K) as number of samples (N), then...
- K-fold runs as Leave One Out method.

Overfitting



Which model?

Machine learning methods can be different to each other according to their approach to problem and therefore may have different success in different problems.



A decorative graphic on the left side of the slide, consisting of a complex, overlapping pattern of blue triangles and polygons in various shades of blue, creating a faceted, crystalline appearance.

Machine Learning

1. week



Thanks for watching