Supervised Machine Learning

CSC 380 - Principles of Data Science

Lecture 6.2

In this lecture

- ML Terminology:
  - Labels, Features, Examples, Models
  - Training, Inference/Testing,
- Linear Regression:
  - Equation, Loss, Update weights based on Loss
  - LASSO, RIDGE
- Gradient Descent
    - Stochastic, Batch, Mini-batch stochastic

In this lecture

- Loss Functions :
  - Esp MAE, MSE
- Learning rate
- Overfitting v/s Underfitting
- Data -
  - Split as (Train, Val, test)
  - K Cross validation
- Regularisation :
  - Reducing Model Complexity:
    - L1/L2 Regularisation
    - Dropout
    - Early stopping
  - Data Augmentation

Terminology

- Labels : The thing we are predicting
- Features : Input variable, there could be more than one.
- Examples (prev Data Point) :
  - {features, label}: (x, y)
  - Labeled and Unlabeled
- Models :
  - Something that defines the relationship between features and labels.

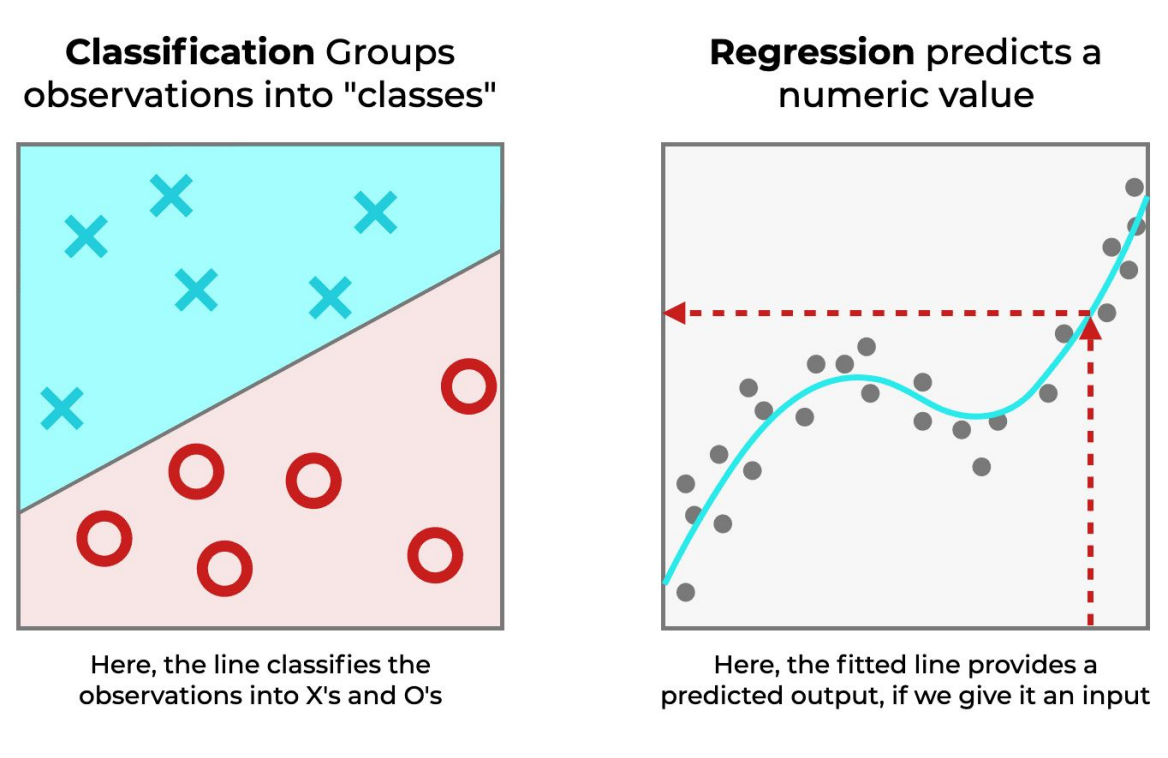
Terminology cont..

Training : Creation of the model.. Ie the process through which the model learns a relationship from different examples.

Inference/Testing : Getting predictions from the model with only the feature provided.

Data : Train, Validation, Test

Regression vs. classification



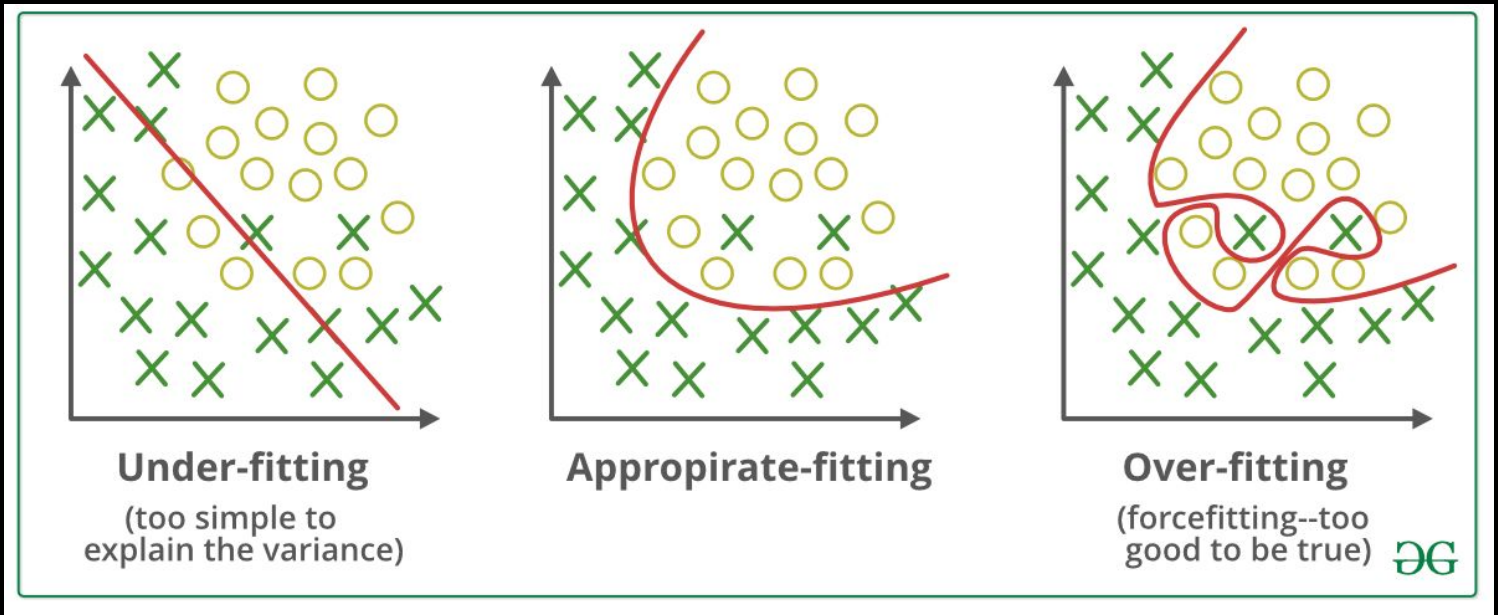
Source

Linear Regression

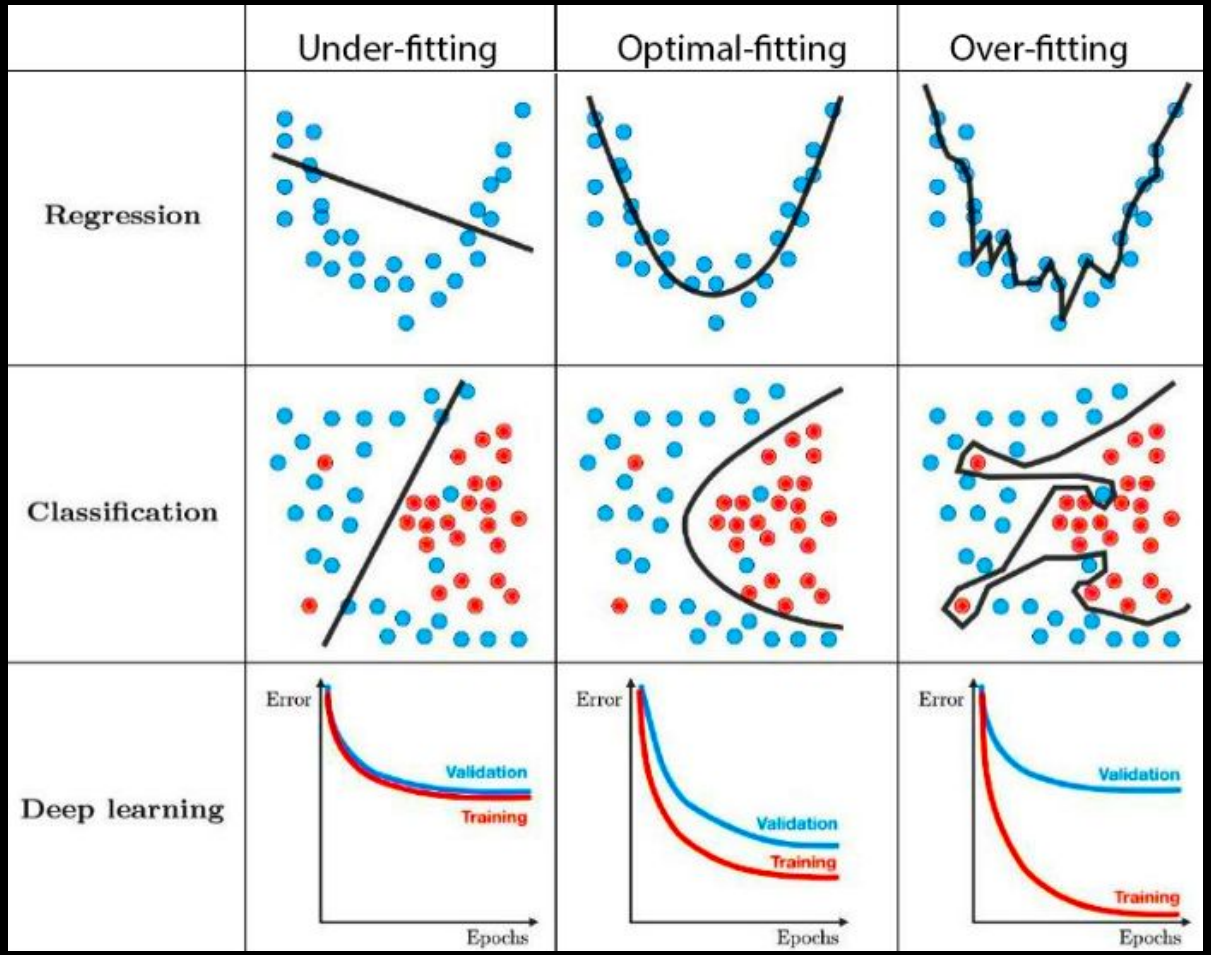
Switch to Board

Fitting - Over , Under and Just Right

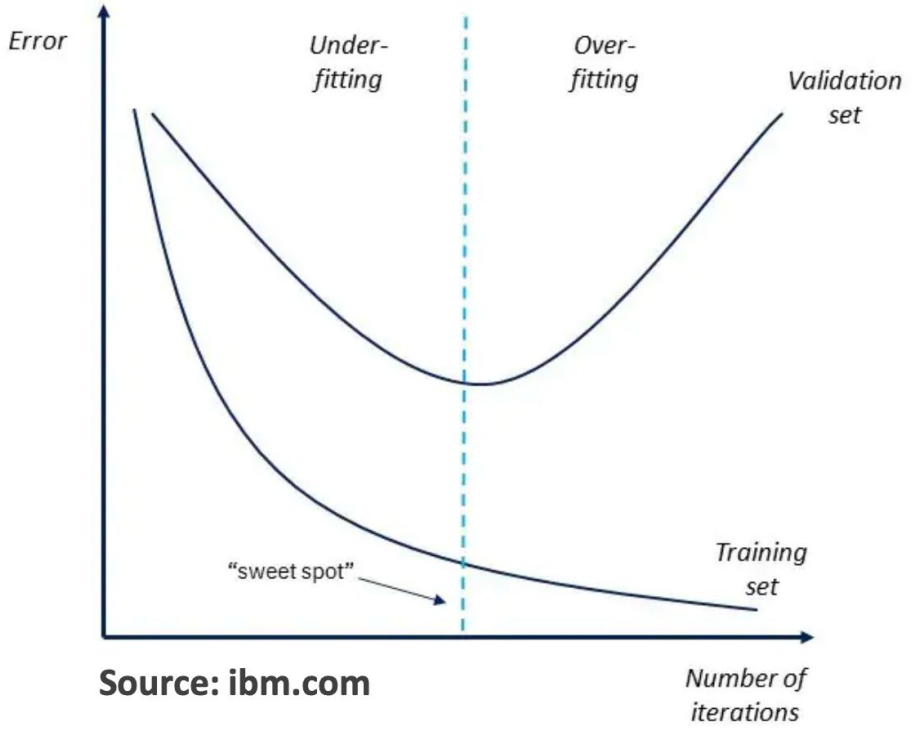
Overfit vs Underfit



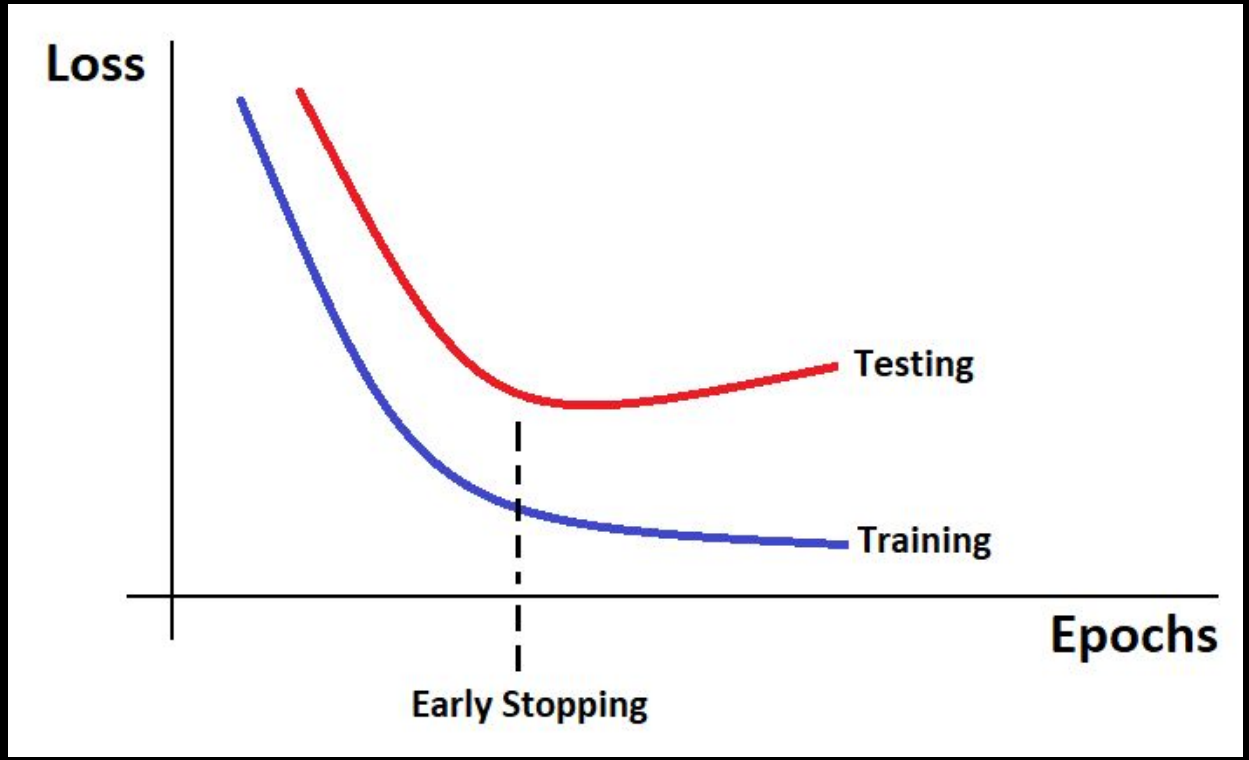
Source - Geeks for Geeks



Towards Data Science

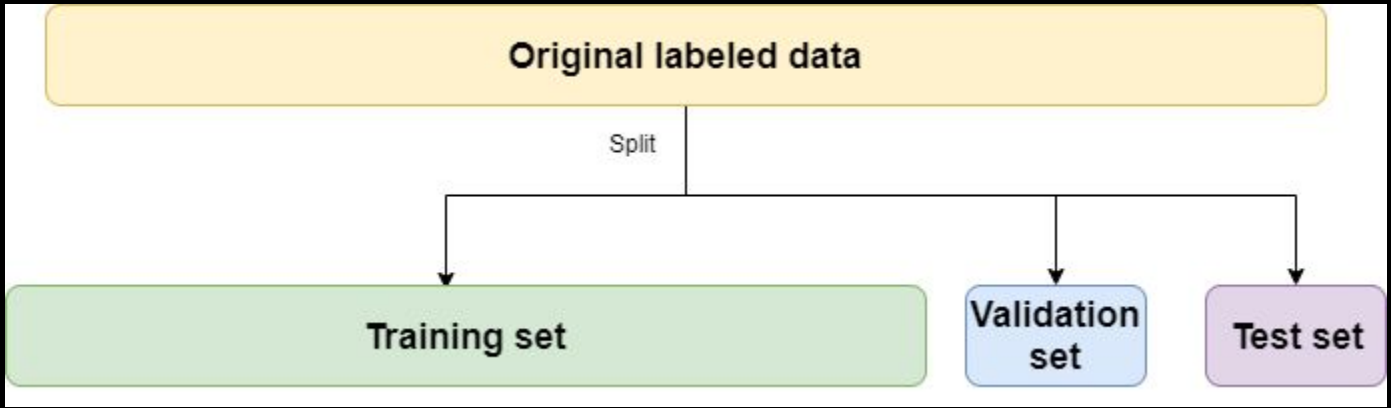


Source: ibm.com

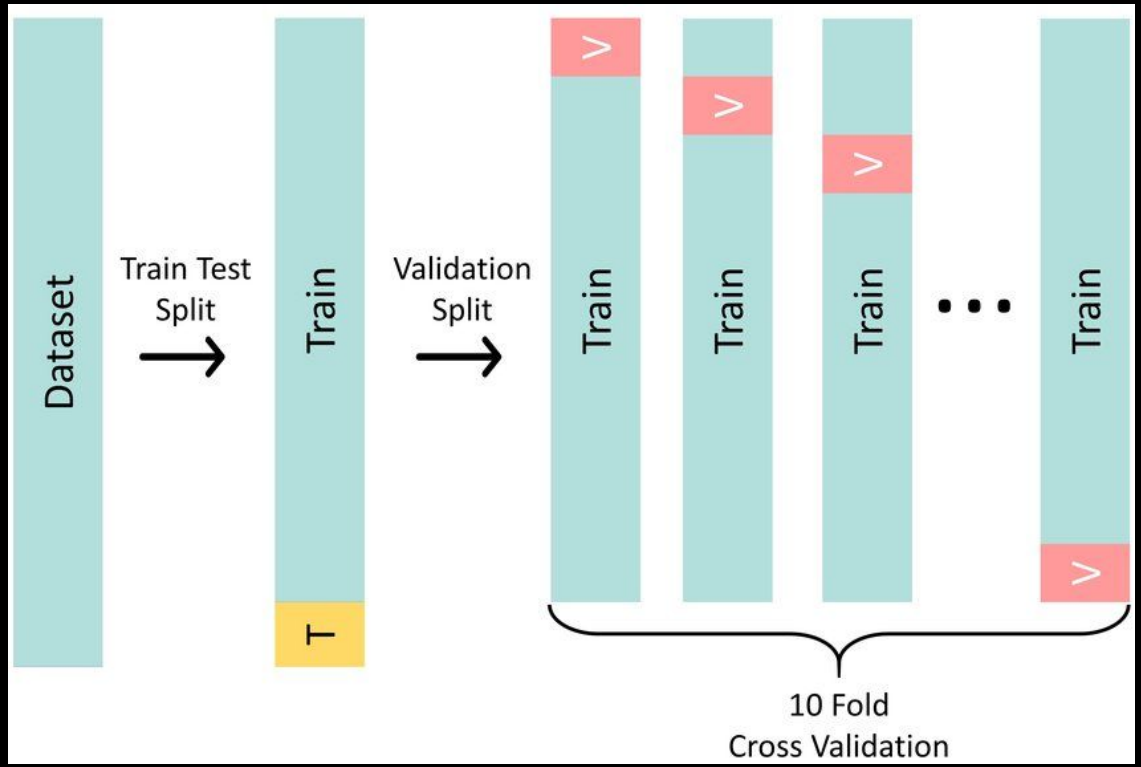


Towards AI

Train - Val - Test Split

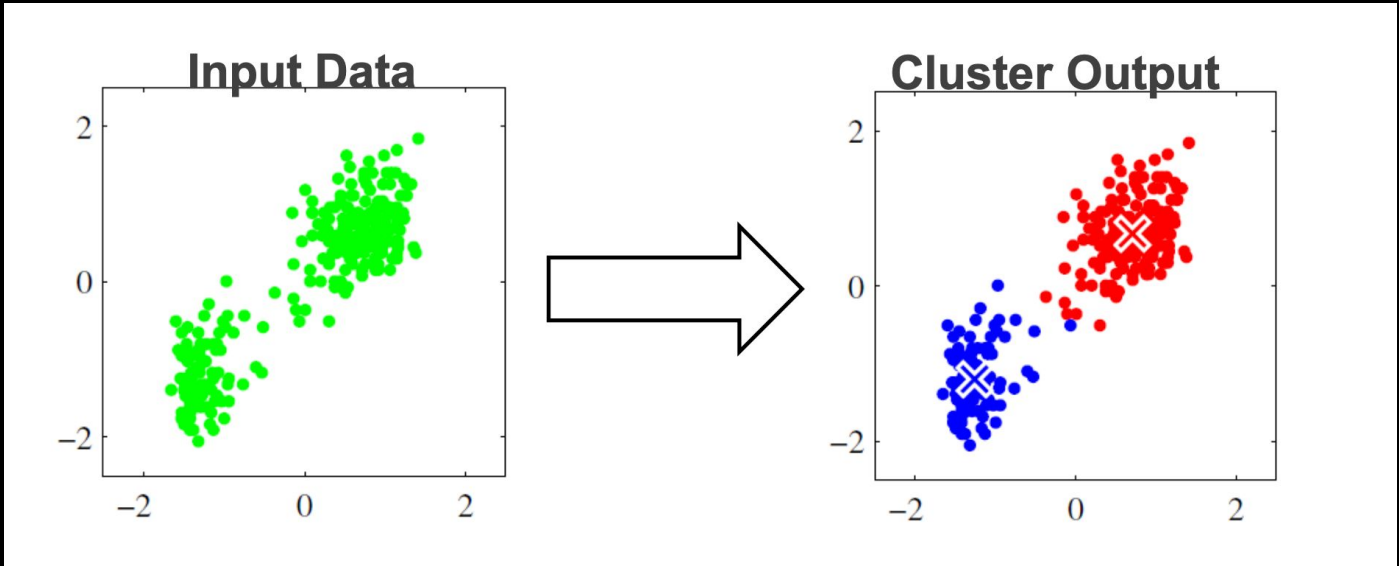

Original labeled data
Split
Training set
Validation set
Test set

K-fold Cross validation

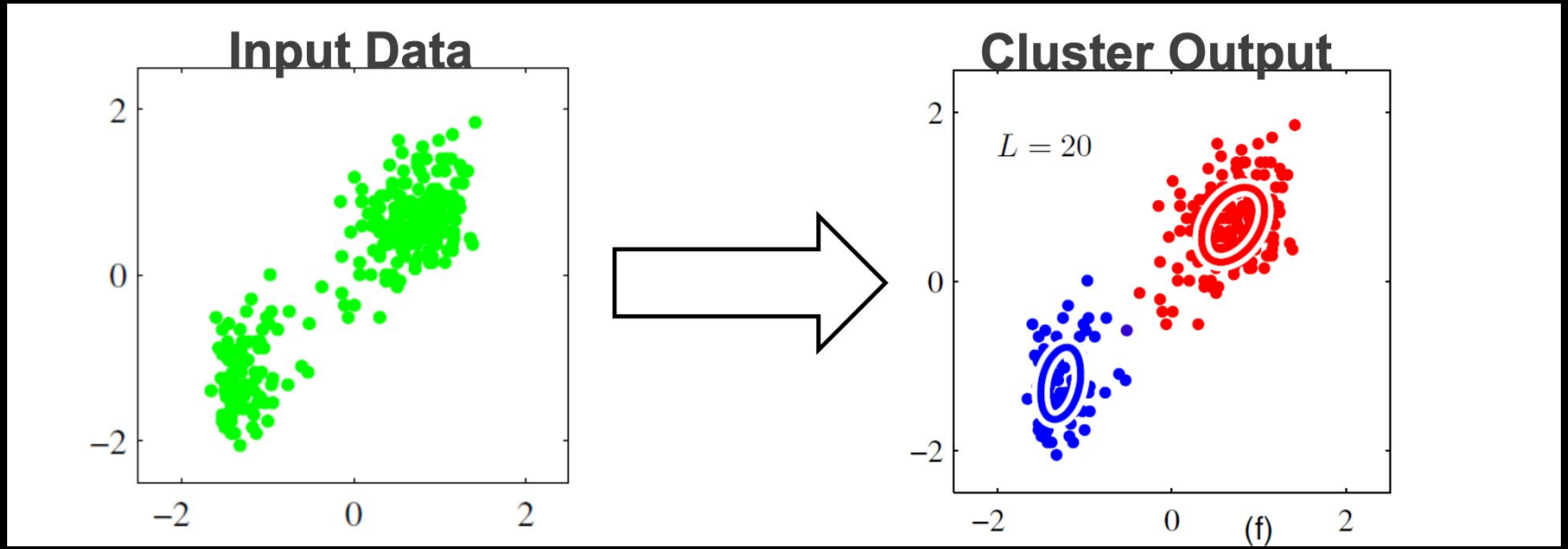


ResearchGate

Non-Probabilistic ex: K-Means



Probabilistic ex: Gaussian Mixture Models



Resources :

- Google ML Crash Course
- Prof. Jason Pacheco Slides