Introduction

Machine Learning

Computer and Systems Engineering

Edwin Puertas, Ph.D(c).
epuerta@utb.edu.co

Definition of Machine Learning

Arthur Samuel (1959): Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed.

Tom Mitchell (1998): 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.

Taxonomy of Machine Learning

Supervised Learning

Unsupervised Learning

Reinforcement Learning

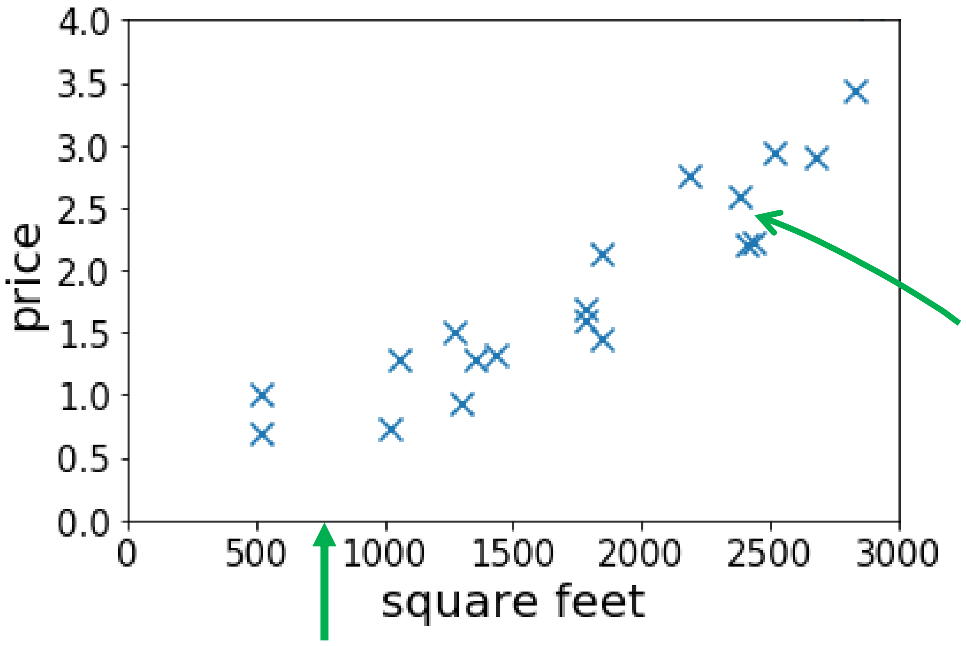
Supervised Learning

Housing Price Prediction

- Given: a dataset that contains n samples

$$\left(x^{(1)}, y^{(1)}\right), \dots \left(x^{(n)}, y^{(n)}\right)$$

- Task: if a residence has x square feet, predict its price?



15th sample
$\left(x^{(15)}, y^{(15)}\right)$

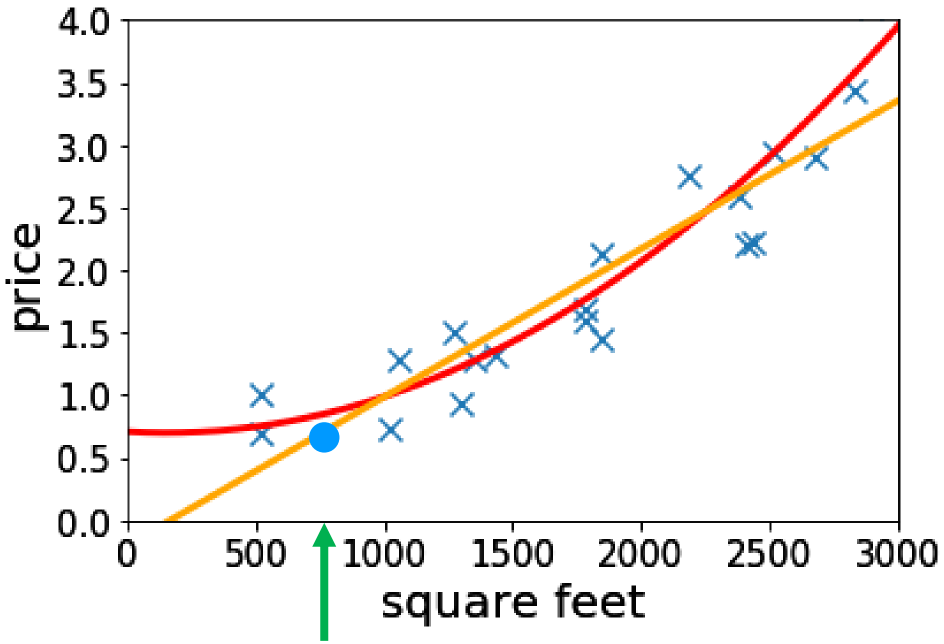
$x = 800$
$y = ?$

Housing Price Prediction

Given: a dataset that contains n samples

$$\left(x^{(1)}, y^{(1)}\right), \ldots \left(x^{(n)}, y^{(n)}\right)$$

Task: if a residence has x square feet, predict its price?

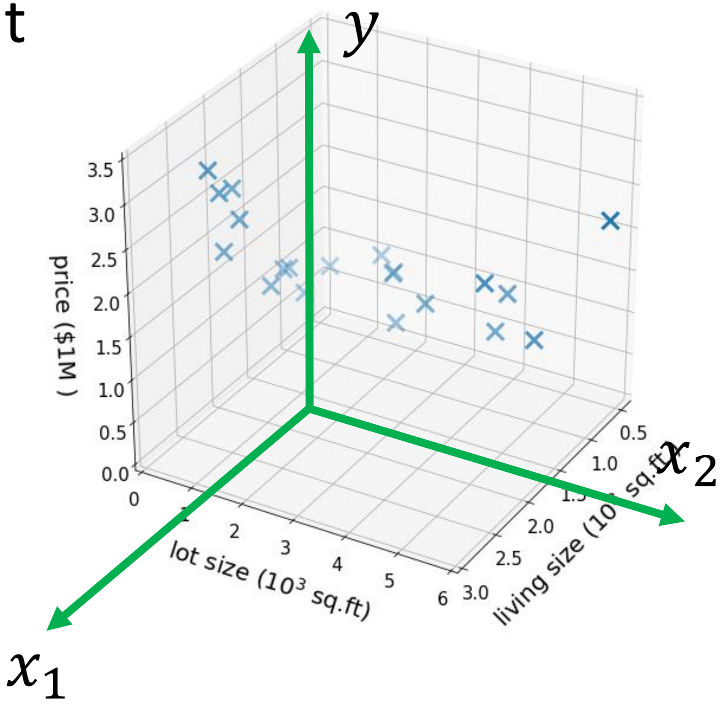


$$x = 800$$
$$y = ?$$

More Features

- Suppose we also know the lot size
- Task: find a function that maps

$$\underbrace{\text{(size, lot size)}}_{\substack{\text{features/input} \\ x \in \mathbb{R}^2}} \quad \rightarrow \quad \underbrace{\text{price}}_{\substack{\text{label/output} \\ y \in \mathbb{R}}}$$



➢ Dataset: $\left(x^{(1)}, y^{(1)}\right), \ldots, \left(x^{(n)}, y^{(n)}\right)$

where $x^{(i)} = (x_1^{(i)}, x_2^{(i)})$

➢ "Supervision" refers to $y^{(1)}, \ldots, y^{(n)}$

High-dimensional Features

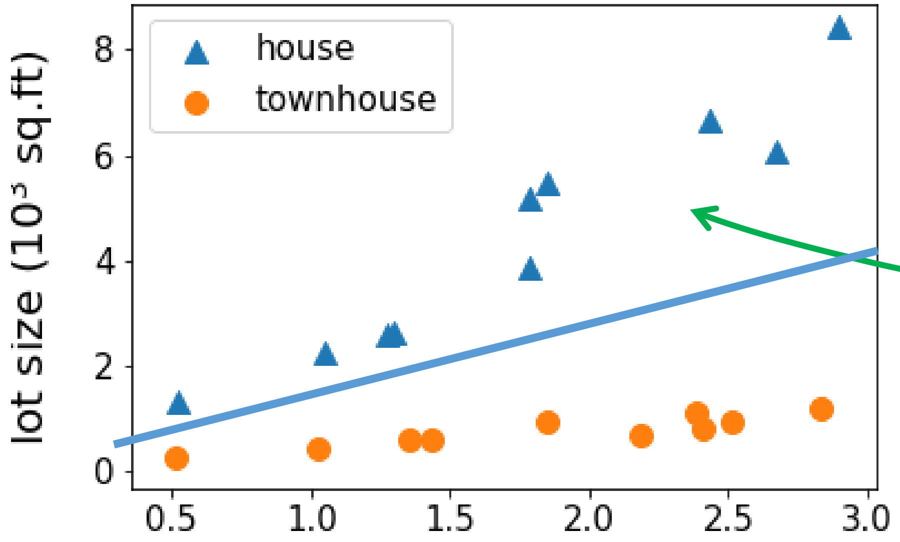
- $x \in \mathbb{R}^d$ for large d
- E.g.,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ \vdots \\ x_d \end{bmatrix} \begin{array}{l} \text{--- living size} \\ \text{--- lot size} \\ \text{--- \# floors} \\ \text{--- condition} \\ \text{--- zip code} \\ \qquad\quad \vdots \end{array} \longrightarrow \quad y\text{--- price}$$

Regression vs Classification

- regression: if $y \in \mathbb{R}$ is a continuous variable
  - e.g., price prediction
- classification: the label is a discrete variable
  - e.g., the task of predicting the types of residence

(size, lot size) $\rightarrow$ house or townhouse?



y = house or townhouse?

Lecture 3&4: classification

Supervised Learning in Computer Vision

Image Classification

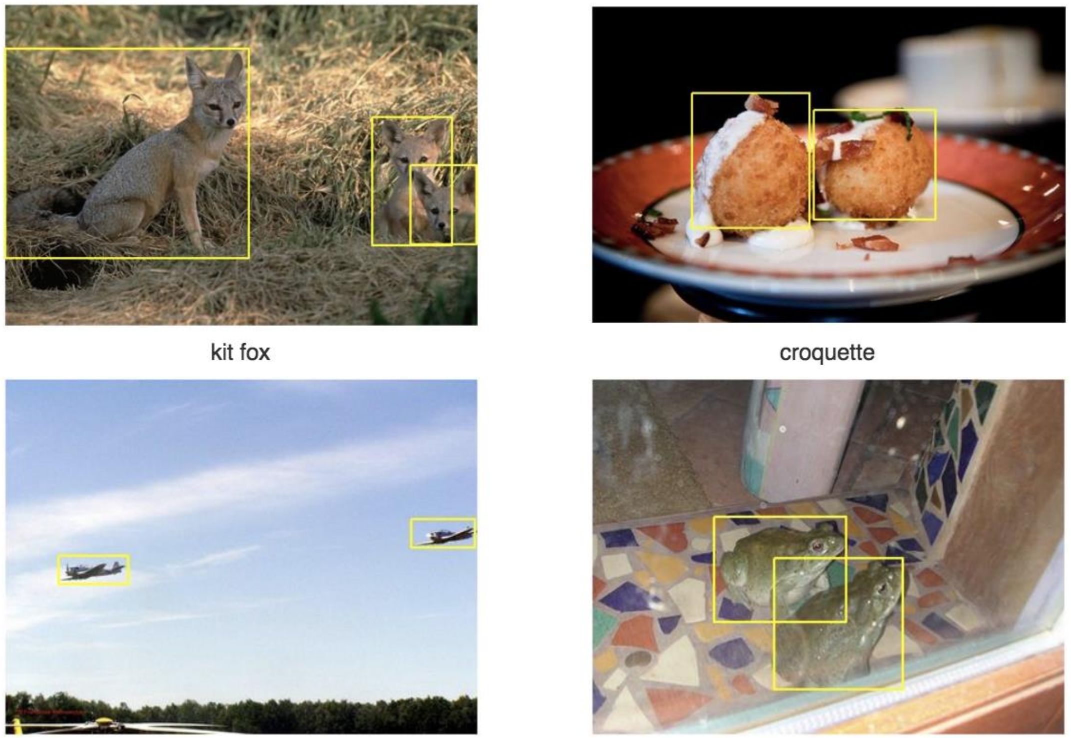
$x =$ raw pixels of the image, $y =$ the main object

ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Supervised Learning in Computer Vision

- Object localization and detection

  x = raw pixels of the image, y = the bounding boxes



kit fox

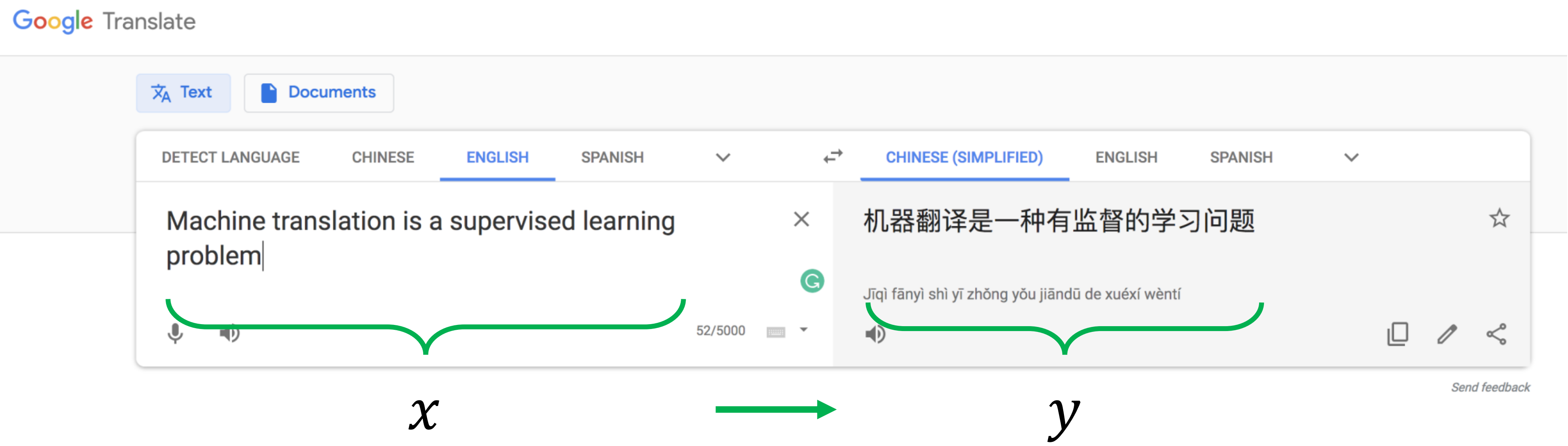
croquette

airplane

frog

ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Supervised Learning in Natural Language Processing



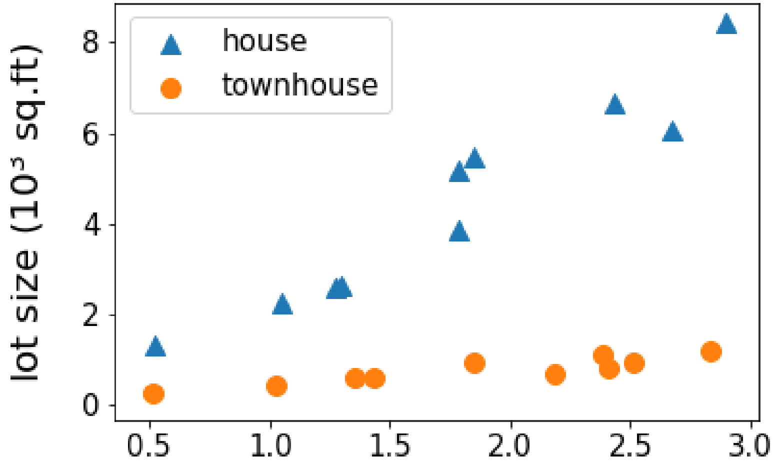
Machine translation

Unsupervised Learning

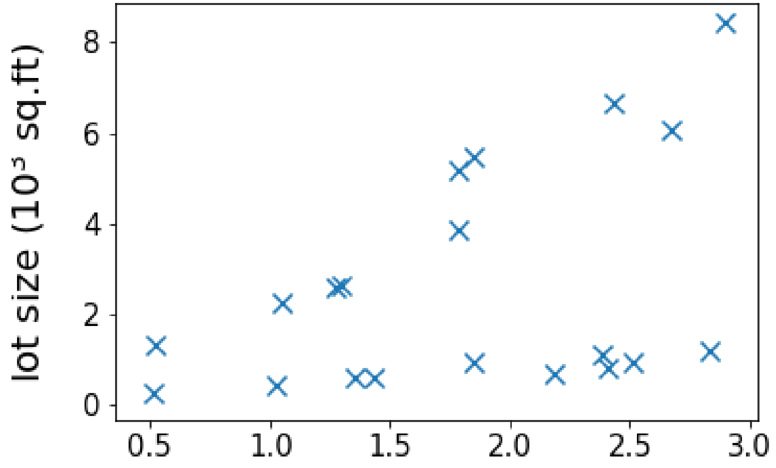
Unsupervised Learning

- Dataset contains no labels: $x^{(1)}, \dots x^{(n)}$
- Goal (vaguely-posed): to find interesting structures in the data

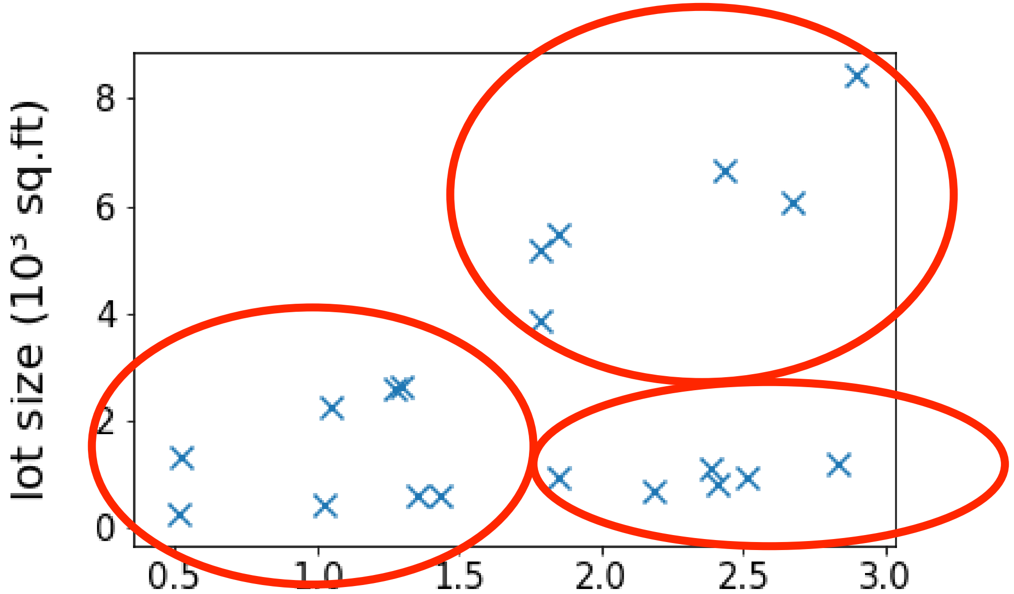
supervised



unsupervised

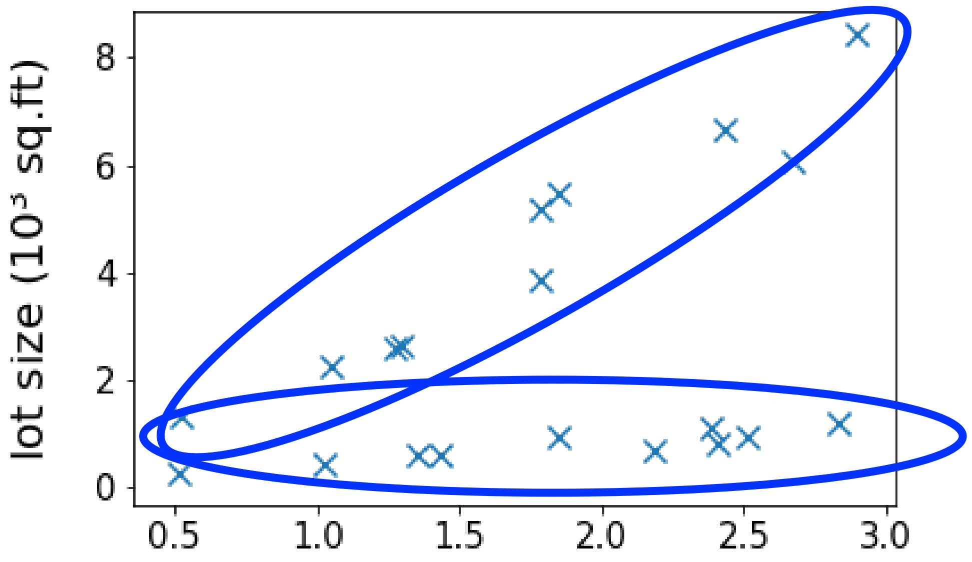


Clustering

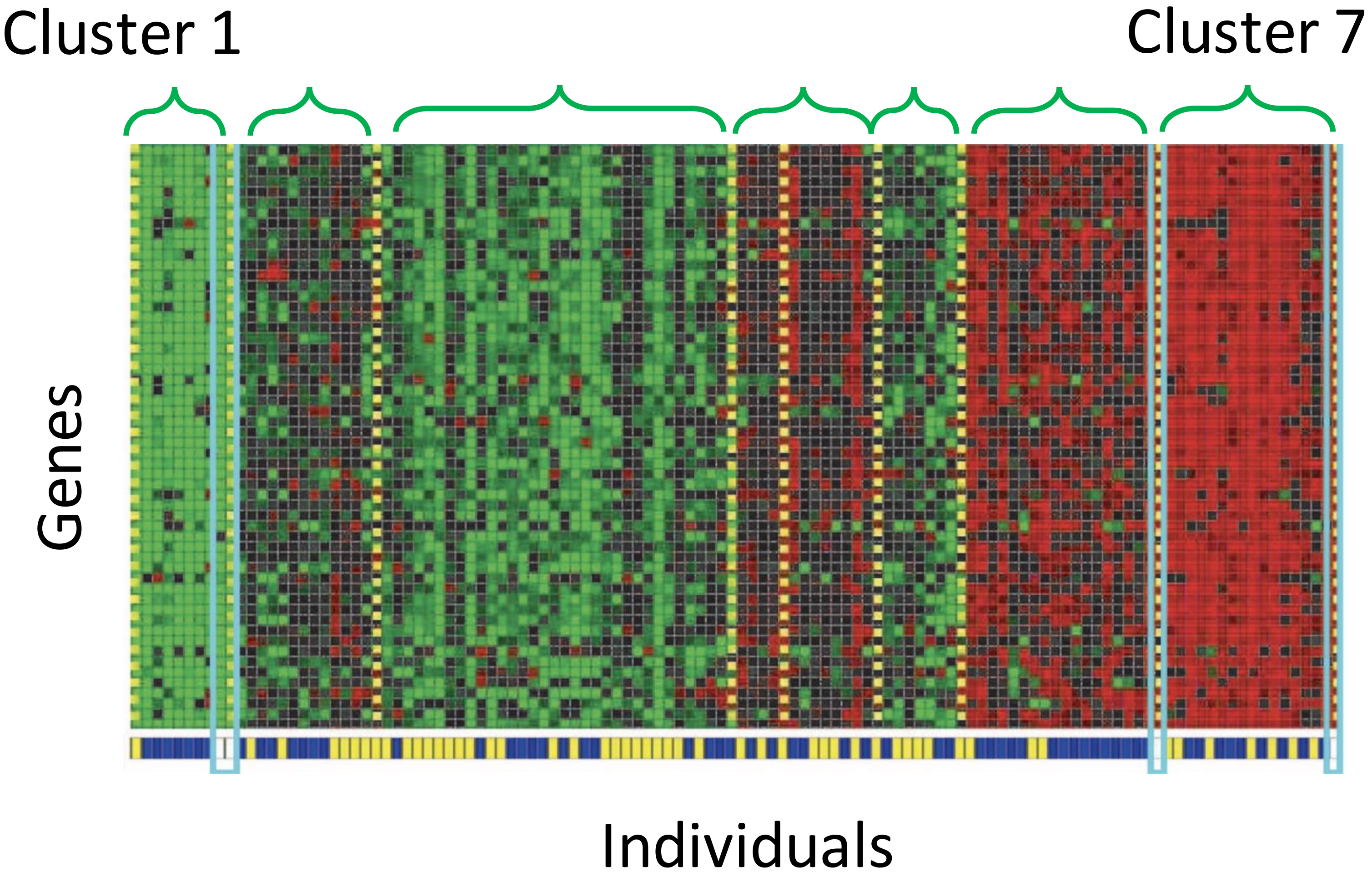


Clustering

k-mean clustering, mixture of Gaussians

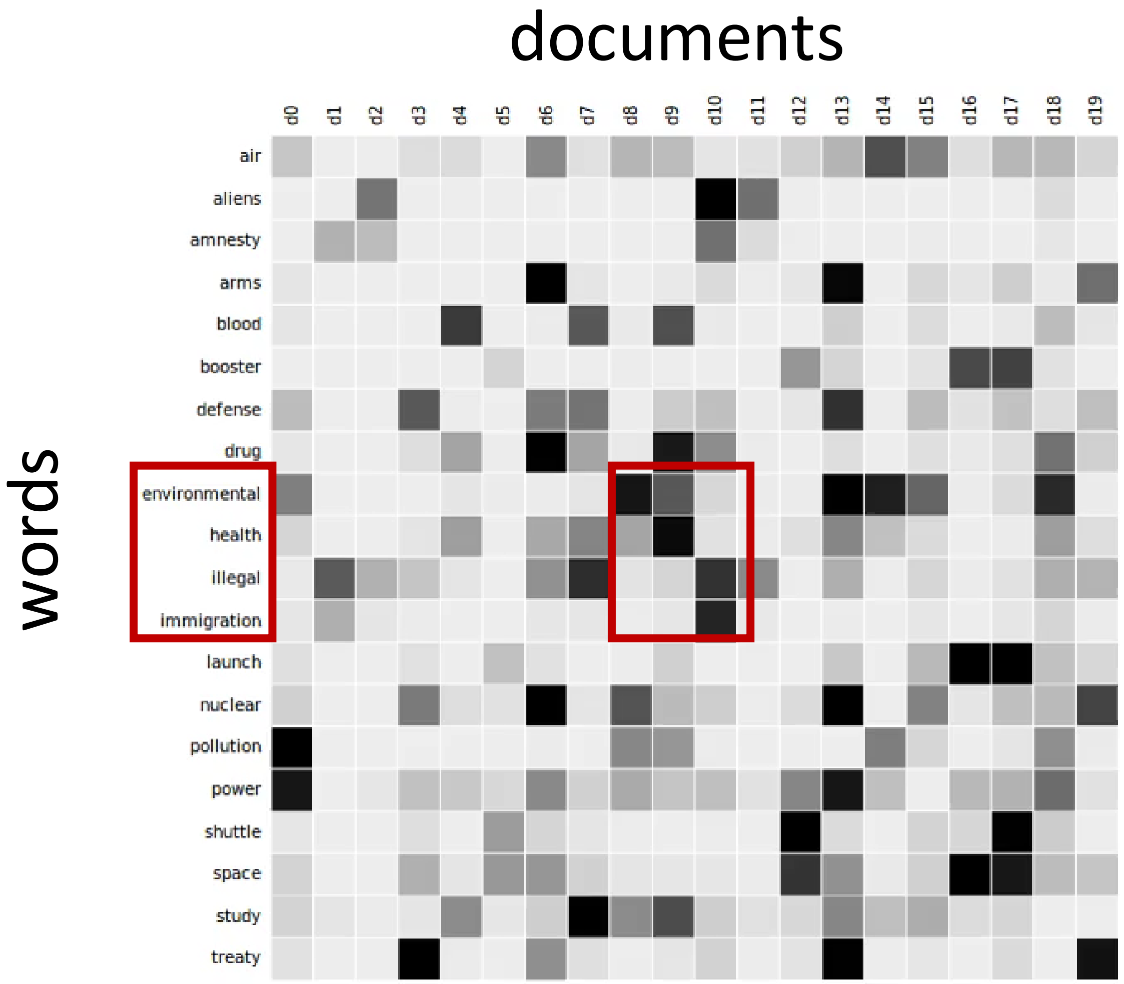


Clustering Genes



Identifying Regulatory Mechanisms using Individual Variation Reveals Key Role for Chromatin Modification. [Su-In Lee, Dana Pe'er, Aimee M. Dudley, George M. Church and Daphne Koller. '06]

Latent Semantic Analysis (LSA)



Principal component analysis (tools used in LSA)

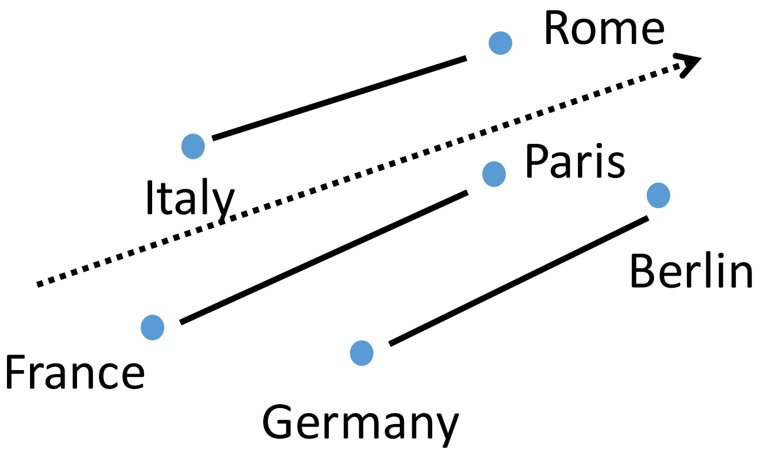
Image credit: https://commons.wikimedia.org/wiki/File:Topic_detection_in_a_document-word_matrix.gif

Word Embeddings

Unlabeled dataset

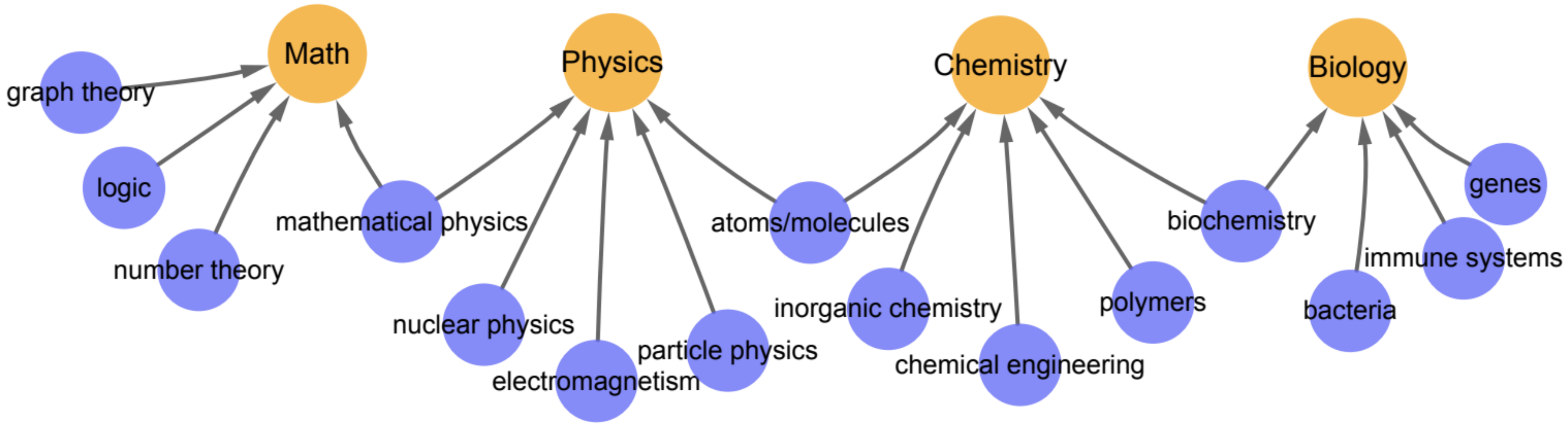
Represent words by vectors

- word $\xrightarrow{\text{encode}}$ vector
- relation $\xrightarrow{\text{encode}}$ direction



Word2vec [Mikolov et al'13]
GloVe [Pennington et al'14]

Clustering Words with Similar Meanings (Hierarchically)

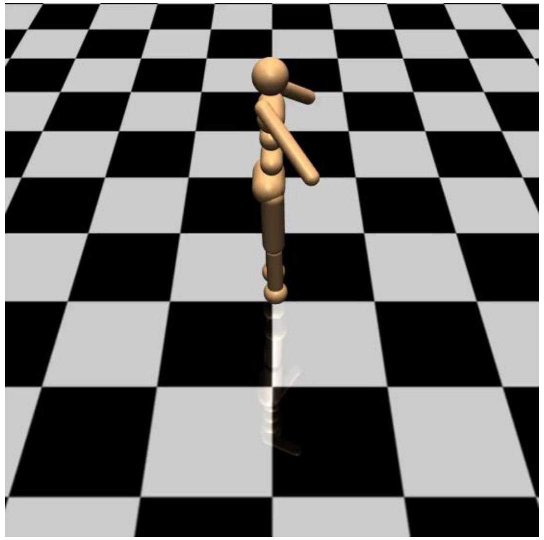


	logic deductive propositional semantics	graph subgraph bipartite vertex	boson massless particle higgs	polyester polypropylene resins epoxy	acids amino biosynthesis peptide
tag	*logic*	*graph theory*	*particle physics*	*polymer*	*biochemistry*

[Arora-Ge-Liang-M.-Risteski, TACL'17,18]

Reinforcement Learning

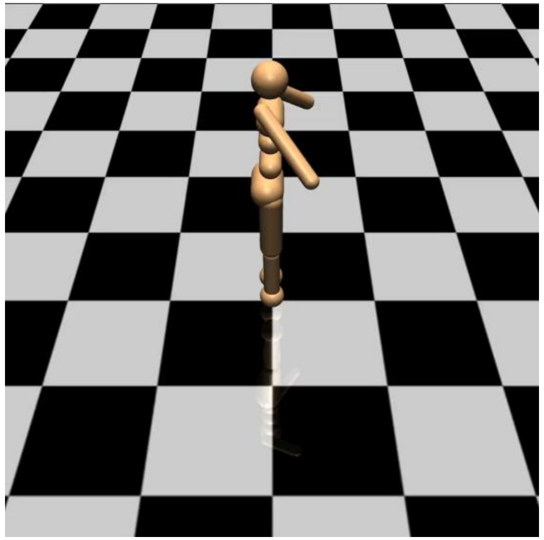
learning to walk to the right



Iteration 10

[Luo-Xu-Li-Tian-Darrell-M.'18]

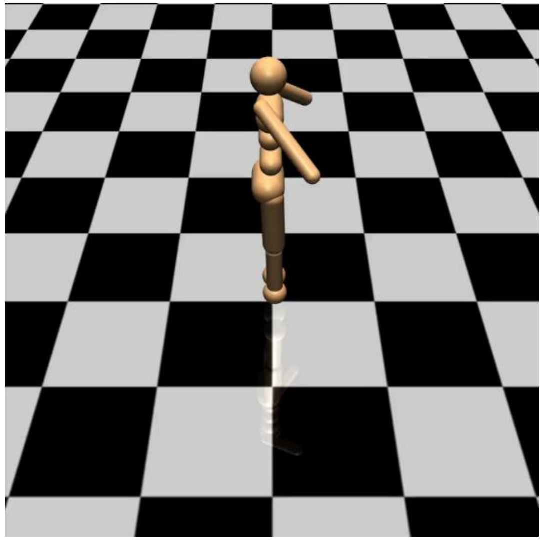
learning to walk to the right



Iteration 20

[Luo-Xu-Li-Tian-Darrell-M.'18]

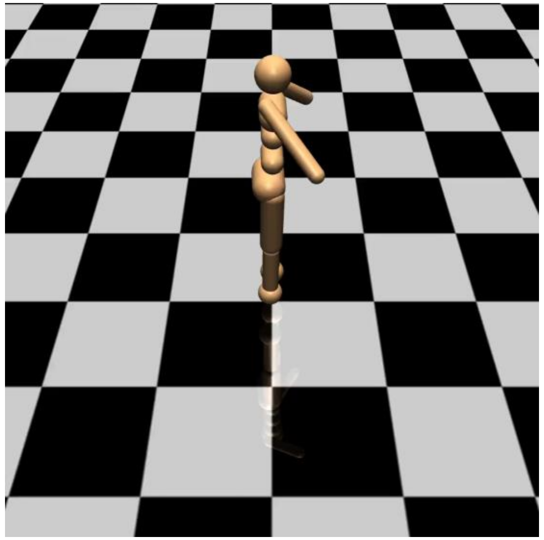
learning to walk to the right



Iteration 80

[Luo-Xu-Li-Tian-Darrell-M.'18]

learning to walk to the right



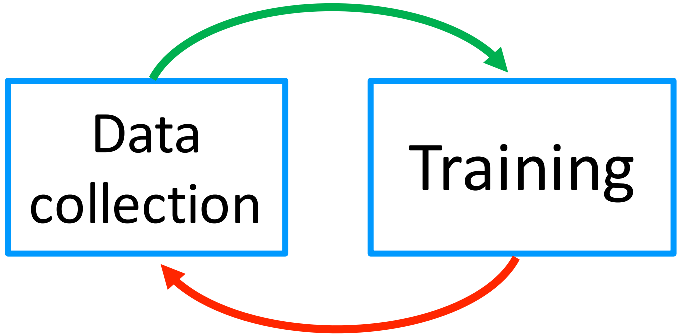
Iteration 210

[Luo-Xu-Li-Tian-Darrell-M.'18]

Reinforcement Learning

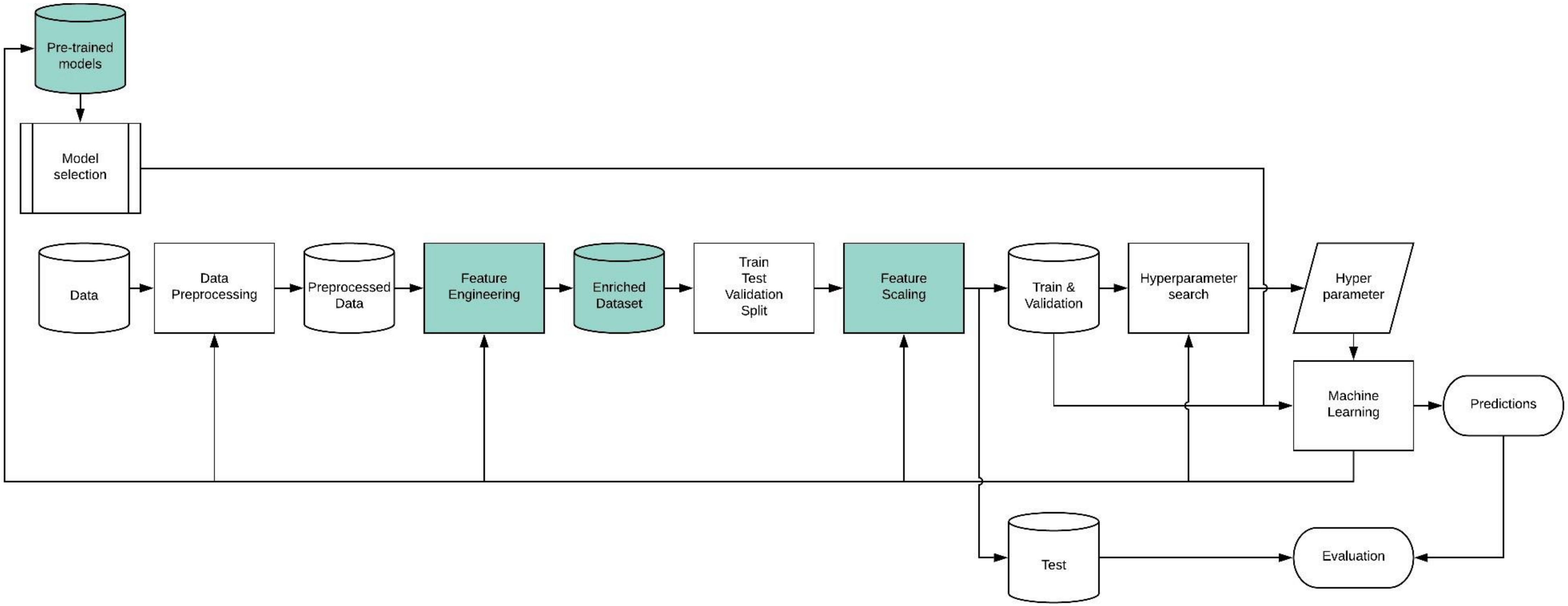
- The algorithm can collect data interactively

Try the strategy and collect feedbacks



Improve the strategy based on the feedbacks

Pipeline Machine Learning Model



Machine Learning as a Process

Define Objectives

- Define measurable and quantifiable goals
- Use this stage to learn about the problem

Data Preparation

- Normalization
- Transformation
- Missing Values
- Outliers

Model Building

- Data Splitting
- Features Engineering
- Estimating Performance
- Evaluation and Model Selection

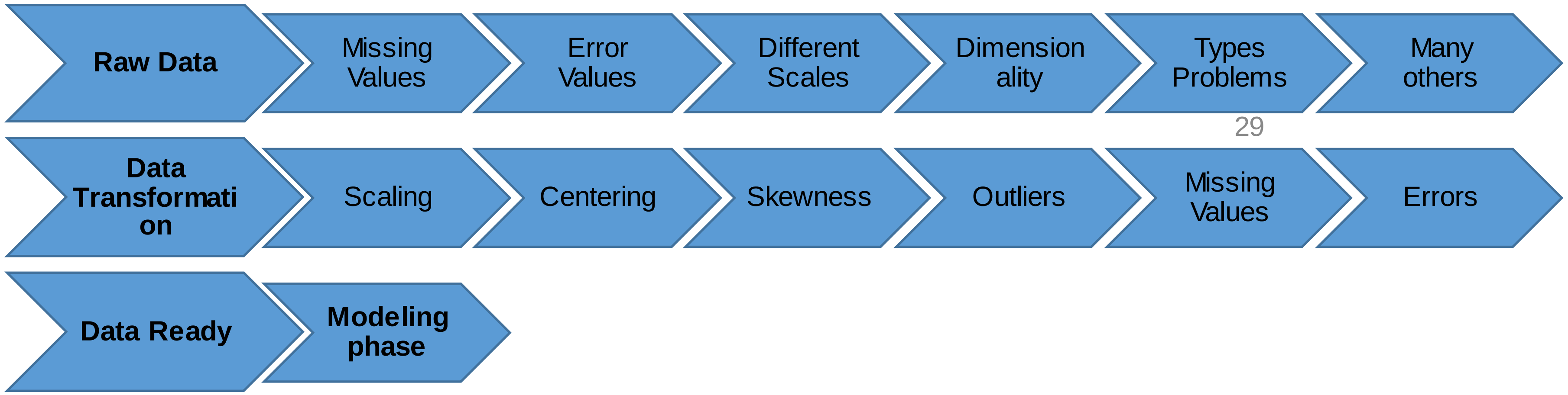
Model Evaluation

- Study models accuracy
  - Work better than the naïve approach or previous system
- Do the results make sense in the context of the problem

Model Deployment

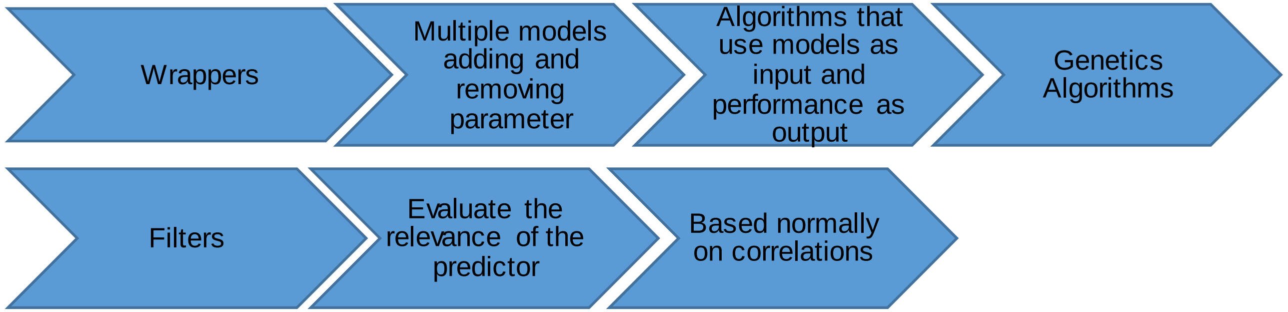
ML as a Process: Data Preparation

- Needed for several reasons
  - Some Models have strict data requirements
    - Scale of the data, data point intervals, etc
  - Some characteristics of the data may impact dramatically on the model performance
- Time on data preparation should not be underestimated



ML as a Process: Feature engineering

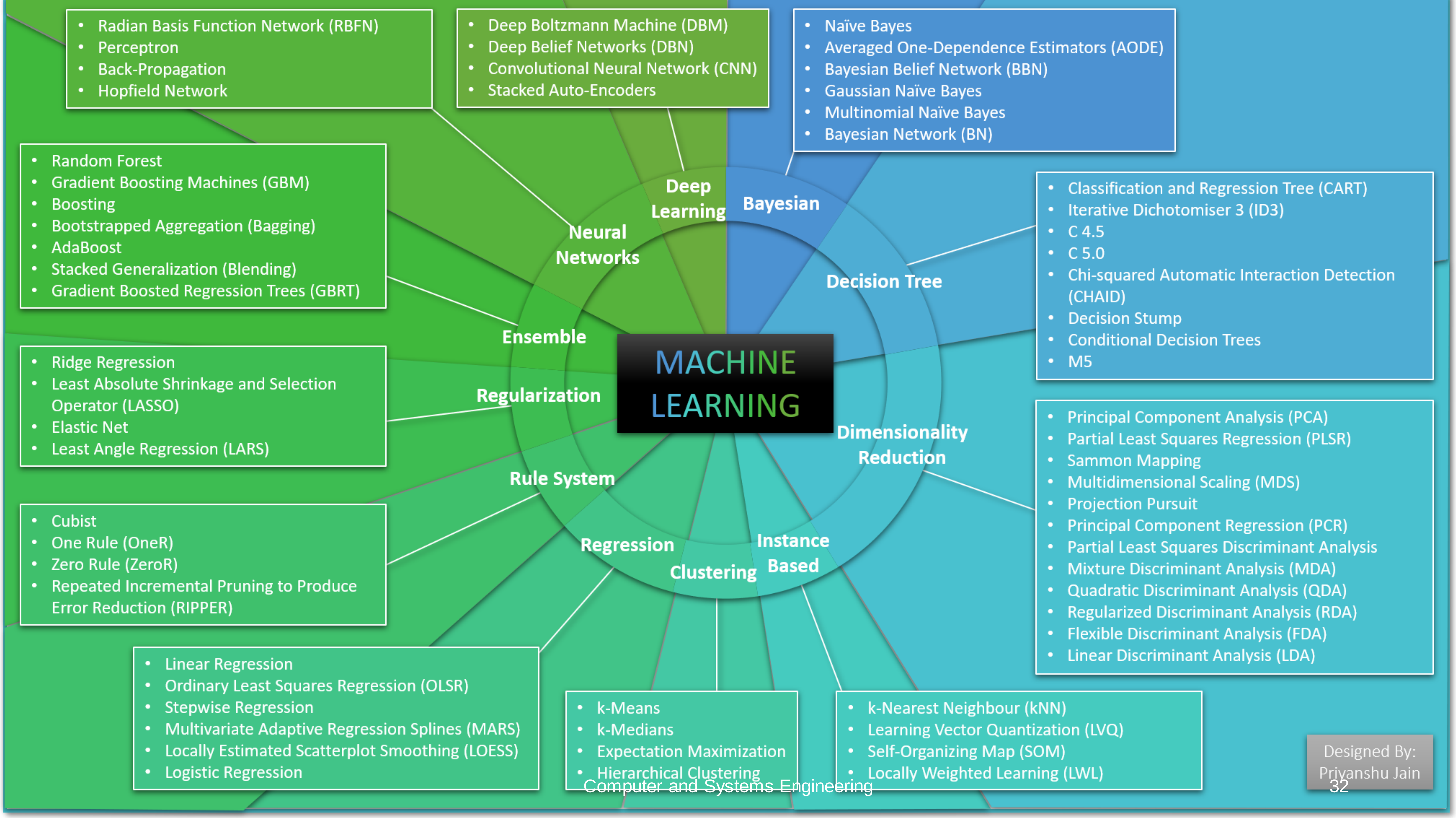
- Determine the predictors (features) to be used is one of the most critical questions
- Sometimes we need to add predictors
- Reduce Number:
  - Fewer predictors more interpretable model and less costly
  - Most of the models are affected by high dimensionality, specially for non-informative predictors



  - Binning predictors

ML as a Process: Model Building

- Data Splitting
  - Allocate data to different tasks
    - model training
    - performance evaluation
  - Define Training, Validation and Test sets
- Feature Selection (Review the decision made previously)
- Estimating Performance
  - Visualization of results – discovery interesting areas of the problem space
  - Statistics and performance measures
- Evaluation and Model selection
  - The ‘no free lunch’ theorem no a priory assumptions can be made
  - Avoid use of favorite models if NEEDED


MACHINE LEARNING
Deep Learning
Bayesian
Neural Networks
Decision Tree
Ensemble
Regularization
Dimensionality Reduction
Rule System
Regression
Clustering
Instance Based
Radian Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Naïve Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naïve Bayes
Multinomial Naïve Bayes
Bayesian Network (BN)
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C 4.5
C 5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR)
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Designed By: Priyanshu Jain