CREWES Data Science
Lecture 8: Regression 1

Marcelo Guarido

www.crewes.org

Today's Agenda

Part 1:	Machine Learning
Part 2:	Linear Regression
Part 3:	SV Regressor
Part 4:	Neural Networks
Part 5:	Live Coding

CREWES DSI

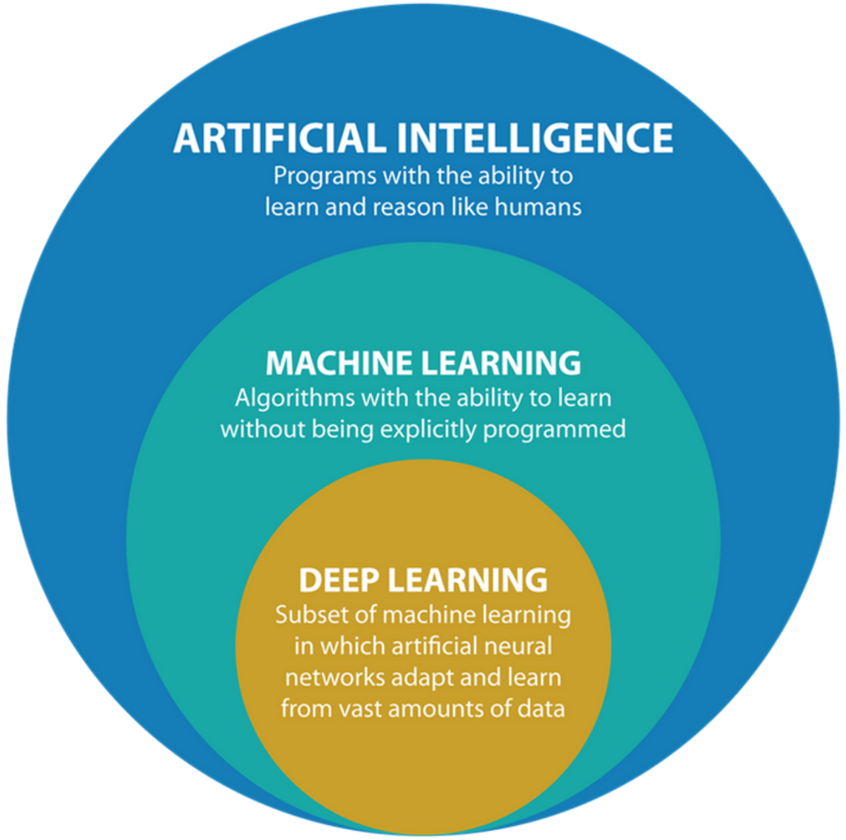
Part 1:

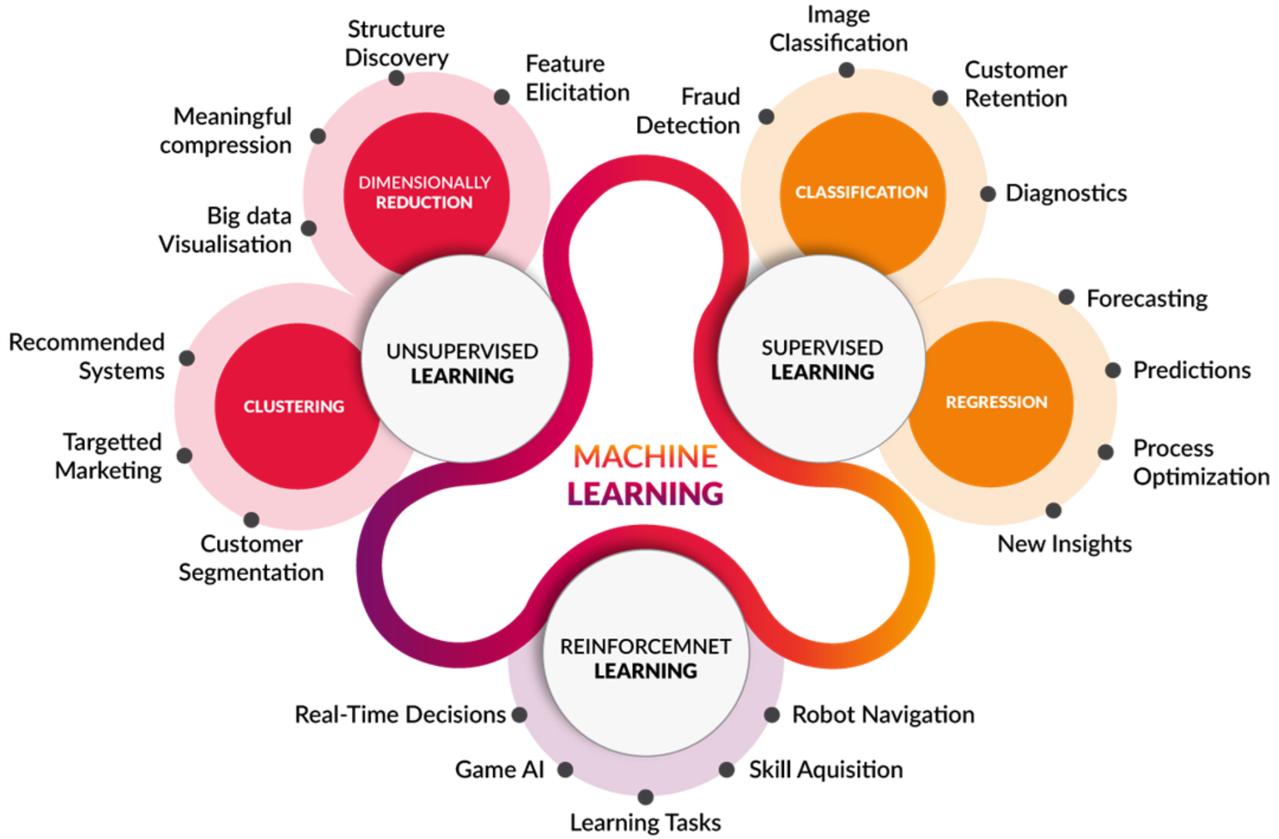
MACHINE LEARNING

Machine Learning

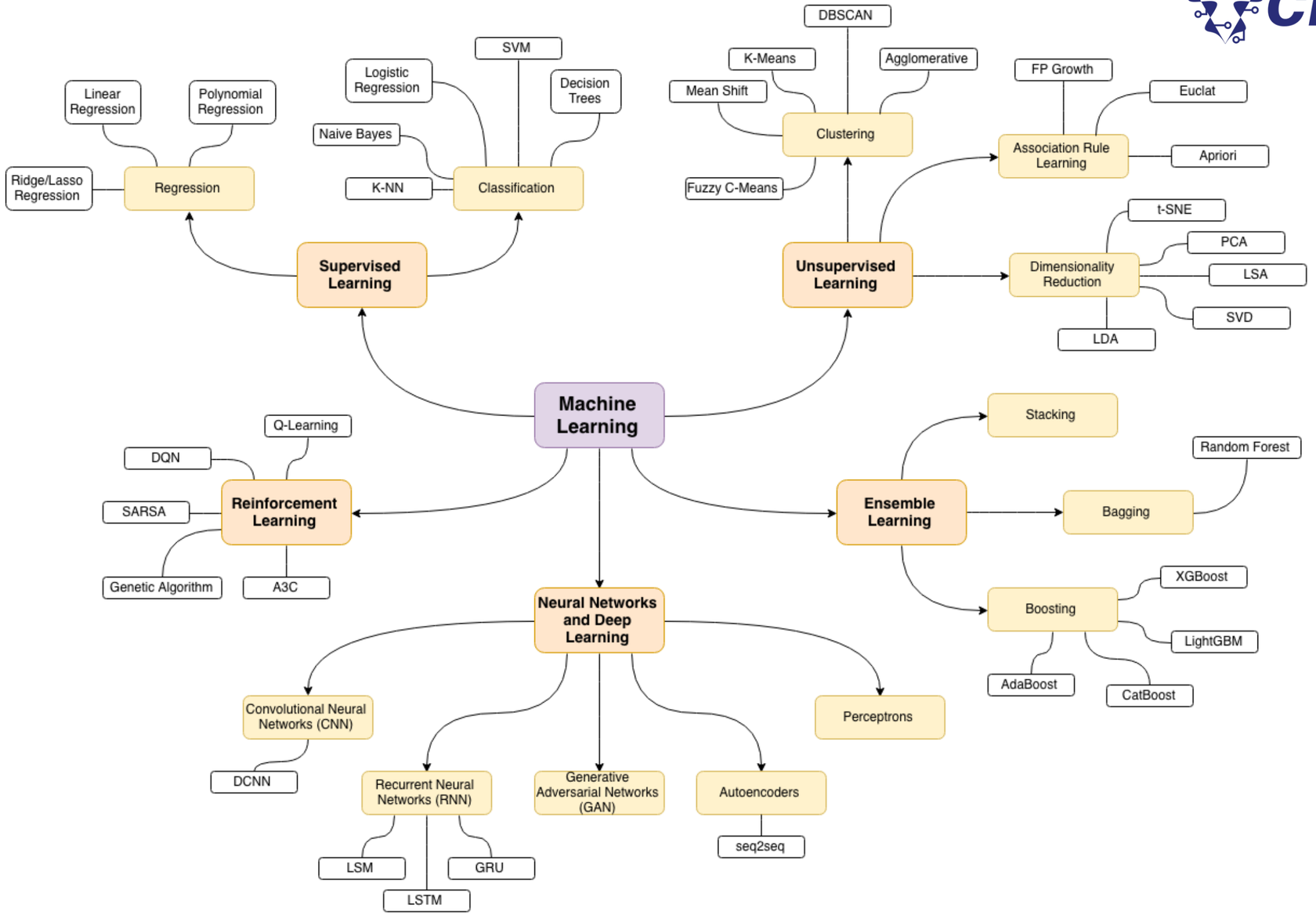
> ***Machine learning*** *is a field of computer science that gives computer systems the ability to "learn" (i.e. progressively improve performance on a specific task) with data, without being explicitly programmed.*

Machine Learning

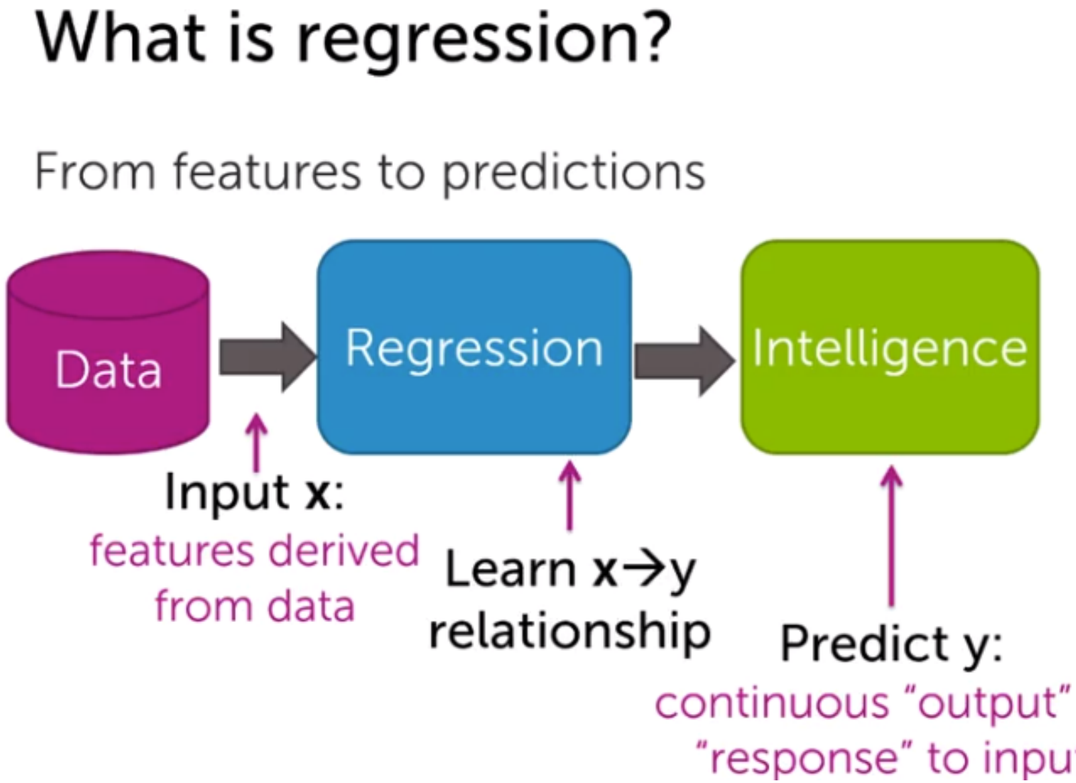




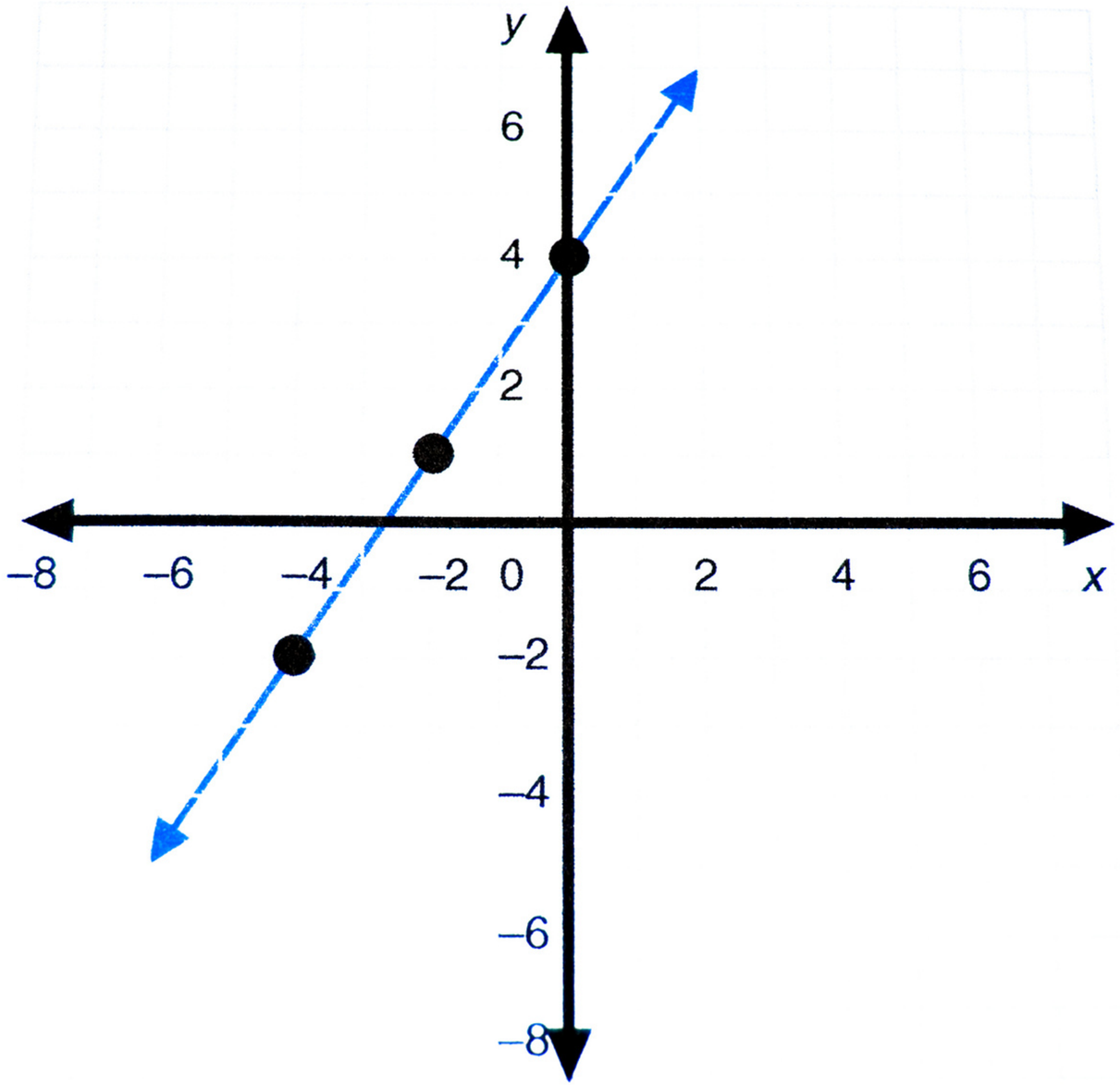
Figures from [Argility]


Machine Learning
Supervised Learning
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
Classification
SVM
Logistic Regression
Decision Trees
Naive Bayes
K-NN
Unsupervised Learning
Clustering
DBSCAN
K-Means
Agglomerative
Mean Shift
Fuzzy C-Means
Association Rule Learning
FP Growth
Euclat
Apriori
Dimensionality Reduction
t-SNE
PCA
LSA
SVD
LDA
Reinforcement Learning
Q-Learning
DQN
SARSA
Genetic Algorithm
A3C
Neural Networks and Deep Learning
Convolutional Neural Networks (CNN)
DCNN
Recurrent Neural Networks (RNN)
LSM
GRU
LSTM
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq
Perceptrons
Ensemble Learning
Stacking
Bagging
Random Forest
Boosting
XGBoost
LightGBM
AdaBoost
CatBoost

Regression



Source

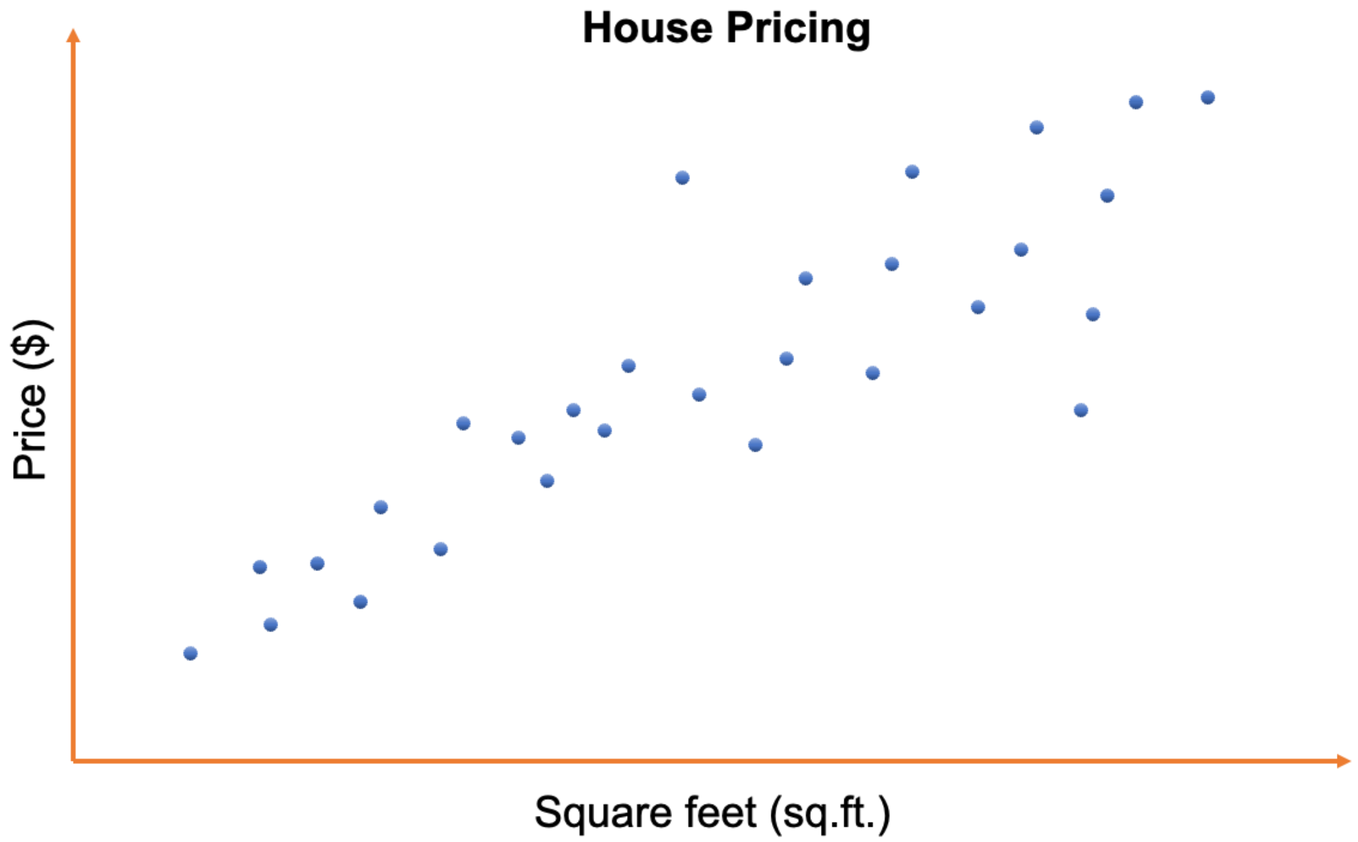


CREWES DSI

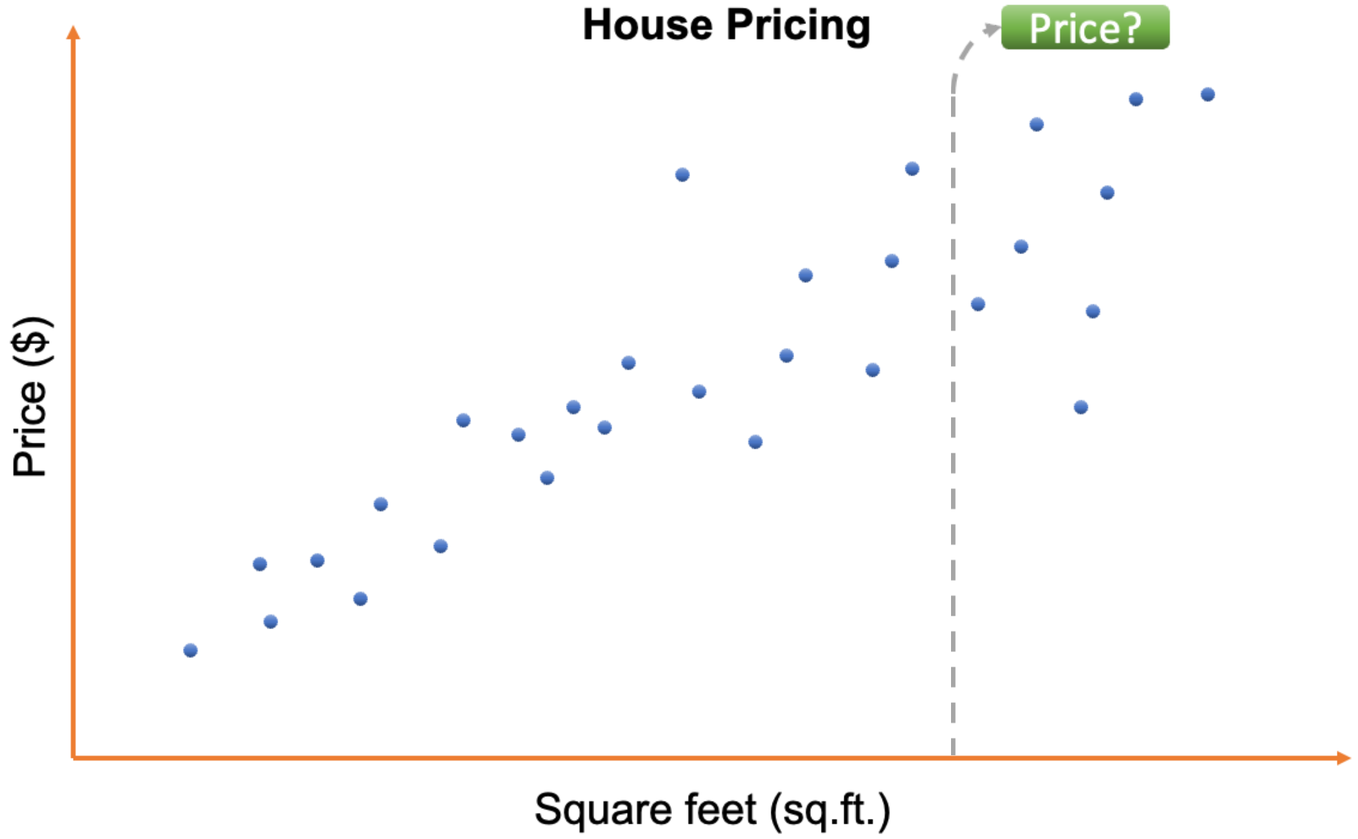
Part 2:

LINEAR REGRESSION

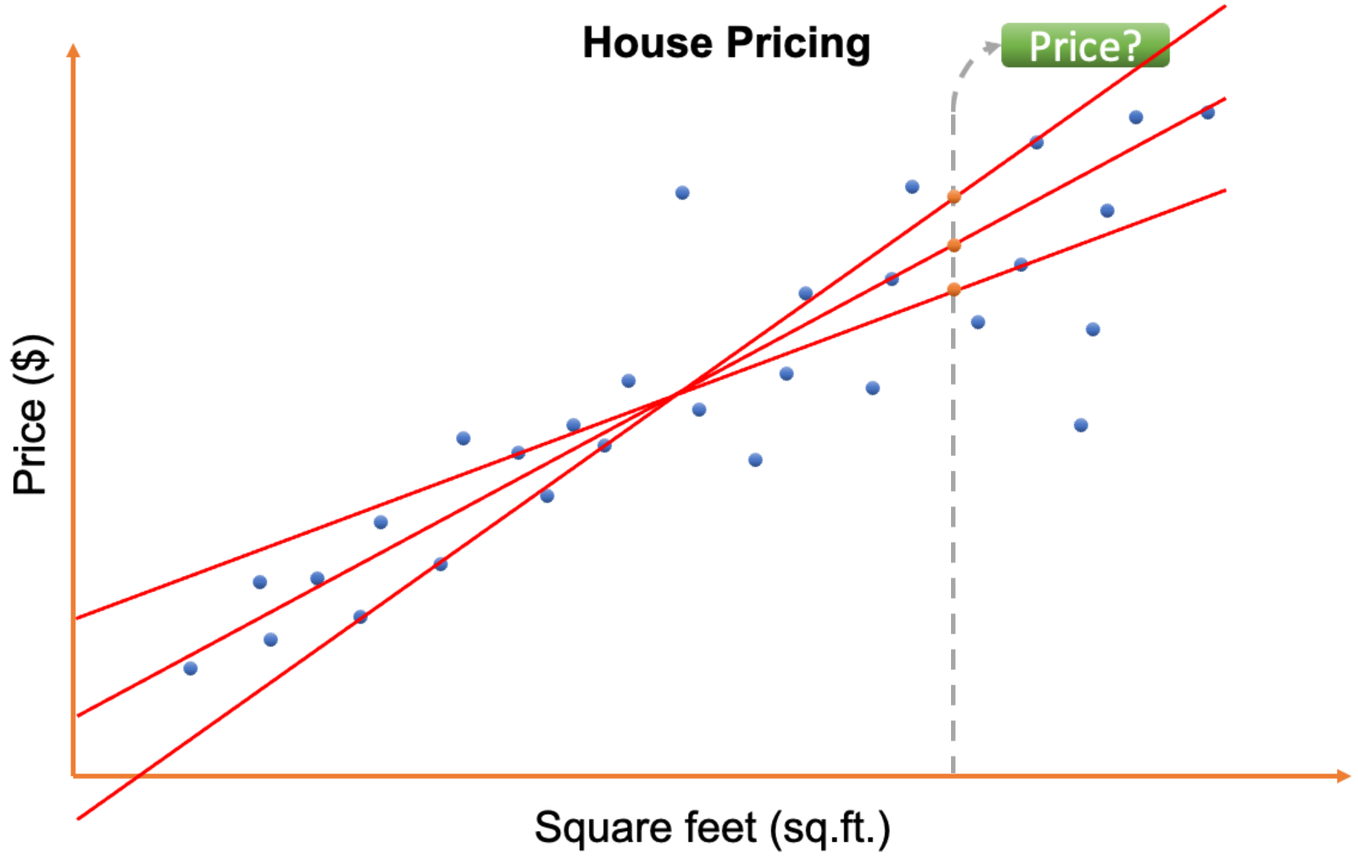
Linear Regression



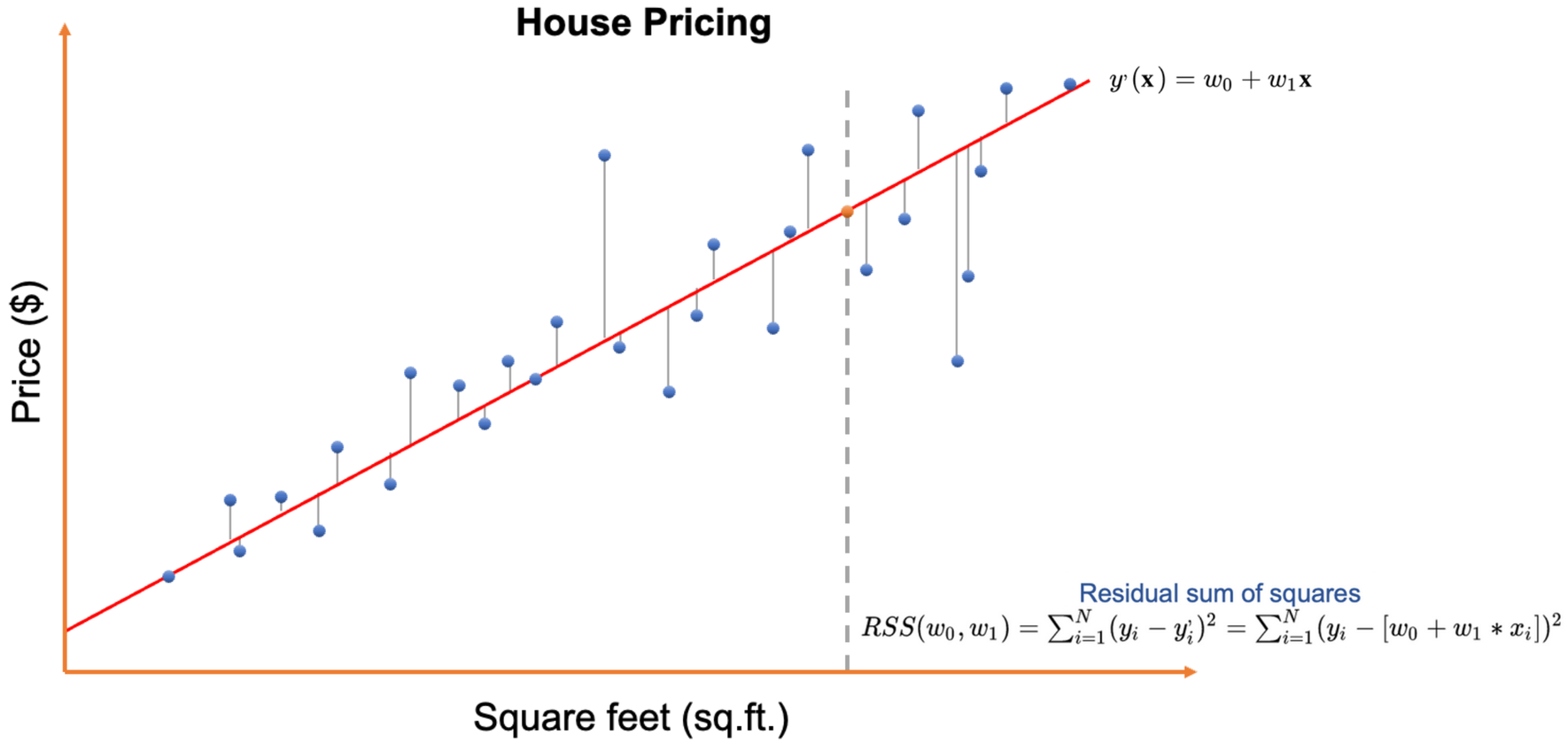
Linear Regression



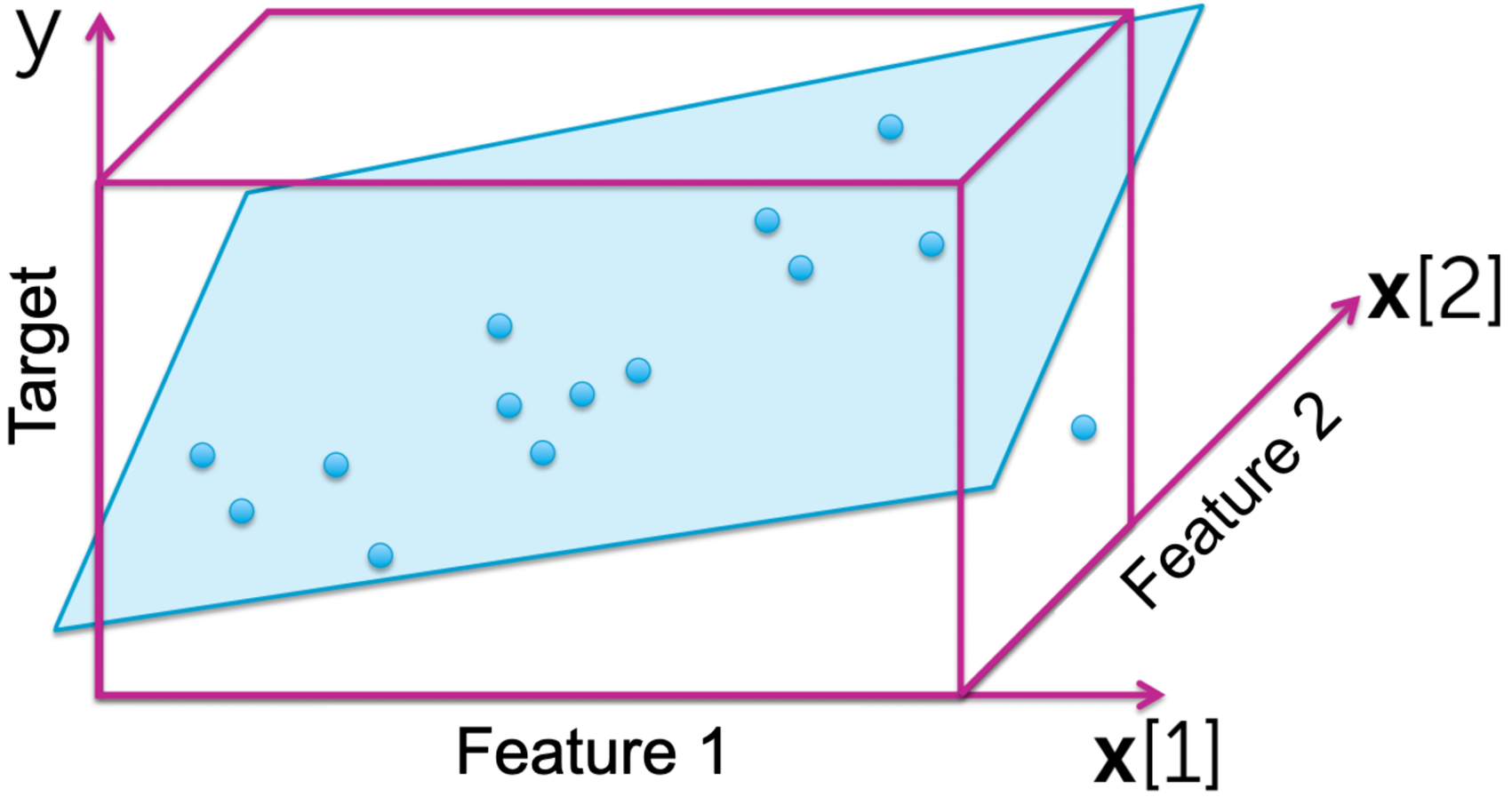
Linear Regression



Linear Regression



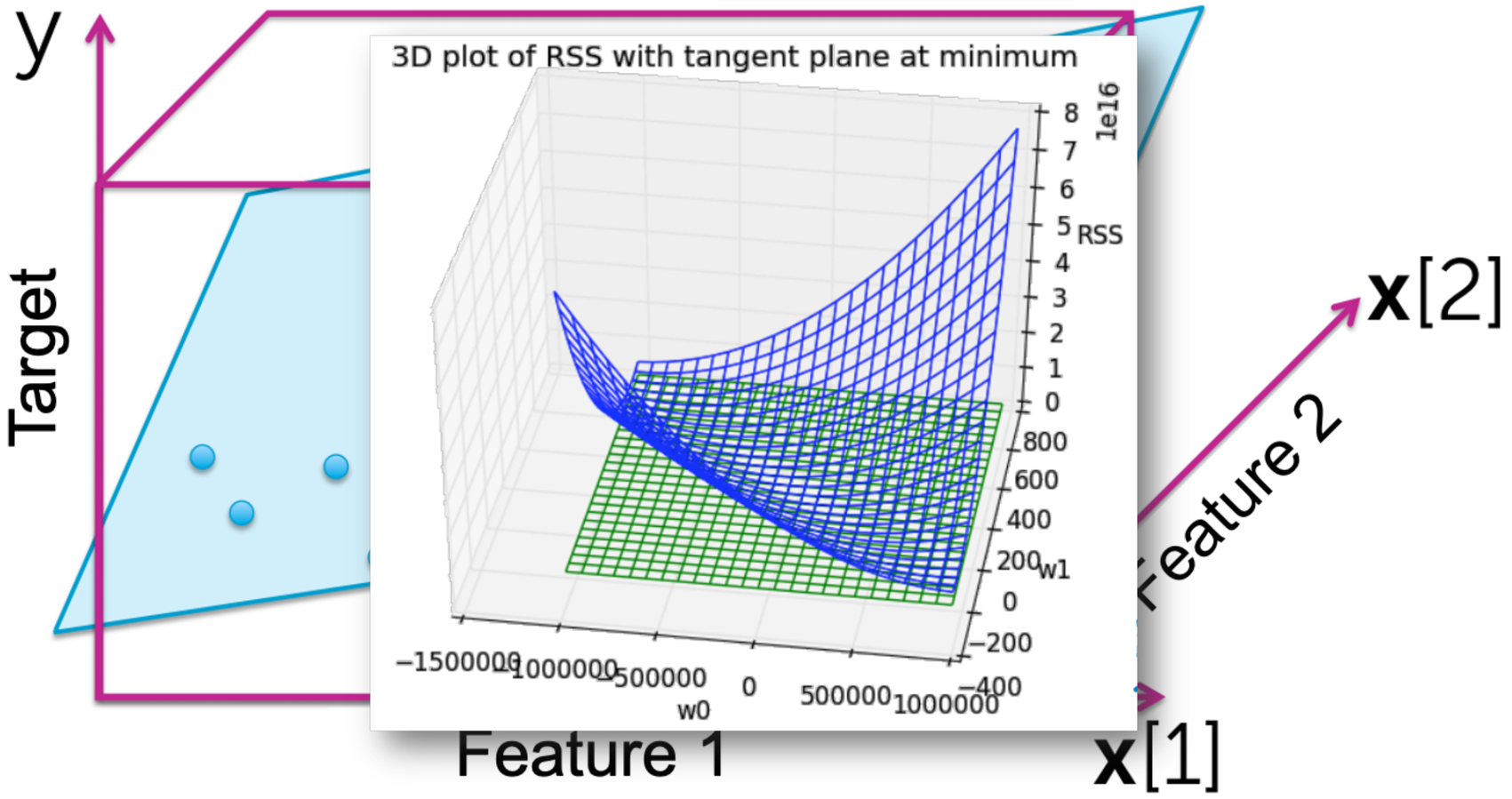
Linear Regression



Linear Regression

Hard to find analytical solution

Gradient Descent



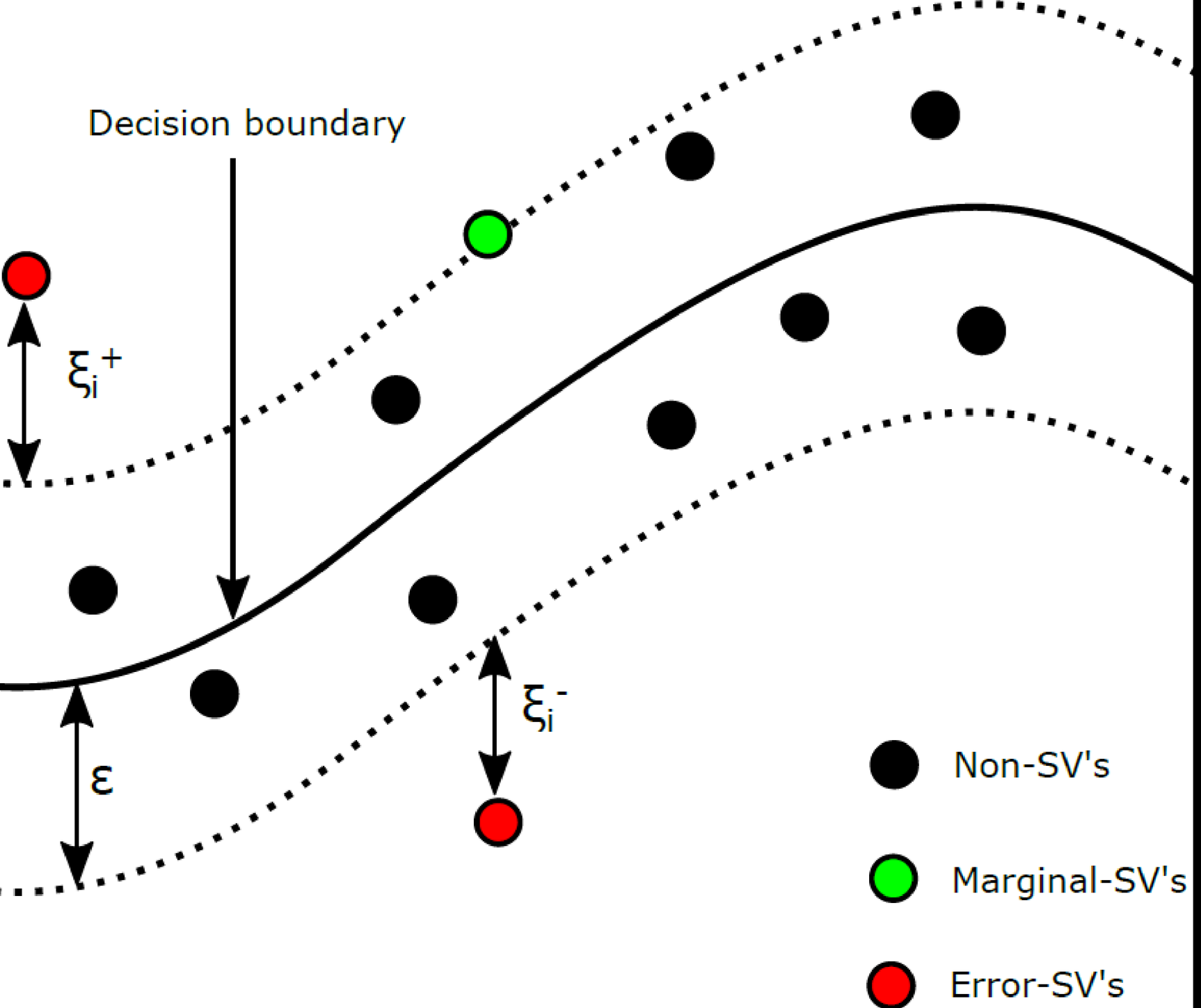
Linear Regression

Pros

- Robust
- Easy to understand and interpret
- Low computing cost

Cons

- Only capture linear correlation
- Affected by outliers
- Requires normalized inputs
- Can't easily handle categorical features



CREWES DSI

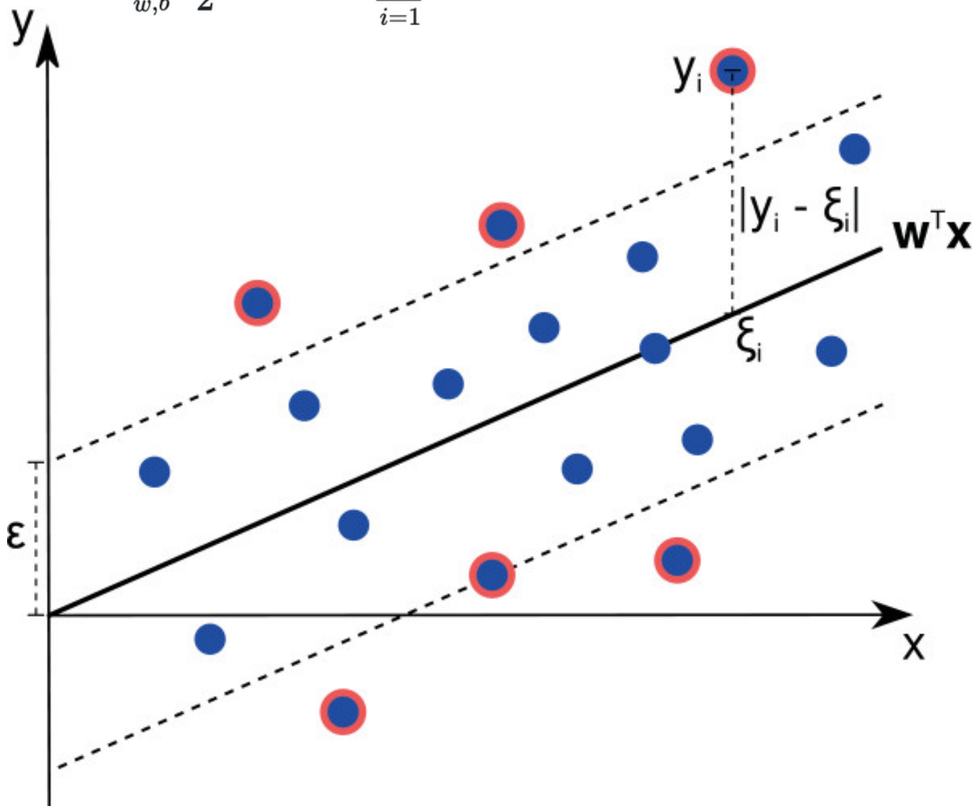
Part 3:

SV REGRESSOR

Suport Vector Regressor (SVR)

- Find a line or curve that better fits the data
- Optimize for more points inside the error interval
- Can use different kernels, such as:
  - Linear: equation of a line
  - Poly: polinomial equation
  - RBF: radial-basis function, also known as "squared exponential"
- Less affected by outliers

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1} \max(0, |y_i - (w^T \phi(x_i) + b)| - \varepsilon)$$



Source

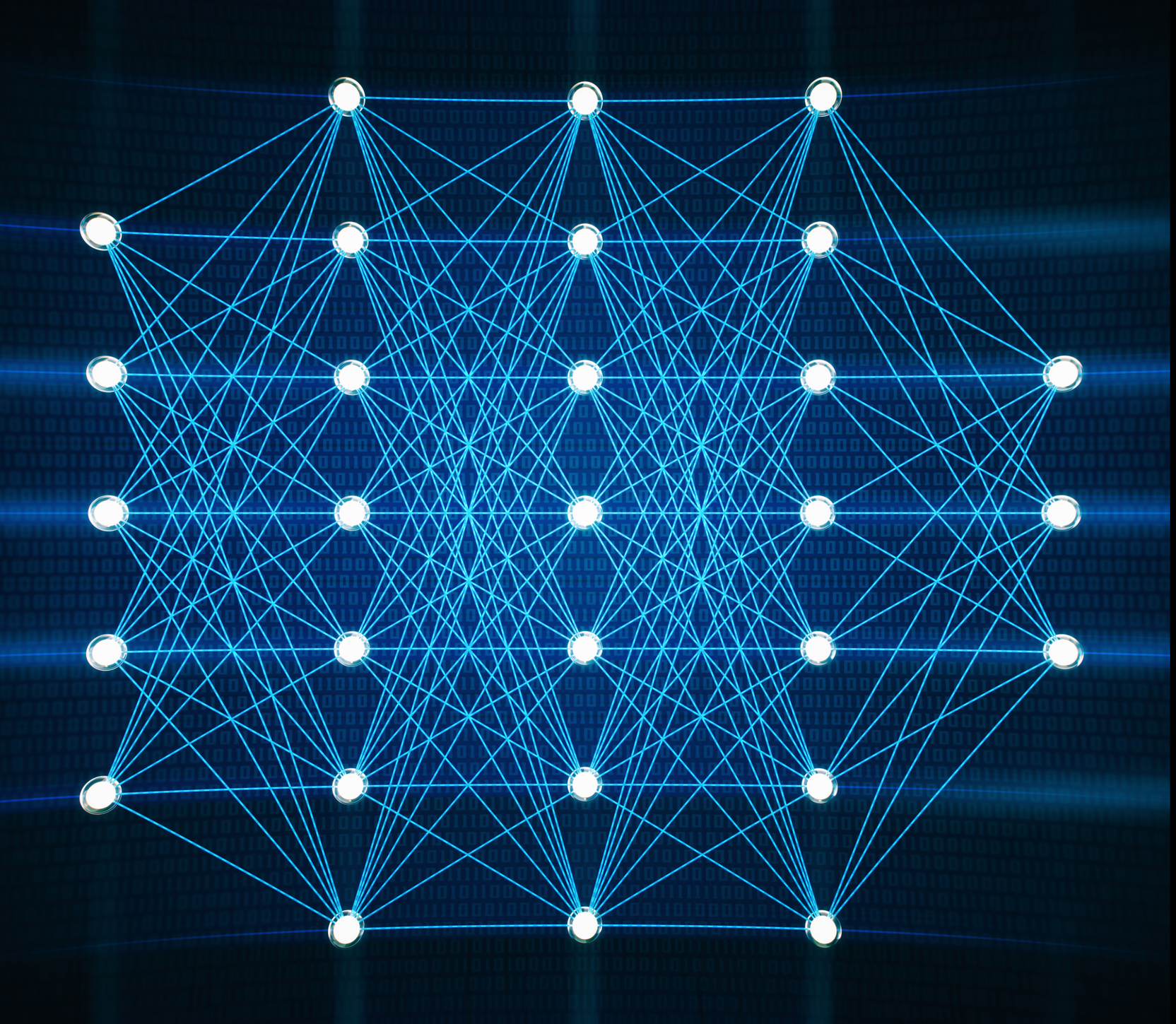
Suport Vector Regressor (SVR)

Pros

- Robust
- Can use different kernels
- Less affected by outliers

Cons

- Hard to parameterize
- Requires normalized inputs
- Can't easily handle categorical features
- High computation cost



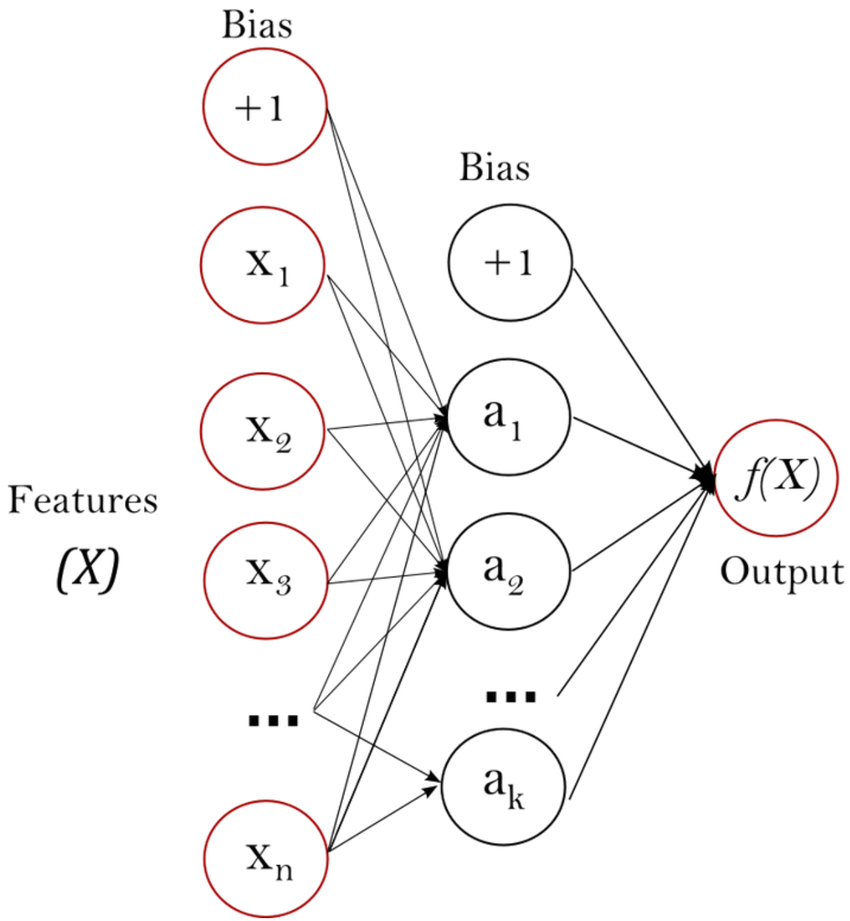
CREWES DSI

Part 4:

NEURAL NETWORKS

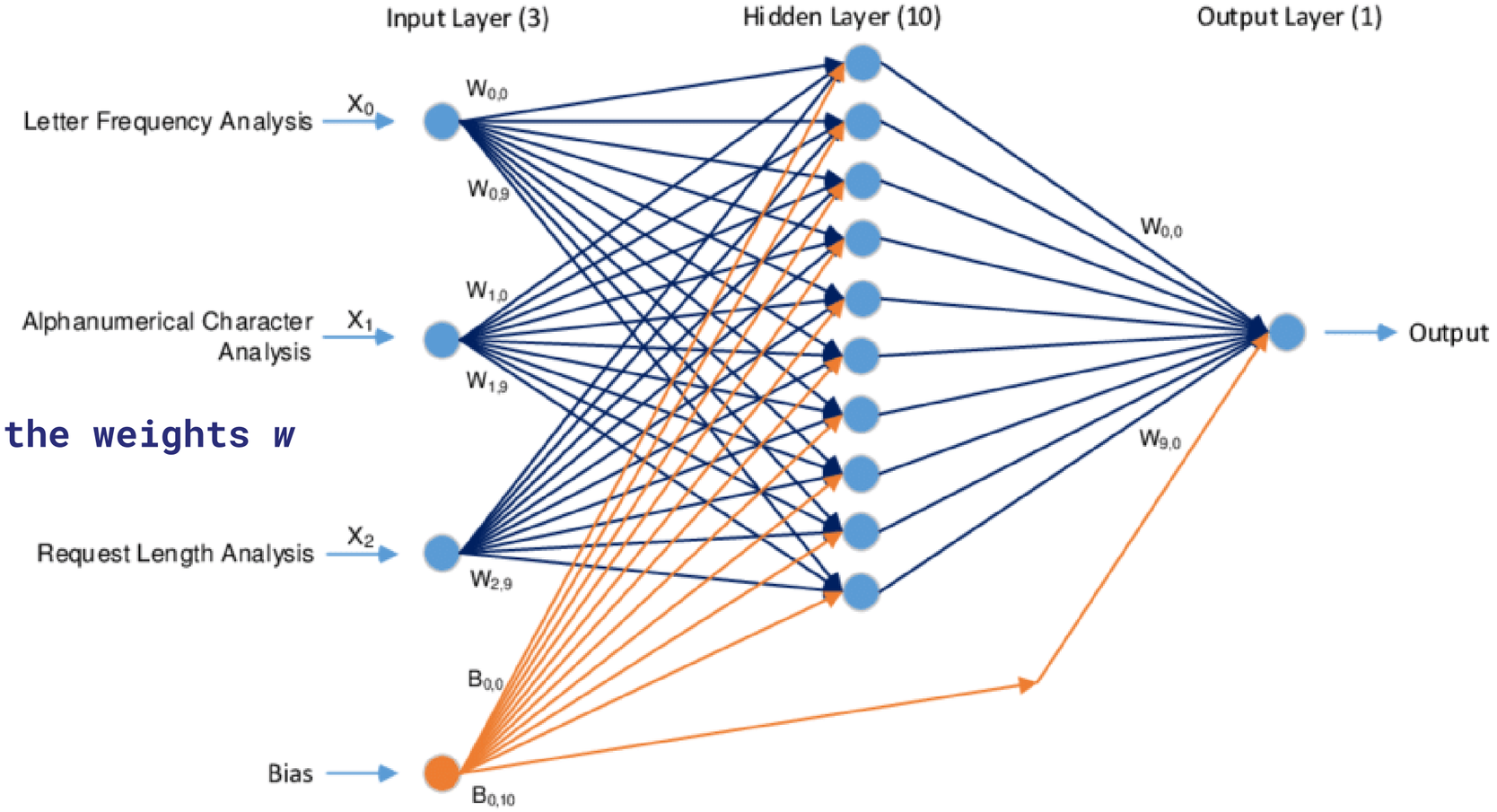
Neural Networks

- Also called "multi-layer perceptron" (MLP)
- Contains different types of layers:
  - *Input layer*: input features
  - *Hidden layers*: internal calculations
  - *Output layer*: predictions
- MLP with 2 or more hidden layers are also called *deep learning*.



Source

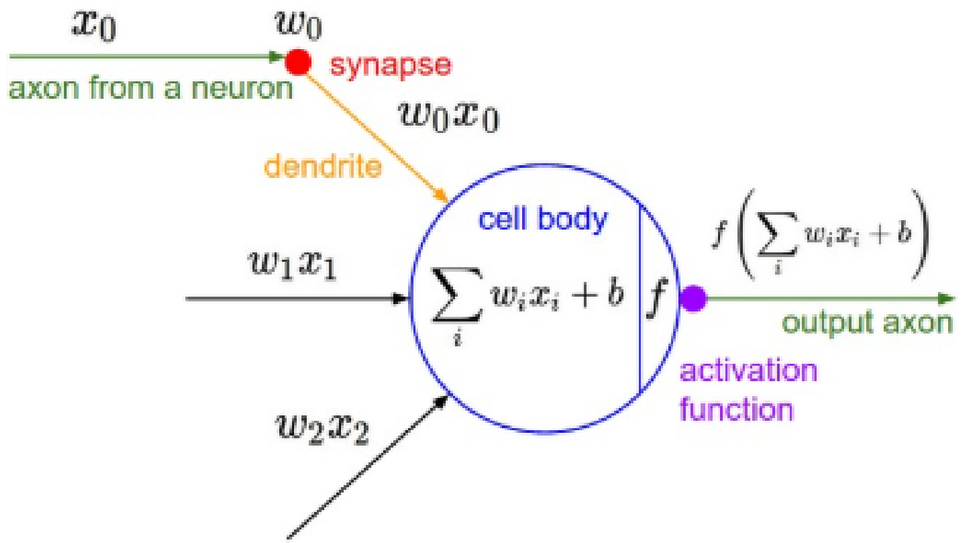
Neural Networks



Solve for the weights *w*

Source

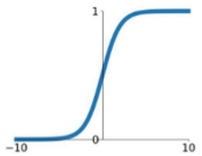
Activation Functions



Source

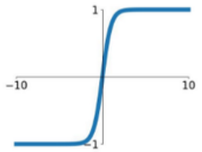
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



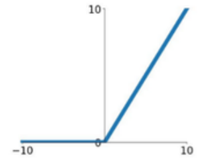
tanh

$\tanh(x)$



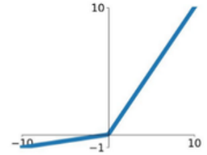
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

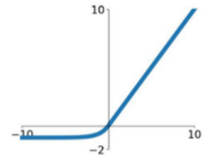


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Source

Activation functions are used to introduce non-linearity to the model

Neural Networks

Pros

- Non-linear
- Flexible
- Works in most scenarios

Cons

- Hard to interpret
- Requires normalized inputs
- Can't easily handle categorical features
- Requires a large amount of data
- High computation cost

CREWES DSI

Part 5:

LIVE CODING (PYTHON)