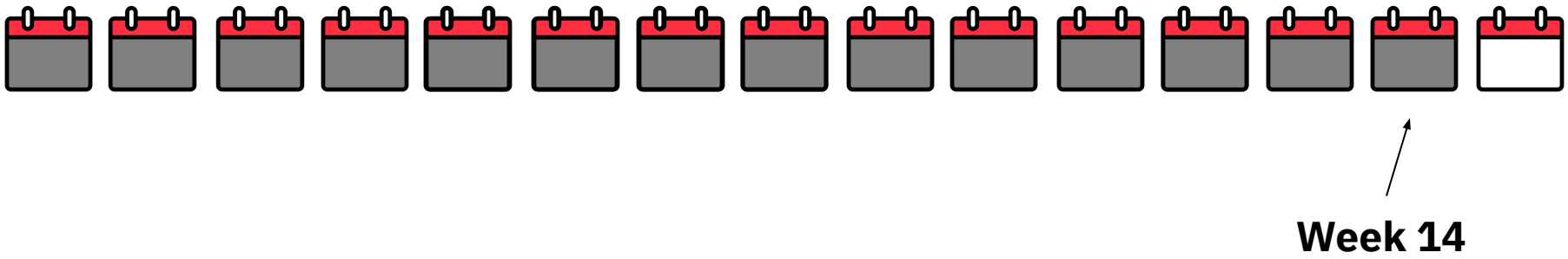
Supervised Learning Workshop

22 April 2019


Week 14

This Week in Business Technology

“no u”

— qualcomm to apple

Netflix says it's testing a shuffle feature for when you don't know what to watch

Sarah Perez @sarahintampa / 2 days ago

Comment

Netflix is testing a new feature that can help you start streaming when you don't know what to watch. The company confirmed it's testing a shuffle mode of sorts, which will allow you to easily click on a popular show to start playing a random episode. The idea with the feature is to offer an experience that's more like traditional TV — where you could just turn the set on, and there would be something to watch.

With today's streaming services, that sort of seamless experience is more difficult to achieve. Instead, viewers now have to first select a streaming app, then scroll through endless menus and recommendations before they can settle on their next title.

The new shuffle feature, instead, offers something closer to the experience of turning on cable TV, when there was always some classic favorite show playing in syndication.

Samsung Galaxy Fold: Multiple Reports Emerge That Review Units Are Breaking

"The only thing they can do now is significantly delay the product."

By Danny Paez on April 18, 2019 Filed Under Smartphones

Might be time to take this one back to the drawing board: Review units for the recently announced $1,980 Galaxy Fold, which shipped to some tech reviewers on Monday, are already breaking. Almost at once, Twitter was flooded

jack
@jack

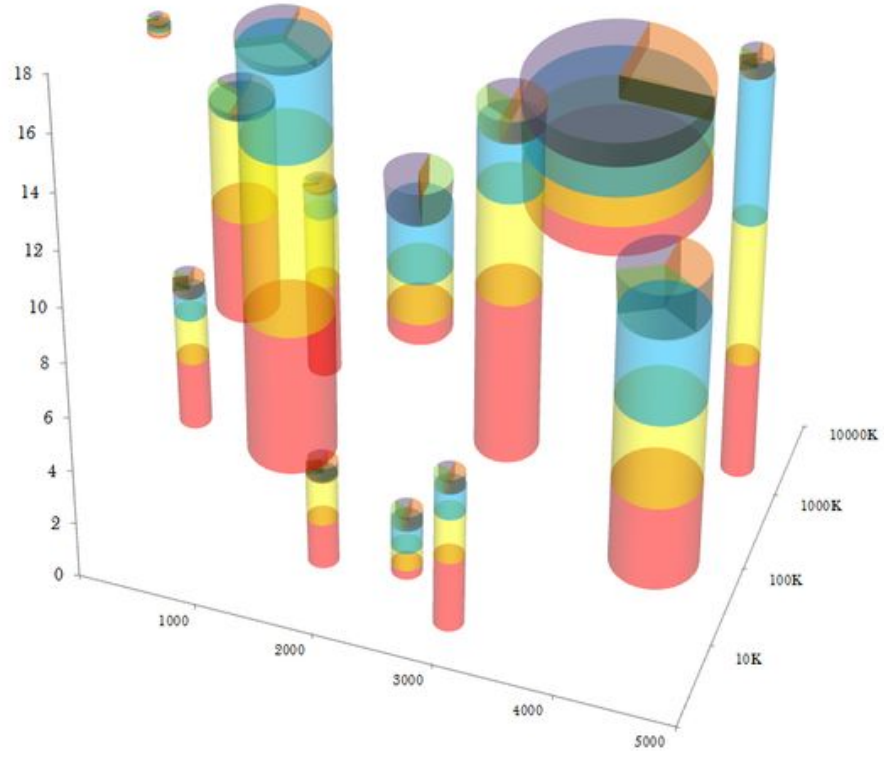
Follow

#BitcoinTwitter and #CryptoTwitter! Square is hiring 3-4 crypto engineers and 1 designer to work full-time on open source contributions to the bitcoin/crypto ecosystem. Work from anywhere, report directly to me, and we can even pay you in bitcoin! Introducing @SqCrypto. Why?

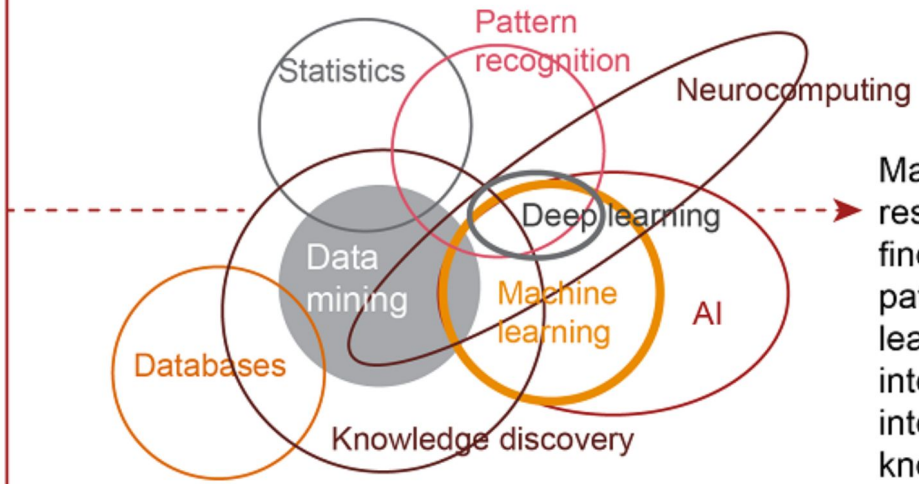
2:58 PM - 20 Mar 2019

4,276 Retweets **12,483** Likes

Supervised Learning


18
16
14
12
10
8
6
4
2
0
1000
2000
3000
4000
5000
10000K
1000K
100K
10K

How does machine learning relate to artificial intelligence?



Machine learning is a category of research and algorithms focused on finding patterns in data and using thos patterns to make predictions. Machine learning falls within the artificial intelligence (AI) umbrella, which in tur intersects with the broader field of knowledge discovery and data mining

Source: SAS, 2014 and PwC, 2016

Supervised learning: a history


the rematch
GARRY
KASPAROV
DEEP
BLUE

Timeline of machine learning

From Wikipedia, the free encyclopedia

This page is a **timeline of machine learning**. Major discoveries, achievements, milestones and other major events are included.

Contents [hide]

Supervised learning: an overview

- Generates a function to map inputs with outputs based on example input-output pairs
- These example input-output pairs are contained within a **training set**
- Ideally, we can train an function to be able to infer an output based on an unfamiliar input (pattern recognition)
- The accuracy of the model can be tested using a set of data distinct from the training set, which is termed a **test set**
- *What do we need to keep in mind?*

Supervised learning

From Wikipedia, the free encyclopedia

See also: Unsupervised learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.[1] It infers a function from *labeled training data* consisting of a set of *training examples*.[2] In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value (also called the *supervisory signal*). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way (see inductive bias).

The parallel task in human and animal psychology is often referred to as concept learning.

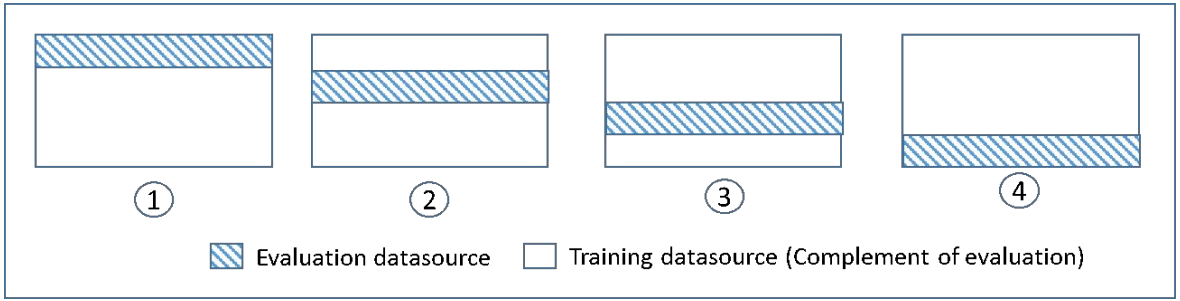
Things to keep in mind

- Overfitting
  - Functions need to be accurate, but how accurate?
- Dimensionality
  - Datasets can contain thousands of variables of data, but how many do we realistically need?
- Bad data
  - Redundant variables are particularly dangerous
  - Different types of variables (categorical, continuous, discrete, etc)
  - Not enough data (see **cross-validation**)

Cross-validation

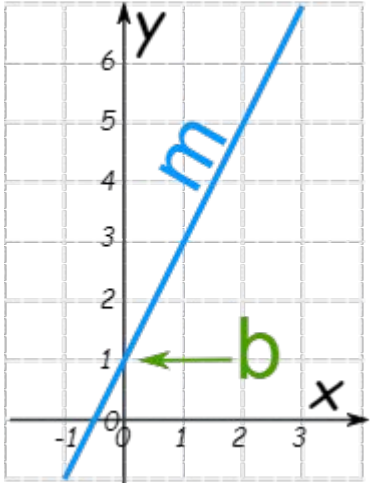
In k-fold cross-validation, you split the input data into k subsets of data (also known as folds).

- You train an ML model on all but one (k-1) of the subsets, and then evaluate the model on the subset that was not used for training.
- This process is repeated k times, with a different subset reserved for evaluation (and excluded from training) each time.



$$y = mx + b$$

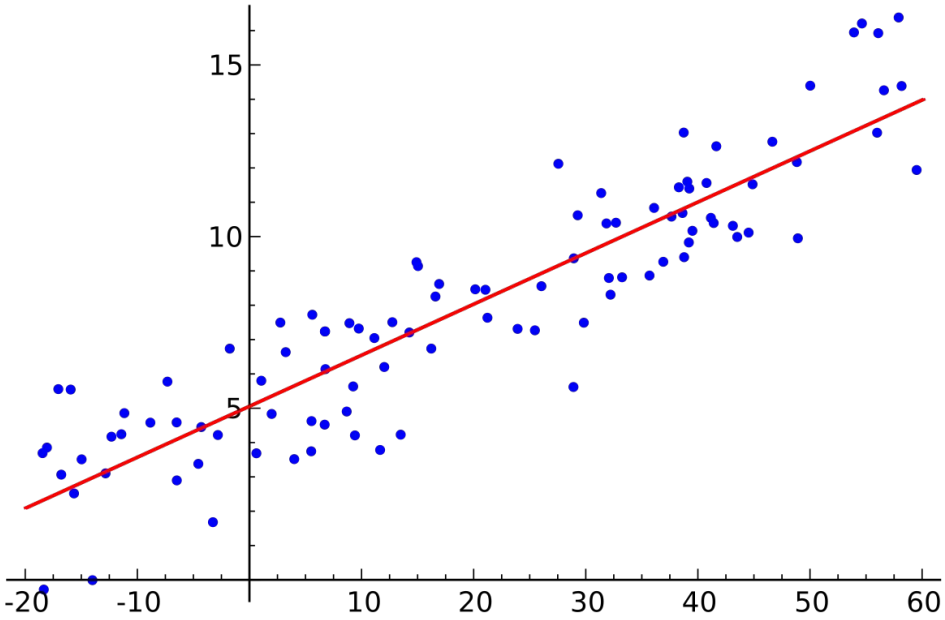
$y = mx + b$



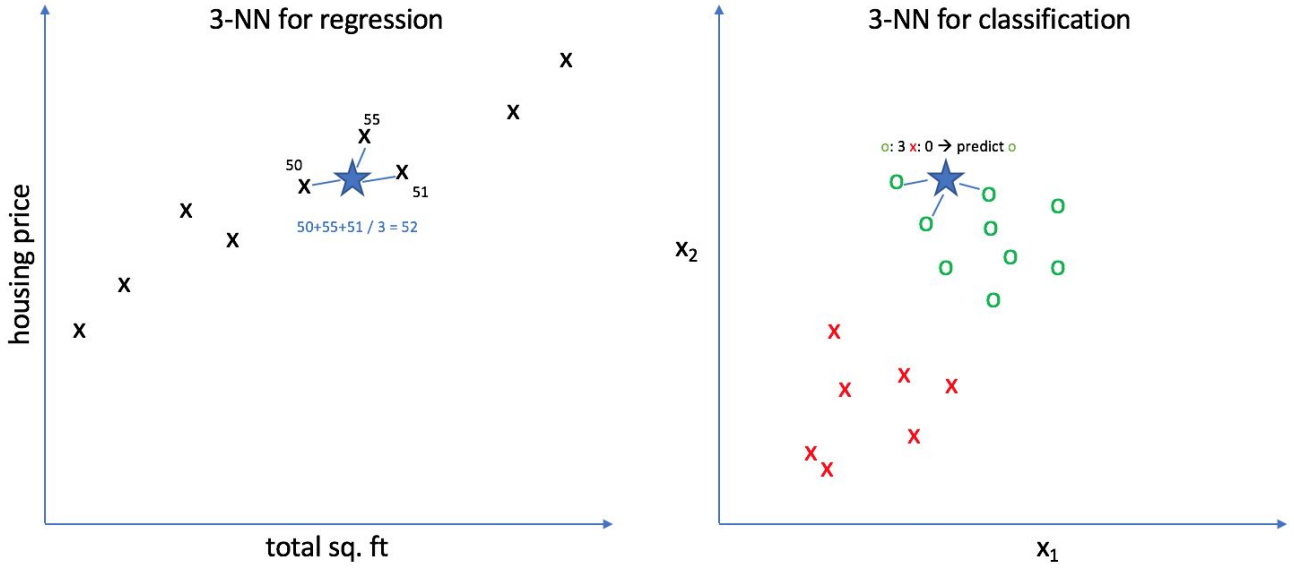
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \ldots, n$$

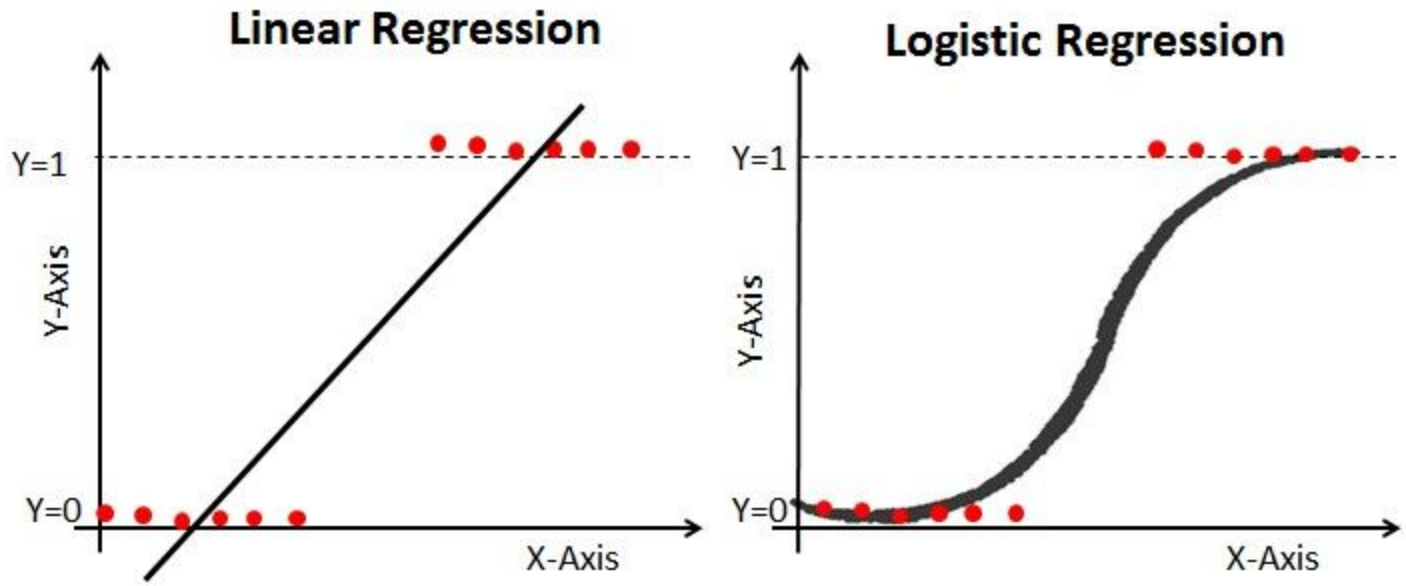
$$Q_{\mathrm{LS}}(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2$$

Linear Regression

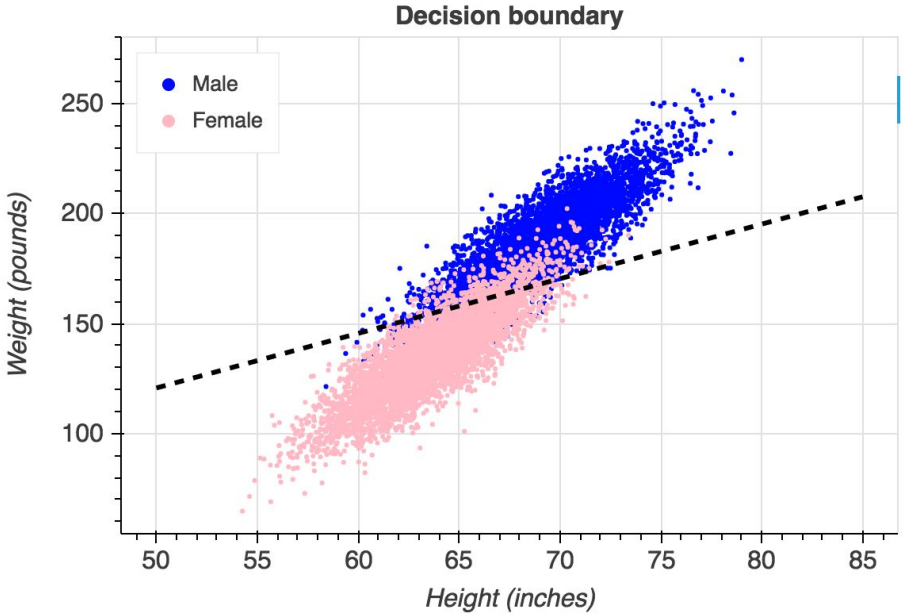


K-Nearest Neighbors

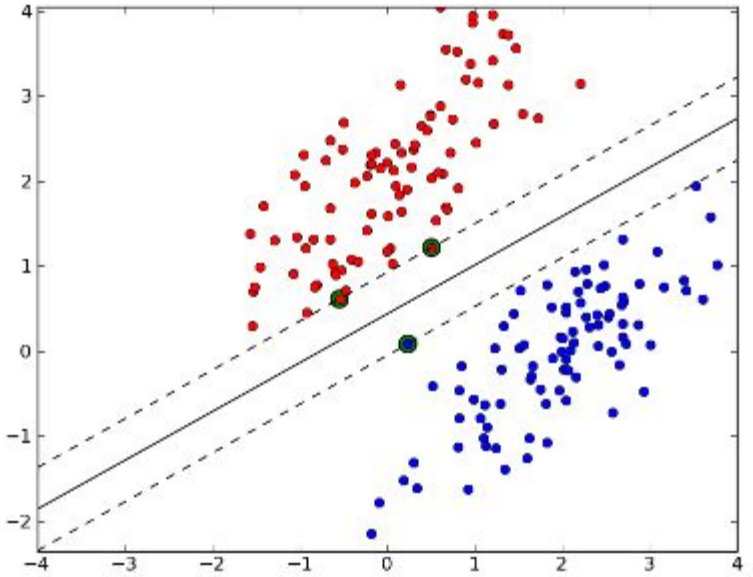


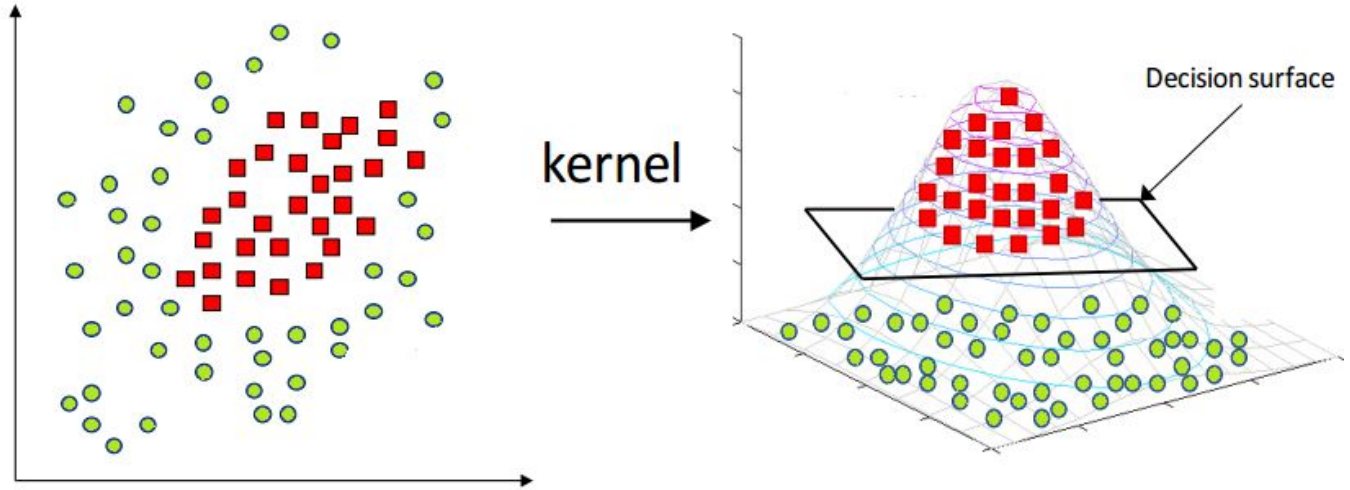

Linear Regression
Y=1
Y-Axis
Y=0
X-Axis
Logistic Regression
Y=1
Y-Axis
Y=0
X-Axis

Logistic Regression

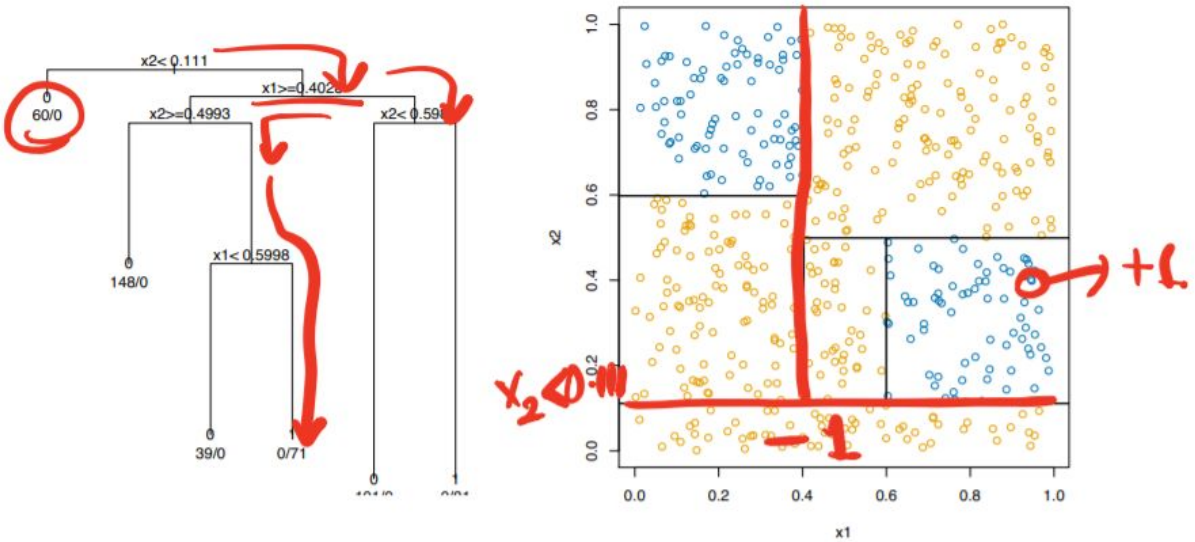


Support Vector Machines



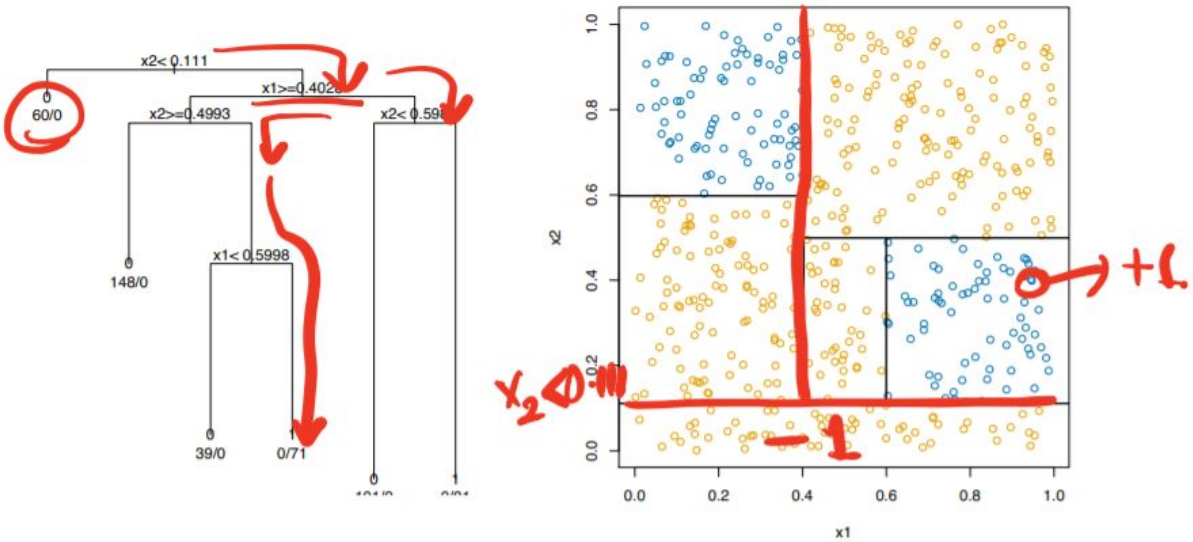

kernel
Decision surface

Decision Trees

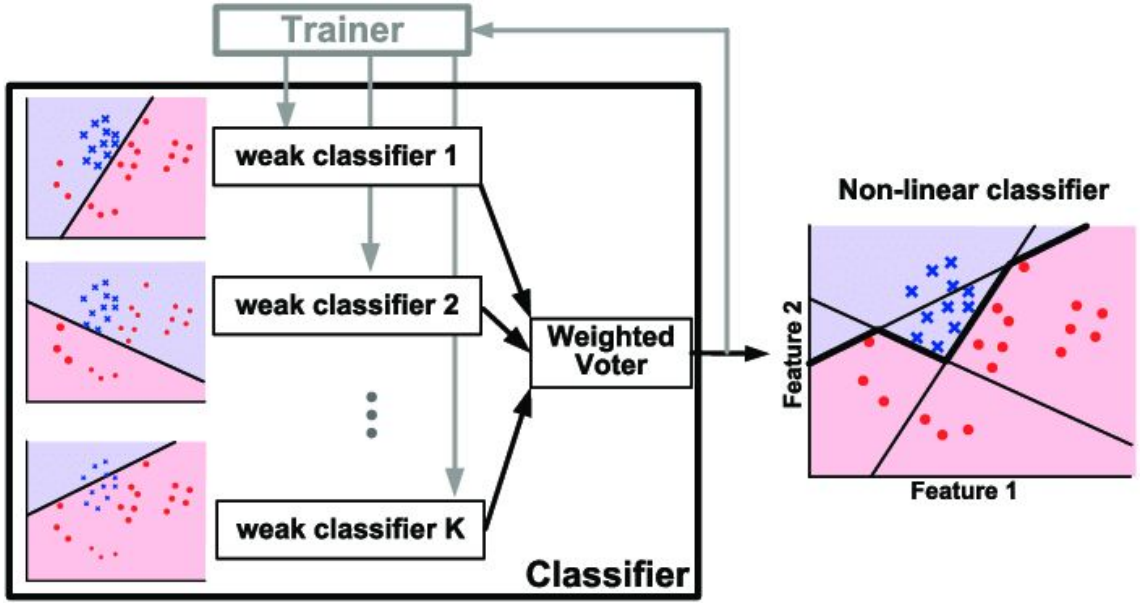


Decision Trees

CART Algorithm

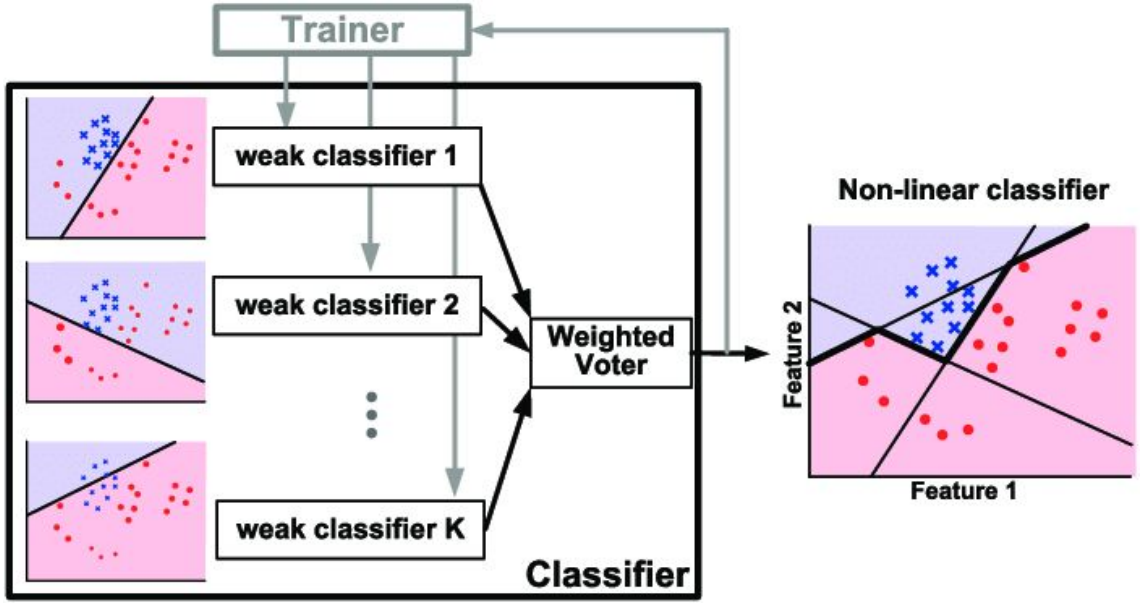


Boosting

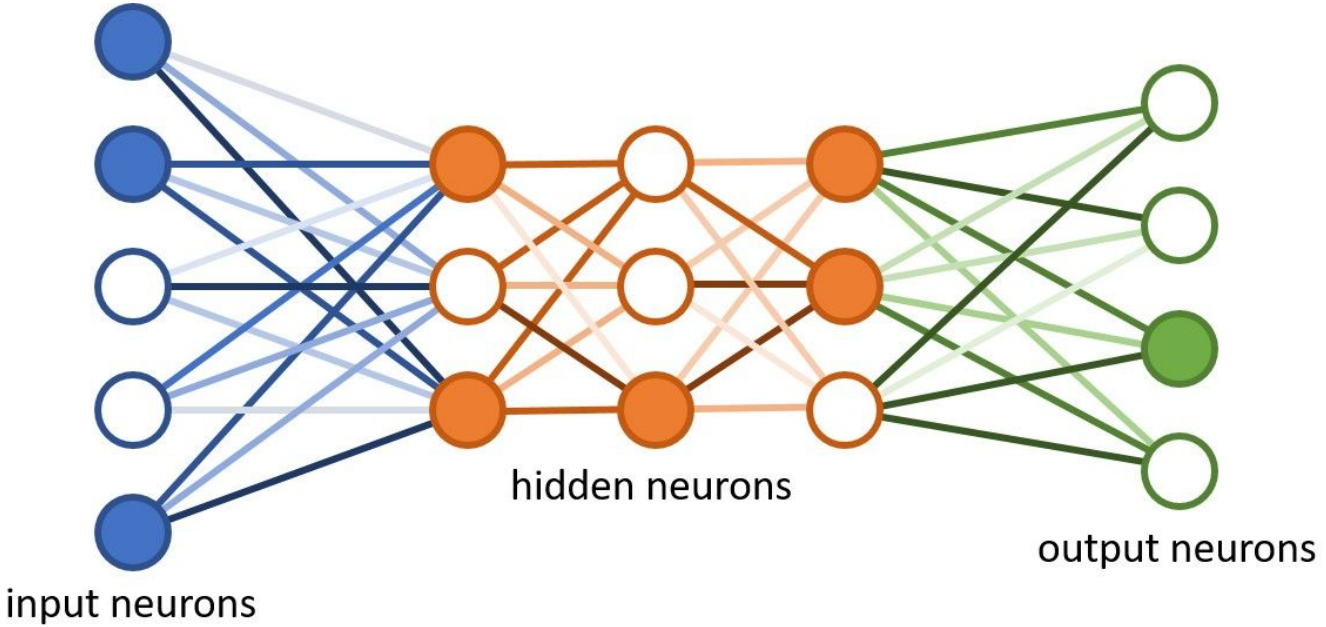

Trainer
weak classifier 1
weak classifier 2
weak classifier K
Weighted Voter
Classifier
Non-linear classifier
Feature 2
Feature 1

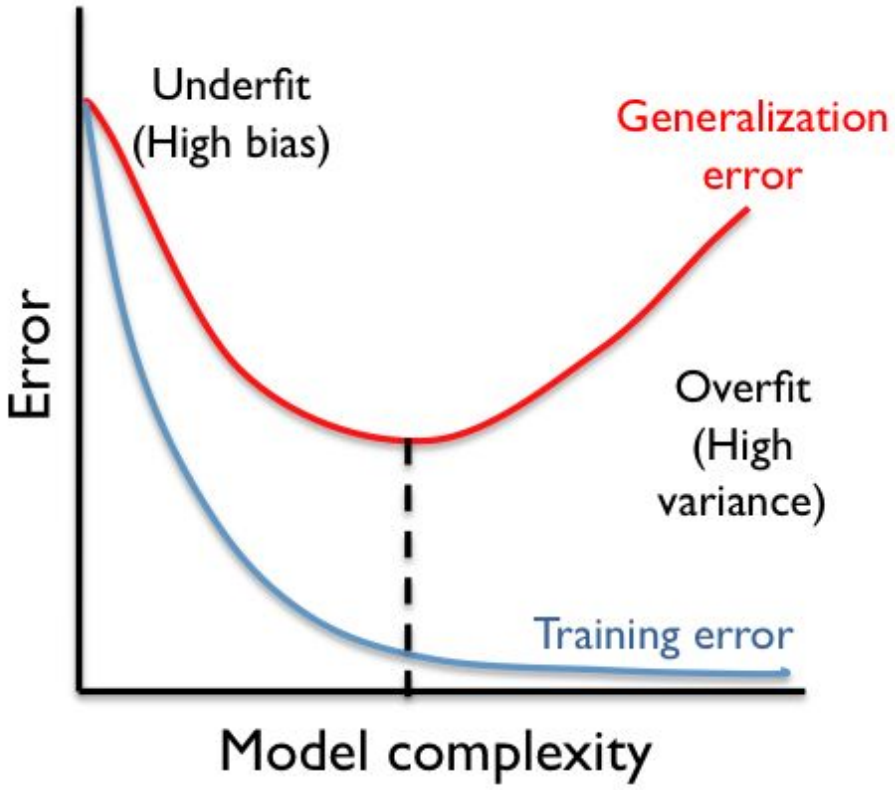
Boosting

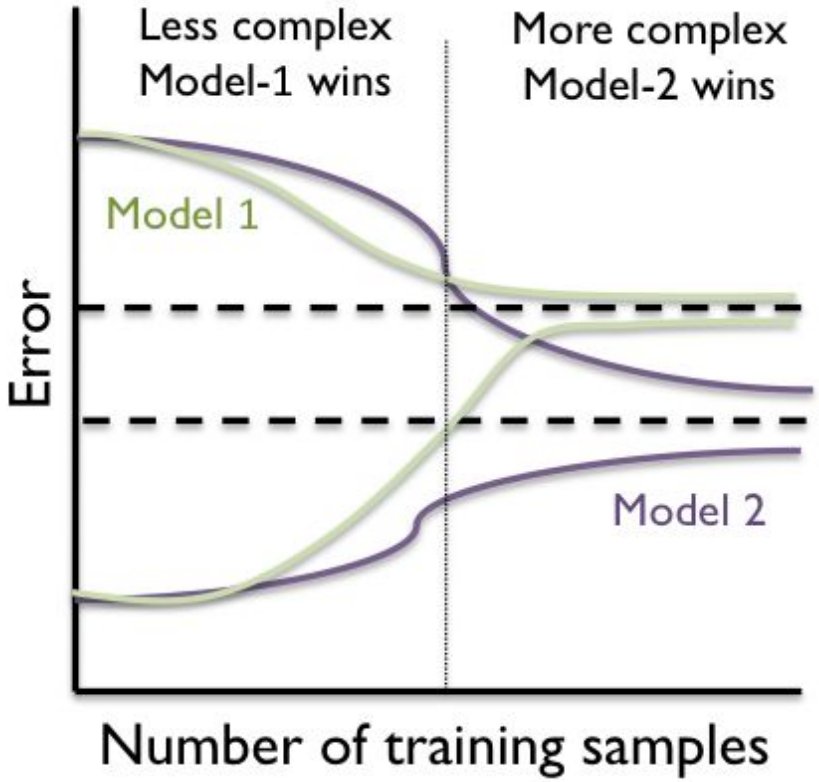
AdaBoost

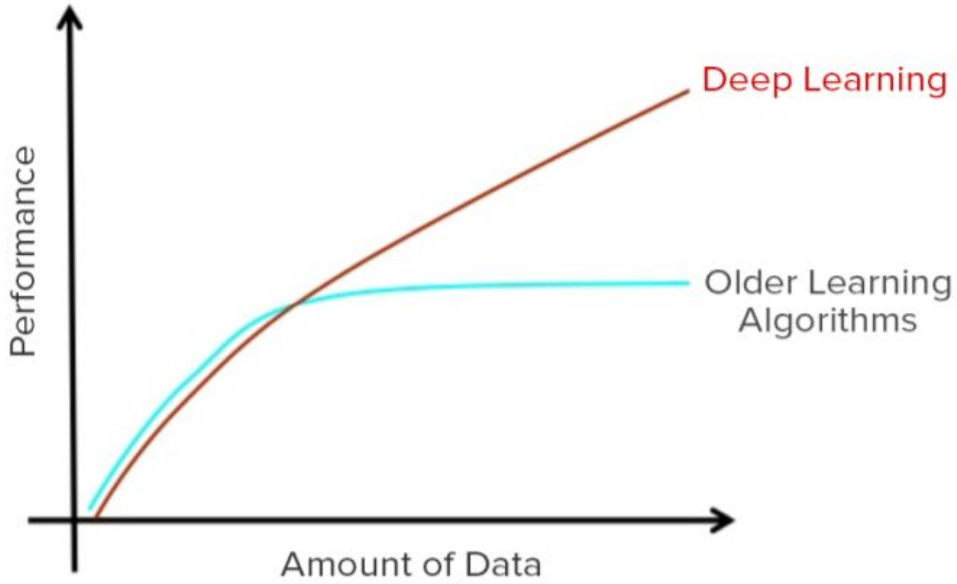


Neural Networks




Underfit
(High bias)
Generalization
error
Error
Overfit
(High
variance)
Training error
Model complexity


Less complex
Model-1 wins
More complex
Model-2 wins
Model 1
Model 2
Error
Number of training samples


Deep Learning
Performance
Older Learning
Algorithms
Amount of Data

tinyurl.com/btg-google-colab

tinyurl.com/btg-s19-super-ml