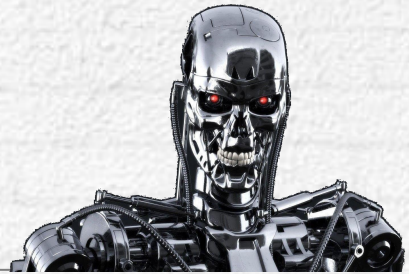
Come with me if you want to learn...



#DataPlatformSummit

Session will start soon...

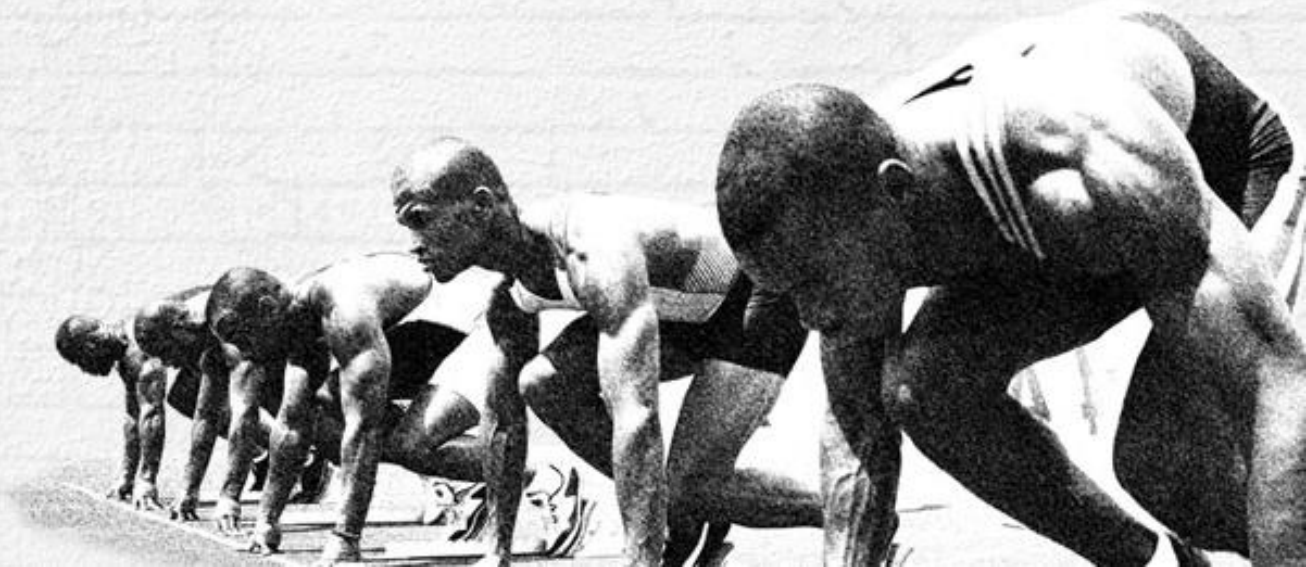
Machines Are Learning. You Should Learn ML Too: An Azure Story



Come with me if you want to learn...

André Melancia

2022-09-19



DATA PLATFORM VIRTUAL SUMMIT 2022

Diamond Knowledge Partner Microsoft

A Lunar Cat production...



From the producer of...
























DATA PLATFORM
VIRTUAL SUMMIT 2022
Machines Are Learning.
You Should Learn ML Too:
An Azure Story
Data Platform Virtual Summit
Andre' Melancia
Online
2022-09-19

Special Thanks To

Microsoft

for supporting

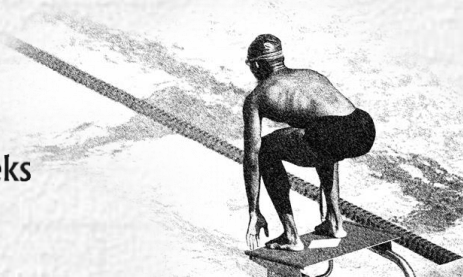
DataPlatformGeeks & SQLServerGeeks

Community Initiatives

SQLServerGeeks

Data Platform Summit

DataPlatformGeeks



Thank You

Three Ways to Win Prizes

Post your selfie with hash tag **#DPS2022**

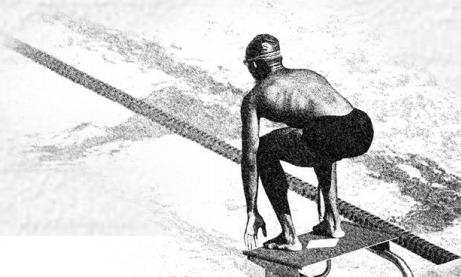
Give Session & Conference Feedback

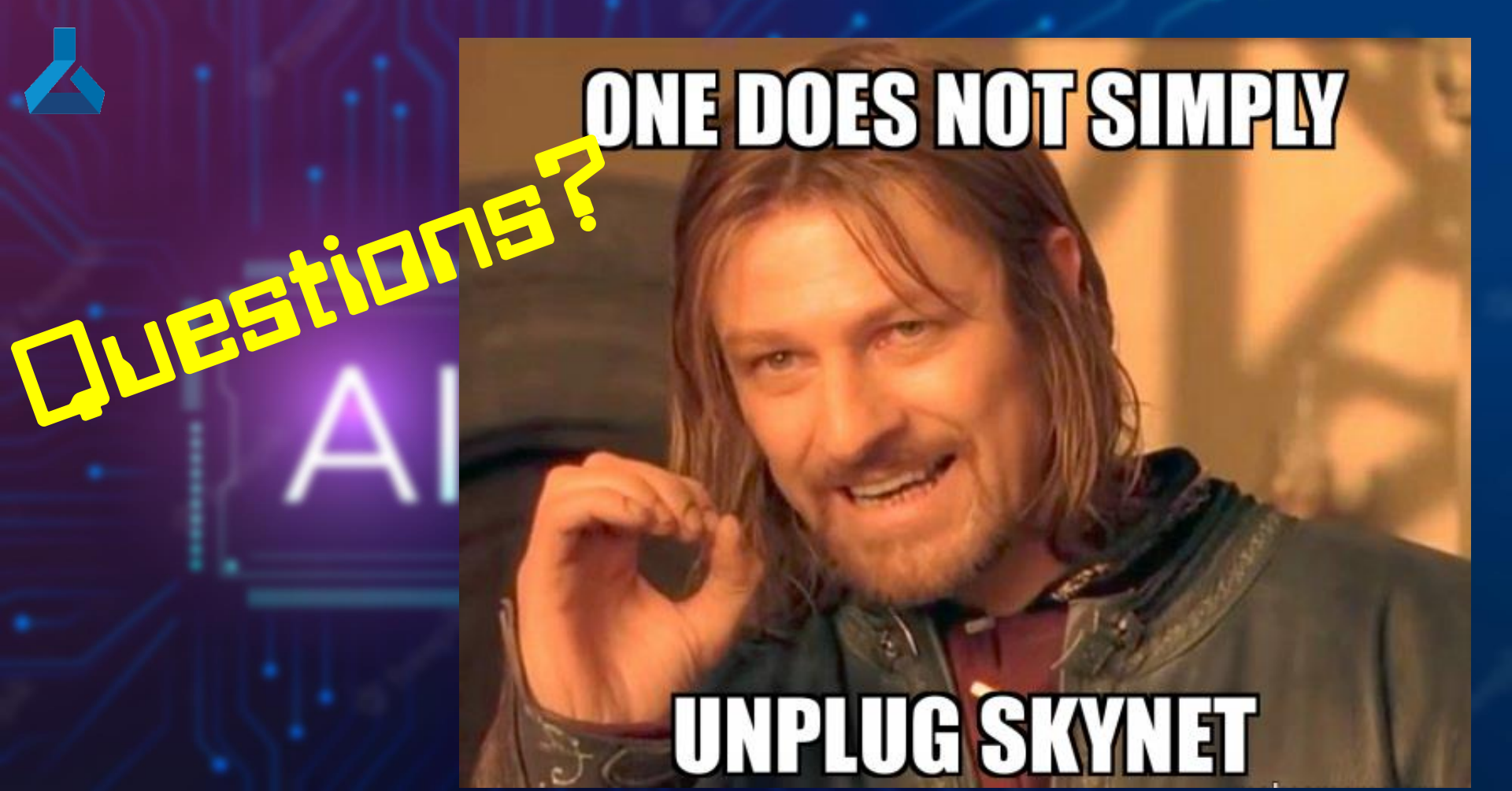
Visit our Sponsors & Exhibitors

Follow us on Twitter **@TheDataGeeks @DataAISummit**

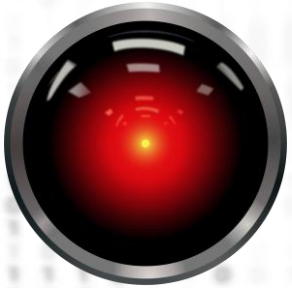
DATA PLATFORM
VIRTUAL SUMMIT 2022

Diamond Knowledge Partner Microsoft



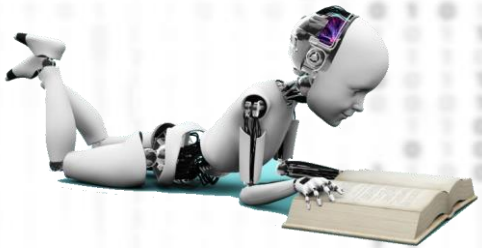

Questions?
ONE DOES NOT SIMPLY
UNPLUG SKYNET

Machine Learning ≠ Artificial Intelligence



Artificial Intelligence

- Rational agent that perceives its environment and takes actions that maximize its chance of success at some goal
- Solve Problems
- Make decisions (and do actions) NOT based on previous human programming



Machine Learning

- Sub-field of Artificial Intelligence
- Study and construction of algorithms that can learn from and make predictions on data.
- Pattern recognition (find patterns in data)
- Predict things



Types of Algorithms

- Supervised (YOU provide the result)
  - Linear regression
  - Classification
- Unsupervised (YOU don't provide the result)
  - Clustering
  - Imaging (e.g. Neural networks)



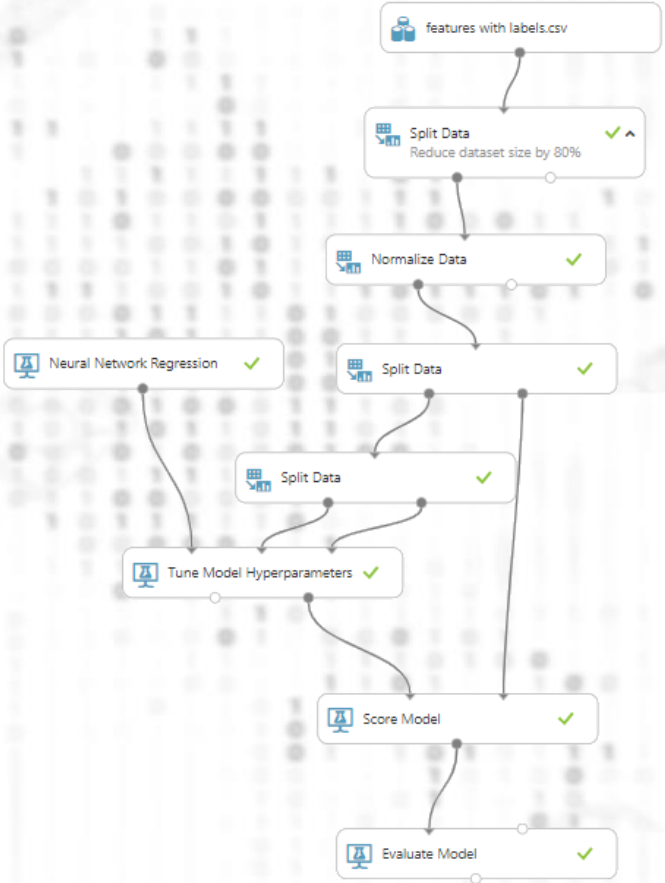
Advanced bovine statistics



I was told workflow was here. Could I see it, please?

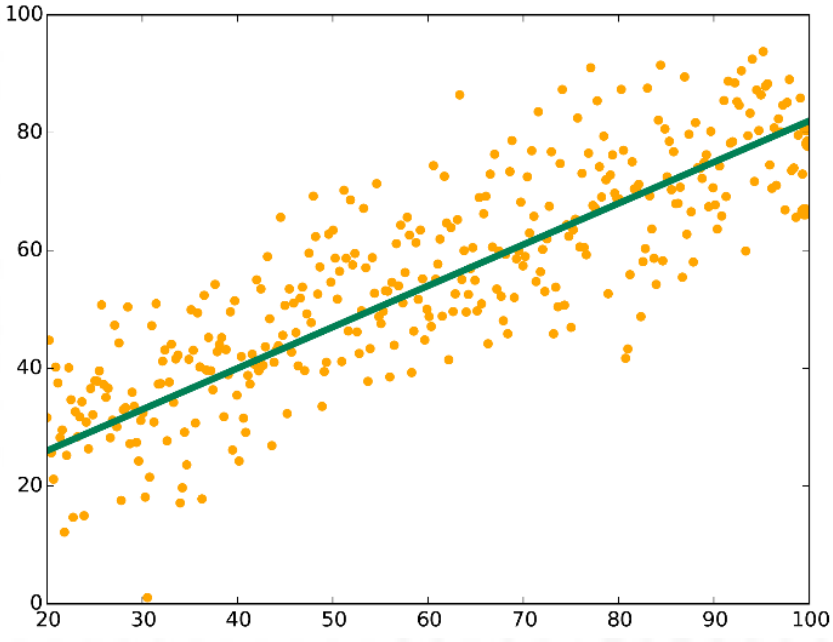
Machine Learning typical workflow:

1. Get datasets
2. Clean, prep, feature engineering
3. Training (Choosing Algorithm and Hyper-Parameters)
   - Regression, Classification (2-class, multi-class), Clustering, Anomaly Detection, etc.
   - Deep Learning, Neural Networks, etc.
4. Scoring/Testing
5. Evaluating
6. Deploying predictive webservices (inference)
7. I'll retrain back!

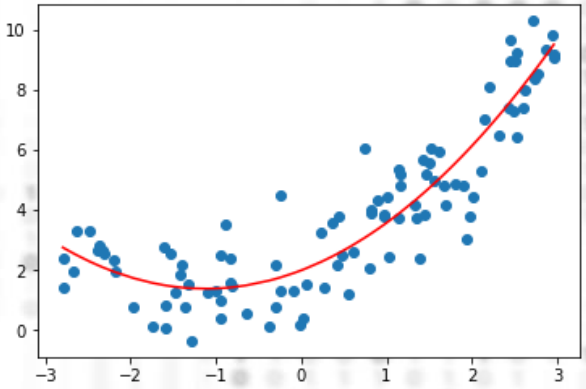


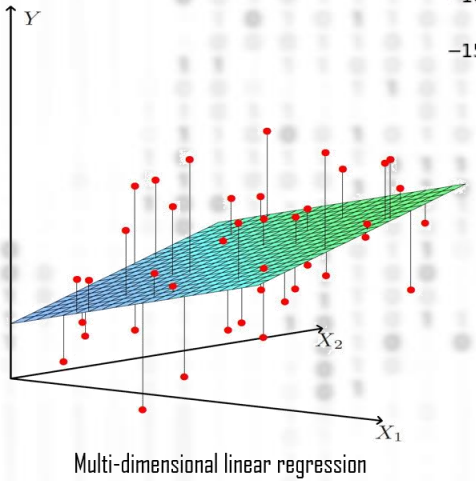
Algorithm: Regression

- Learn by finding a linear pattern in data
  - https://en.wikipedia.org/wiki/Linear_regression
- Predict numeric values (linear or not)
  - Trends
  - Weather
  - Economics
  - Biology
  - [...]

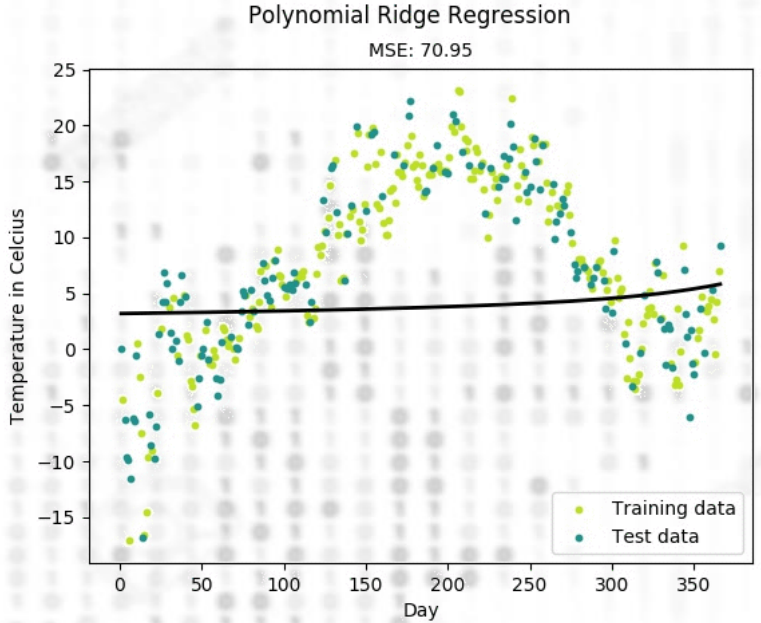


Algorithm: Regression



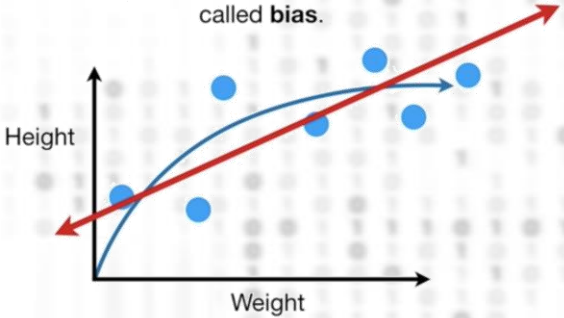


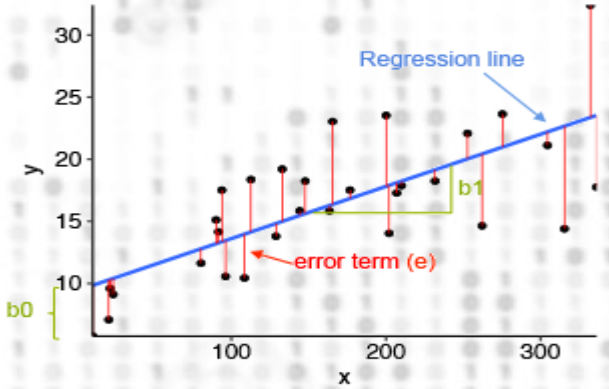
Multi-dimensional linear regression

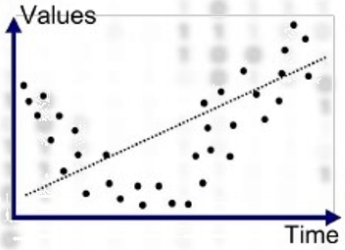


Algorithm: Regression

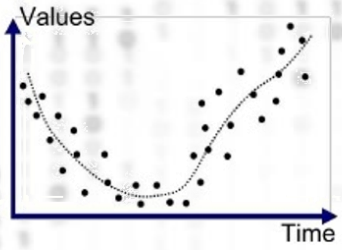
The inability for a machine learning method (like linear regression) to capture the true relationship is called **bias**.



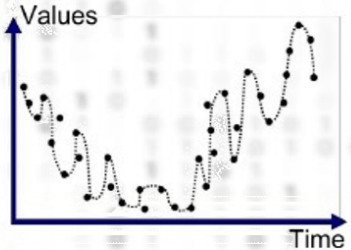




Underfitted

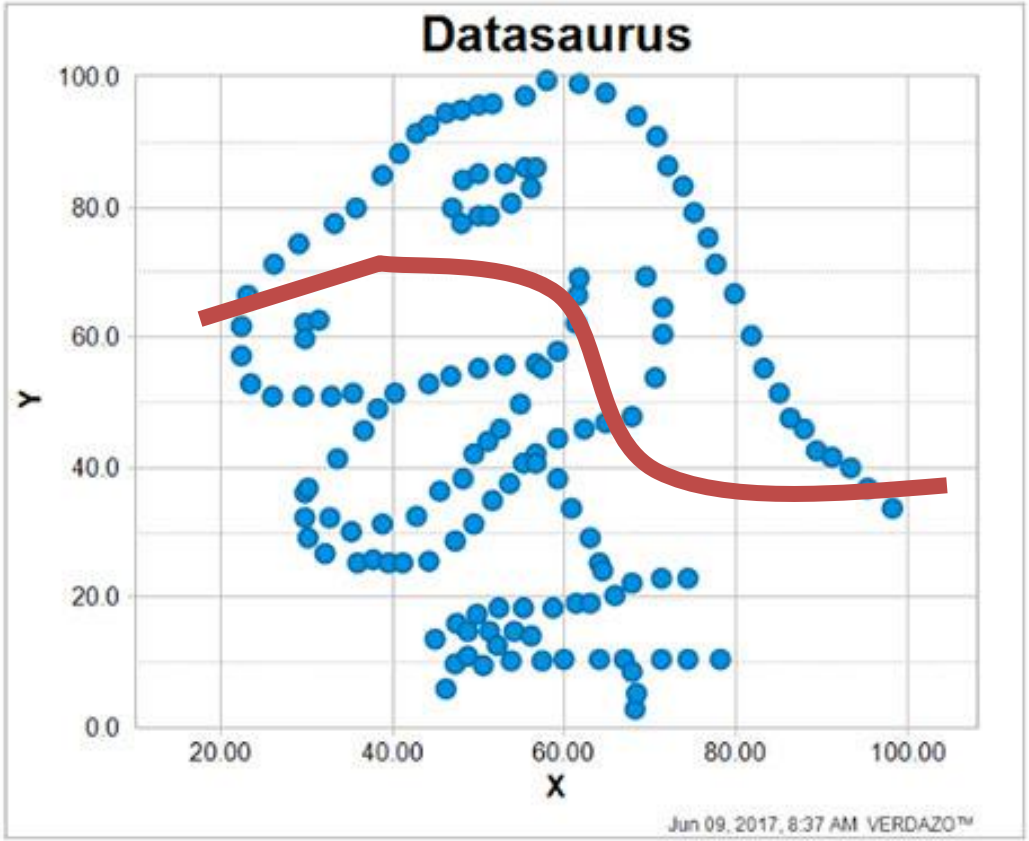


Good Fit/Robust



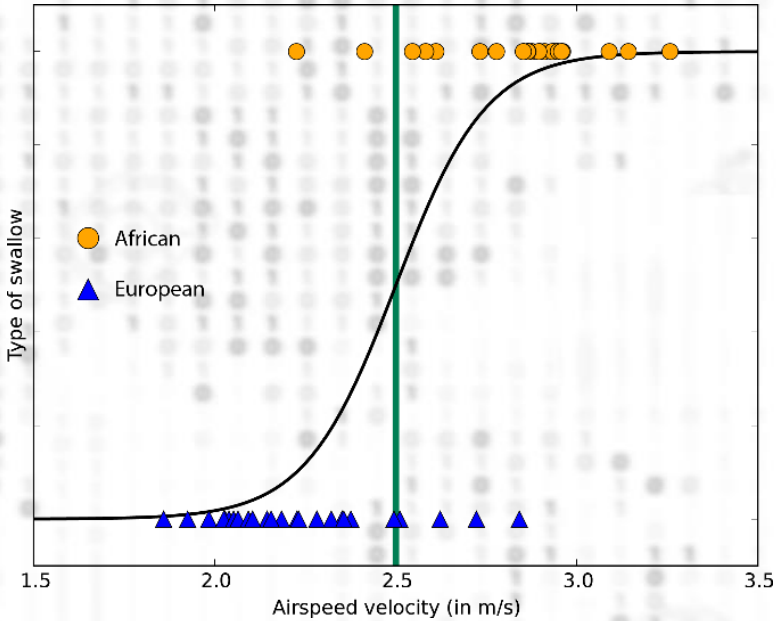
Overfitted

Beware of the statistics monsters...



Algorithm: Logistic Regression

- Classify in 2-class model
- Learn by finding a binary pattern in data
  - Either the positive or the negative (or A and B)
  - Returns a numeric value (e.g. percentage, where 0% is most likely false and 100% is most likely true)
- https://en.wikipedia.org/wiki/Logistic_regression

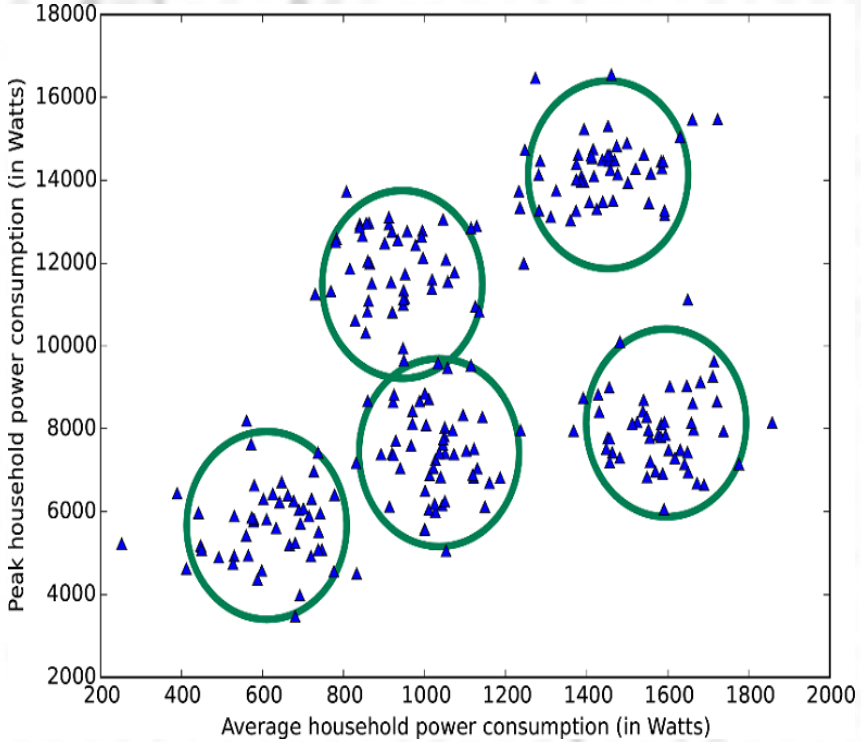


Data Scientist driven torture...



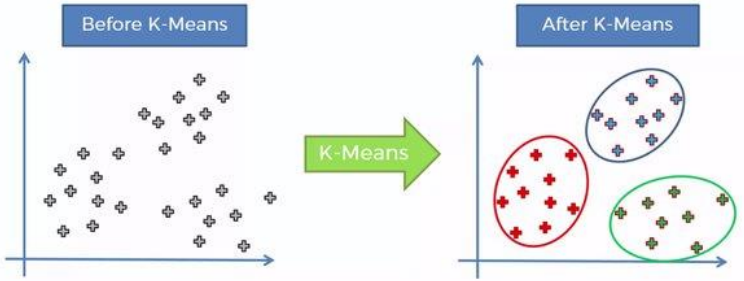
Algorithm: Clustering

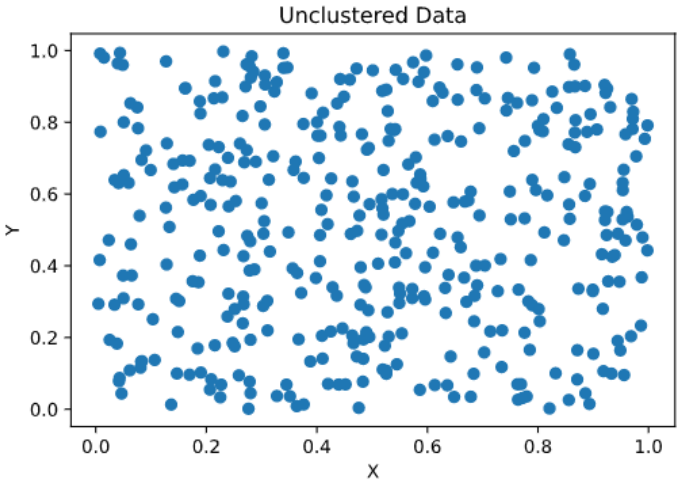
- Multiple algorithms for clustering
  - K-means, etc.
- Find grouping patterns in data
  - Shopping trends
  - Social network analysis
  - Crime analysis (fraud detection)

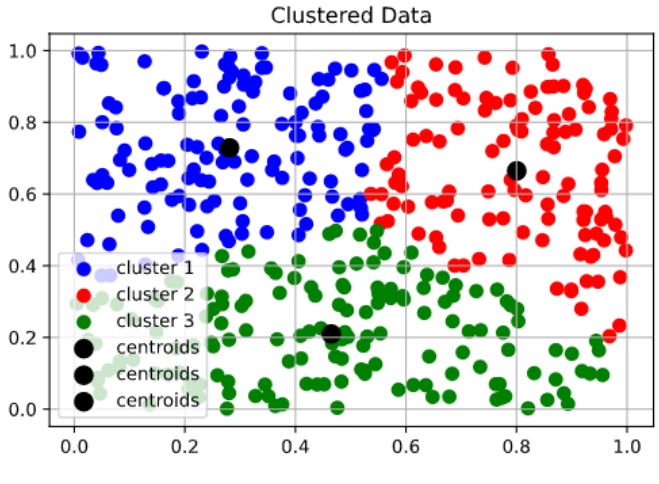


- https://en.wikipedia.org/wiki/K-means_clustering

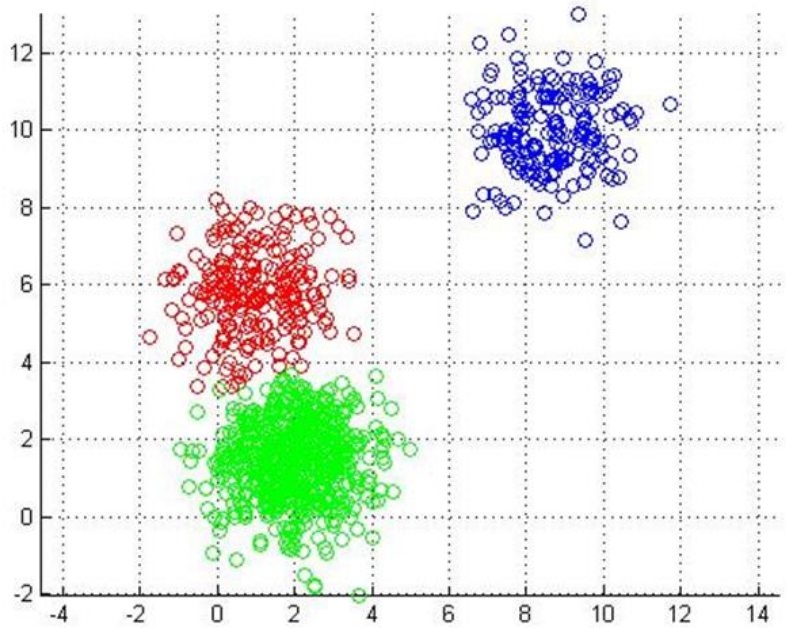
Algorithm: Clustering with K-means

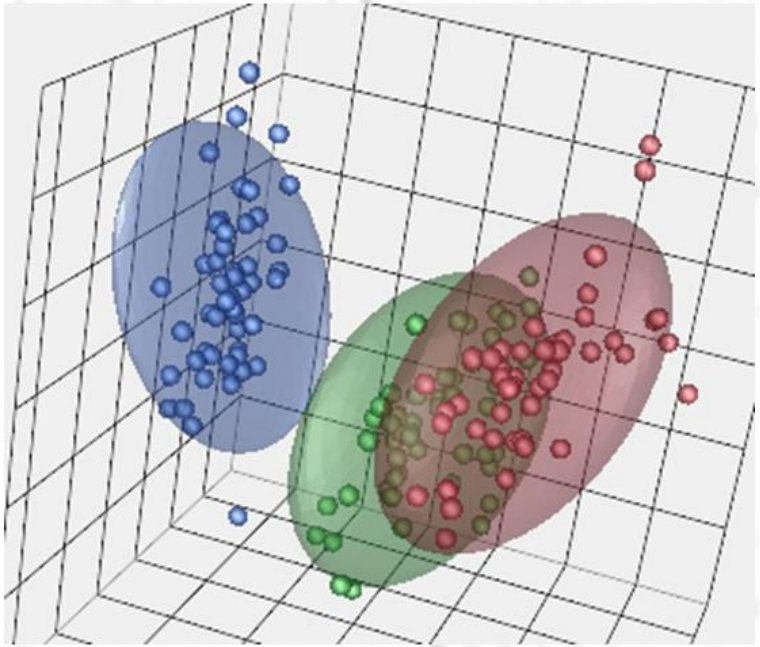






Algorithm: Clustering in N-dimensions



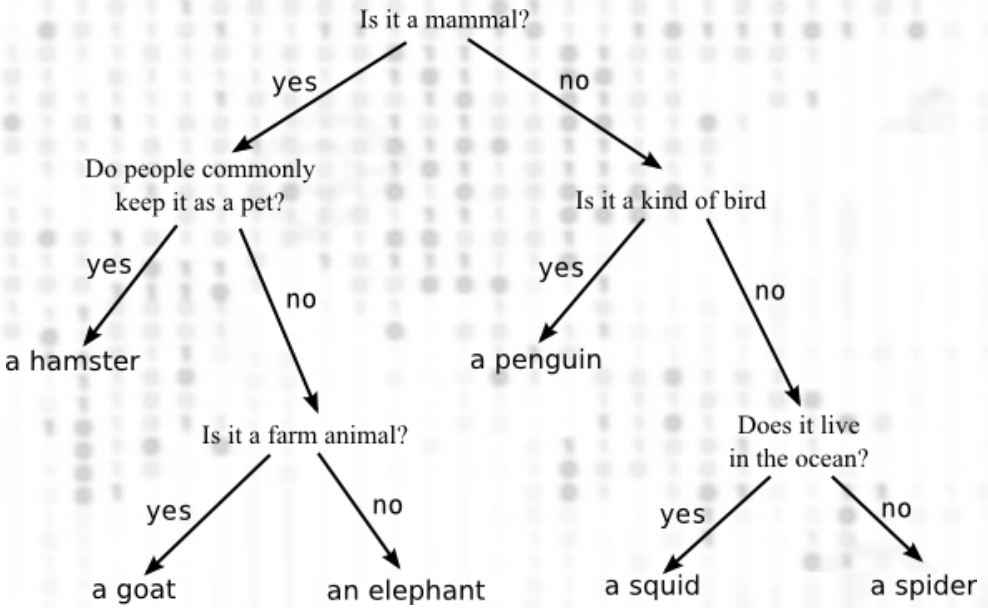


Explanation + code : https://stanford.edu/~cpiech/cs221/handouts/kmeans.html

Algorithm techniques: Decision tree

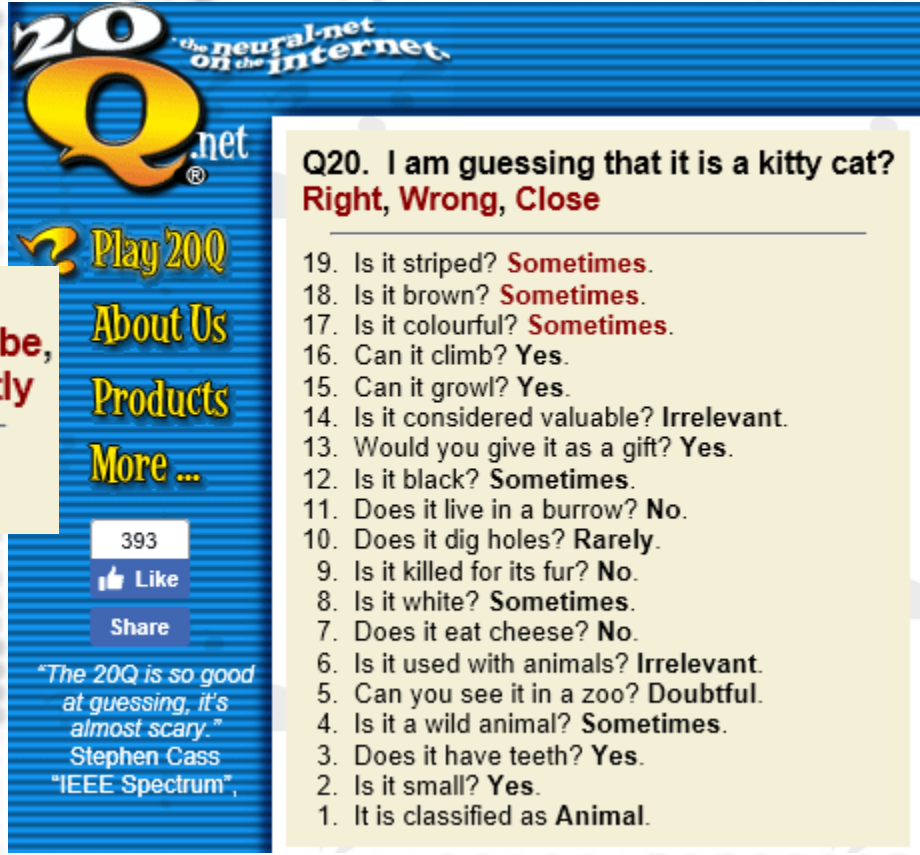
- E.g.Multi-class model
- Both for numerical and categorical data.
- Conditions explained by simple Boolean logic.
- Easy to expand the tree if more detail is needed

- https://en.wikipedia.org/wiki/Decision_tree_learning



Decision tree in the wild (since the 1990s)

- http://20Q.NET

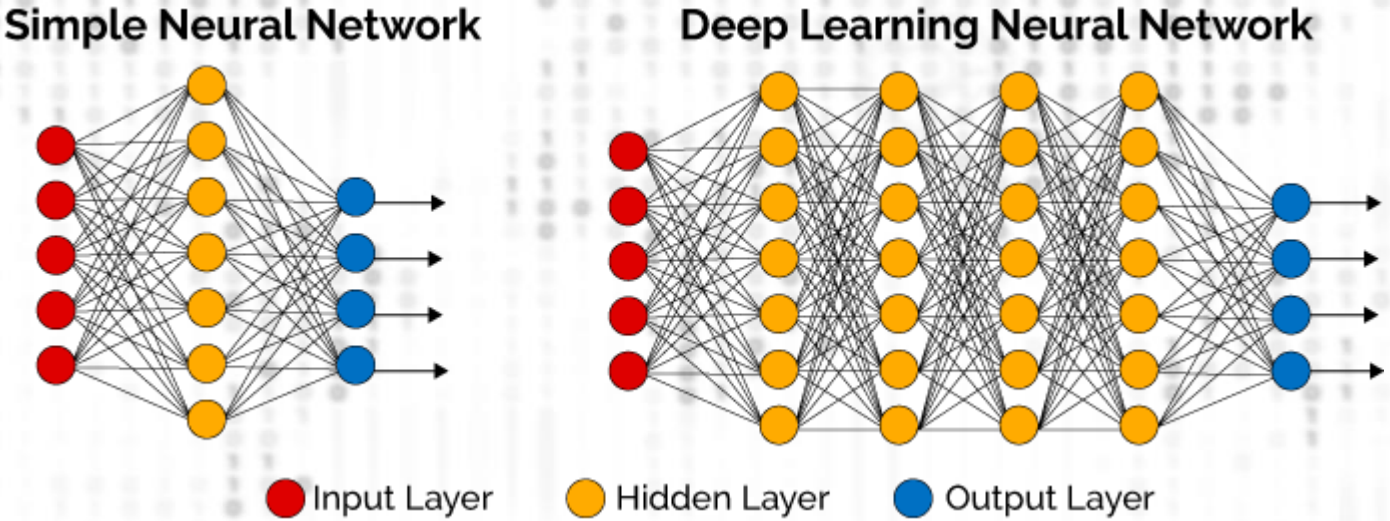


Q3. Will it eat almost anything?
Yes , No , Unknown, Irrelevant, Sometimes, Maybe, Probably, Doubtful, Usually, Depends, Rarely, Partly

2. Can it scratch? Yes.
1. It is classified as Animal .

Algorithm techniques: Neural Networks

- E.g.Imaging, Deep Learning
- Slowest but most accurate
- https://en.wikipedia.org/wiki/Artificial_neural_network
- https://en.wikipedia.org/wiki/Deep_learning



Machine Learning Algorithm Cheat Sheet

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution.
Your decision is driven by both the nature of your data and the goal you want to achieve with your data.

Microsoft Azure

What do you want to do?

Extract information from text

Text Analytics

Derives high-quality information from text

Answers questions like: What info is in this text?

- Latent Dirichlet Allocation ← Unsupervised topic modeling, group texts that are similar
- Extract N-Gram Features from Text ← Creates a dictionary of n-grams from a column of free text
- Feature Hashing ← Converts text data to integer encoded features using the Vowpal Wabbit library
- Preprocess Text ← Performs cleaning operations on text, like removal of stop-words, case normalization
- Word2Vector ← Converts words to values for use in NLP tasks, like recommender, named entity recognition, machine translation

Predict between several categories

Multiclass Classification

Answers complex questions with multiple possible answers

Answers questions like: Is this A or B or C or D?

- Multiclass Logistic Regression ← Fast training times, linear model
- Multiclass Neural Network ← Accuracy, long training times
- Multiclass Decision Forest ← Accuracy, fast training times
- One-vs-All Multiclass ← Depends on the two-class classifier
- One-vs-One Multiclass ← Depends on binary classifier, less sensitive to an imbalanced dataset with larger complexity
- Multiclass Boosted Decision Tree ← Non-parametric, fast training times and scalable

Predict between two categories

Two-Class Classification

Answers simple two-choice questions, like yes or no, true or false

Answers questions like: Is this A or B?

- Two-Class Support Vector Machine ← Under 100 features, linear model
- Two-Class Averaged Perceptron ← Fast training, linear model
- Two-Class Decision Forest ← Accurate, fast training
- Two-Class Logistic Regression ← Fast training, linear model
- Two-Class Boosted Decision Tree ← Accurate, fast training, large memory footprint
- Two-Class Neural Network ← Accurate, long training times

Generate recommendations

Recommenders

Predicts what someone will be interested in

Answers the question: What will they be interested in?

- Use the Train Wide & Deep Recommender module ← Hybrid recommender, both collaborative filtering and content-based approach
- SVD Recommender ← Collaborative filtering, better performance with lower cost by reducing dimensionality

Predict values

Regression

Makes forecasts by estimating the relationship between values

Answers questions like: How much or how many?

- Fast Forest Quantile Regression ← Predicts a distribution
- Poisson Regression ← Predicts event counts
- Linear Regression ← Fast training, linear model
- Bayesian Linear Regression ← Linear model, small data sets
- Decision Forest Regression ← Accurate, fast training times
- Neural Network Regression ← Accurate, long training times
- Boosted Decision Tree Regression ← Accurate, fast training times, large memory footprint

Discover structure

Clustering

Separates similar data points into intuitive groups

Answers questions like: How is this organized?

- K-Means ← Unsupervised learning

Find unusual occurrences

Anomaly Detection

Identifies and predicts rare or unusual data points

Answers the question: Is this weird?

- One Class SVM ← Under 100 features, aggressive boundary
- PCA-Based Anomaly Detection ← Fast training times

Classify images

Image Classification

Classifies images with popular networks

Answers questions like: What does this image represent?

- ResNet ← Modern deep learning neural network
- DenseNet ← Modern deep learning neural network

© 2021 Microsoft Corporation. All rights reserved.

Share this poster: aka.ms/mlcheatsheet



Humans...

Using Machine Learning in the Microsoft Universe doesn't require Data Science background!

Anyone can do it!



"This is the world now. Logged on, plugged in, all the time."

1. Azure Machine Learning (Service / Workspace, not "Classic")
   - SDK includes a lot of stuff (AutoML, HyperDrive, Model registry, SDK support for webservices in containers, etc.) – Supports Python, R (incomplete), MLFlow, etc.
2. Azure Databricks (Spark clusters, SQL, Python, R, Scala)
3. Azure HDInsight Spark
4. Azure HDInsight R Server
5. Azure Synapse Analytics (with SQL or Spark – SQL, Python, R, Scala, .Net)
6. SQL Server 2016 (SQL, .Net or R Services)
7. SQL Server 2017/2019/2022 (SQL, .Net, ML Services using R or Python)
8. Any platform with .Net Core and libraries for ML.Net
9. Azure Cognitive Services + Bots (managed webservices) + M365 Power Virtual Agents
10. Etc...

- Some of these include their own notebooks (Jupyter, Zeppelin, ODC, etc.) and you can use your tools like Azure Data Studio, etc.

Magic is in the libraries, not the languages!


"Come with me if you want to DEMO..."

Special Thanks To

Microsoft

for supporting

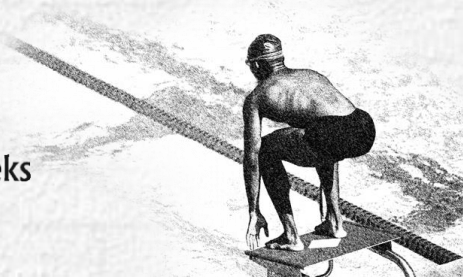
DataPlatformGeeks & SQLServerGeeks

Community Initiatives

SQLServerGeeks

Data Platform Summit

DataPlatformGeeks



Thank You

Three Ways to Win Prizes

Post your selfie with hash tag **#DPS2022**

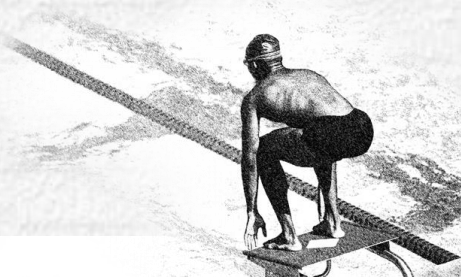
Give Session & Conference Feedback

Visit our Sponsors & Exhibitors

Follow us on Twitter **@TheDataGeeks @DataAISummit**

DATA PLATFORM
VIRTUAL SUMMIT 2022

Diamond Knowledge Partner Microsoft





"The unknown future rolls toward us.
I face it for the first time with a sense of hope, because **if a machine, a Terminator, can learn the value of human life, maybe we can, too.**"

Thank you!
धन्यवाद्
Tack så mycket!
Благодаря!
¡Muchas gracias!
Дуже дякую!
ধন্যবাদ
Vielen Danke!
Obrigado, pá!
Hvala vam!
ඔයාට ස්තුතියි
Ευχαριστώ!
Merci beaucoup!
Terima kasih!
Grazie mille!
Ďakujem!
Mulțumesc!
Labai ačiū!
Dziękuję Wam!
Mockrát děkuju!
Mange tak!
Kiitos!
Takk fyrir!
Dank u wel!
Takk!
Dank je!
Köszönöm!
Go raibh maith agaibh!
Diolch!



André Melancia
Andy.PT
LunarCat.PT