Come with me
if you want to
learn...

🐾 A Lunar Cat production...

AI Subtitles?

Machines Are Learning, You Should Learn ML Too: An Azure Story *by* André Melancia

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
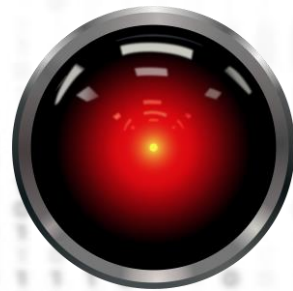
→ Machine Learning

- → 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

techforum technical community
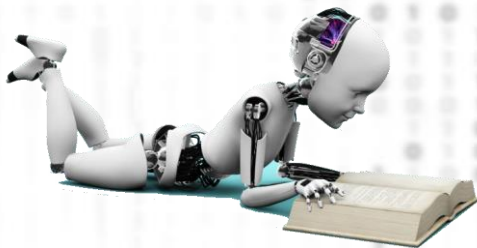
Types of Algorithms

→ Supervised (YOU provide the result)

  → Linear regression

  → Classification

→ Unsupervised (YOU don't provide the result)

  → Clustering

  → Imaging (e.g. Neural networks)

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!
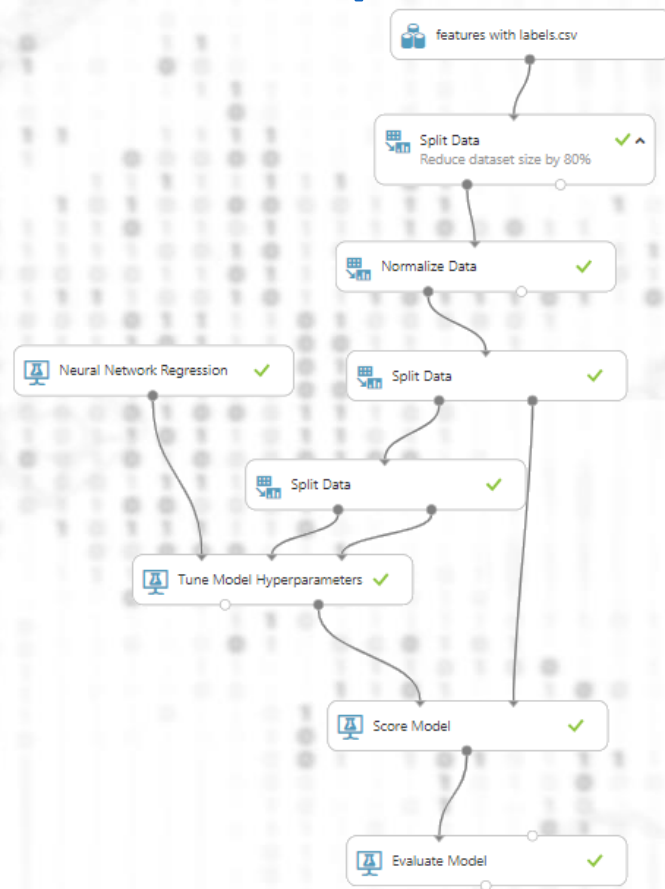
Algorithm: Regression

→ # 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

Polynomial Ridge Regression

MSE: 70.95
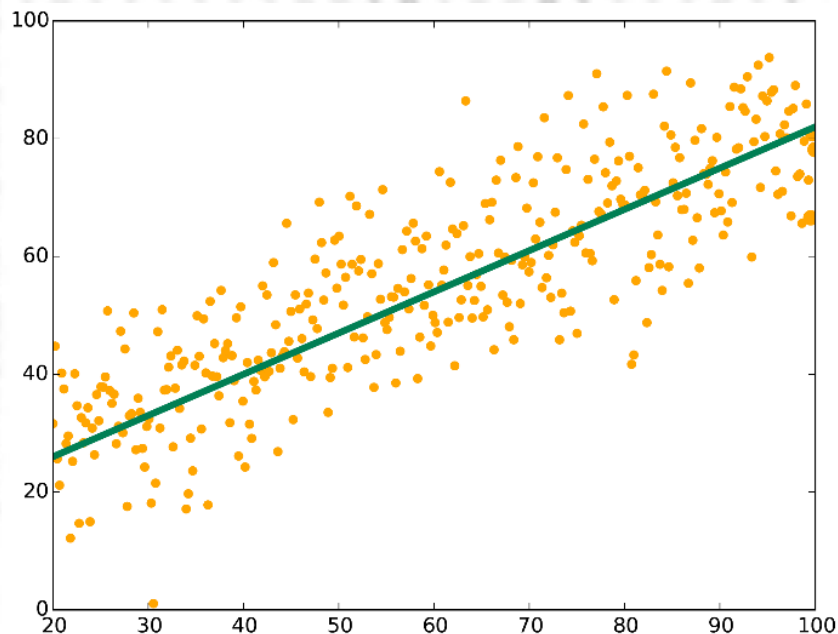
Multi-dimensional linear regression
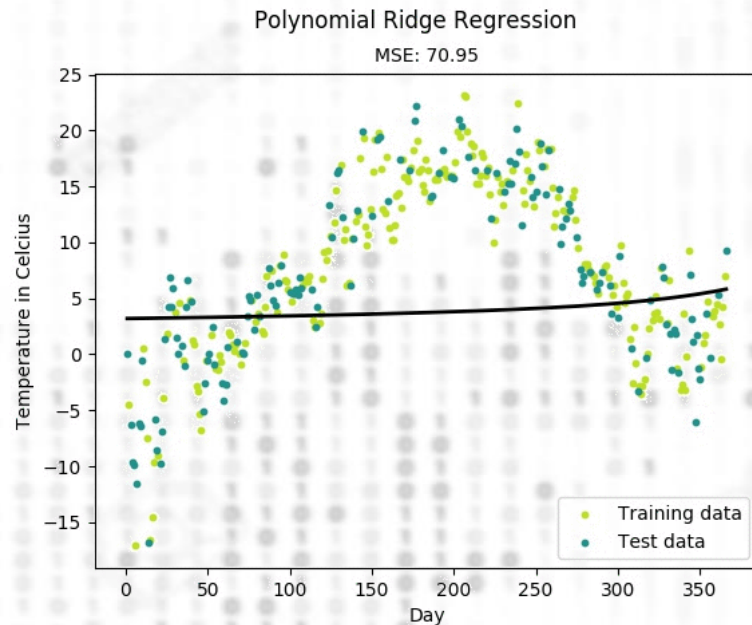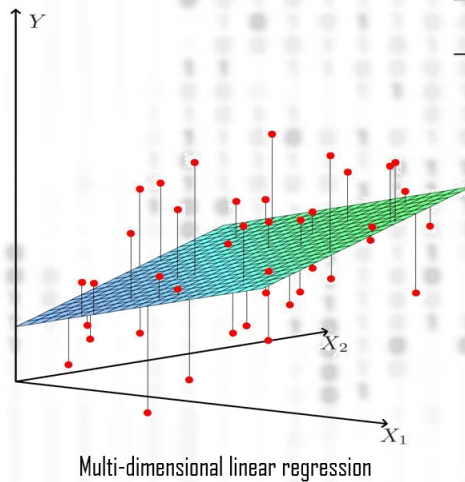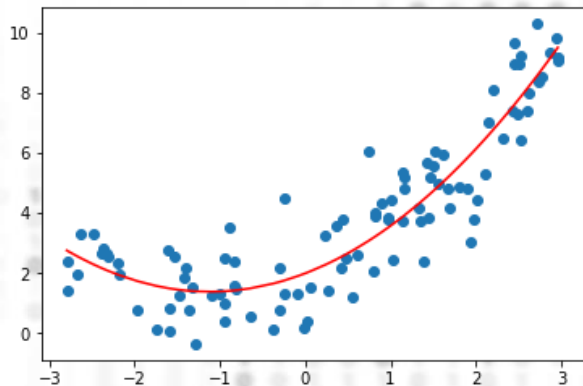
Algorithm: Regression

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

Height

Weight

Regression line

y

30

25

20

15

10

b1

error term (e)

b0

100 200 300

x

Values

Time

Underfitted

Values

Time

Good Fit/Robust

Values

Time
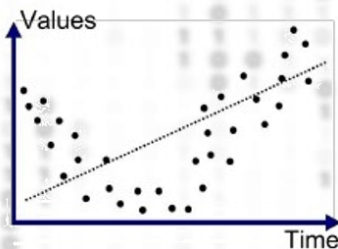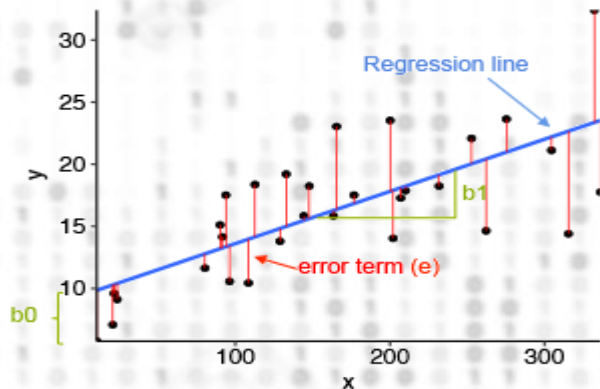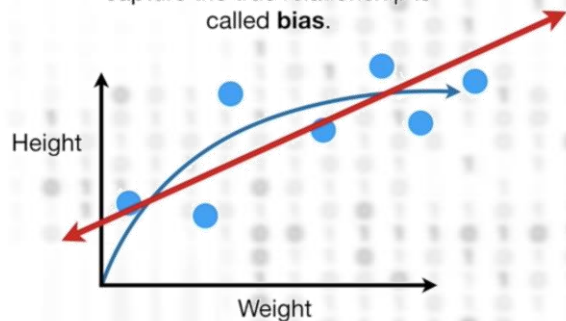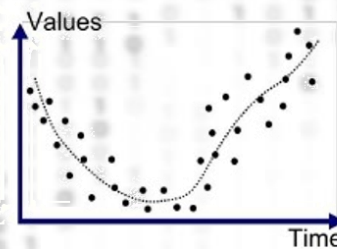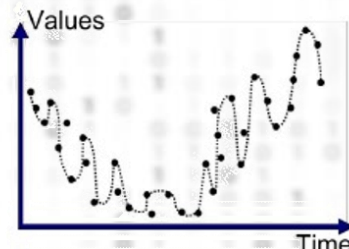
Overfitted

Beware of the statistics monsters...

Datasaurus
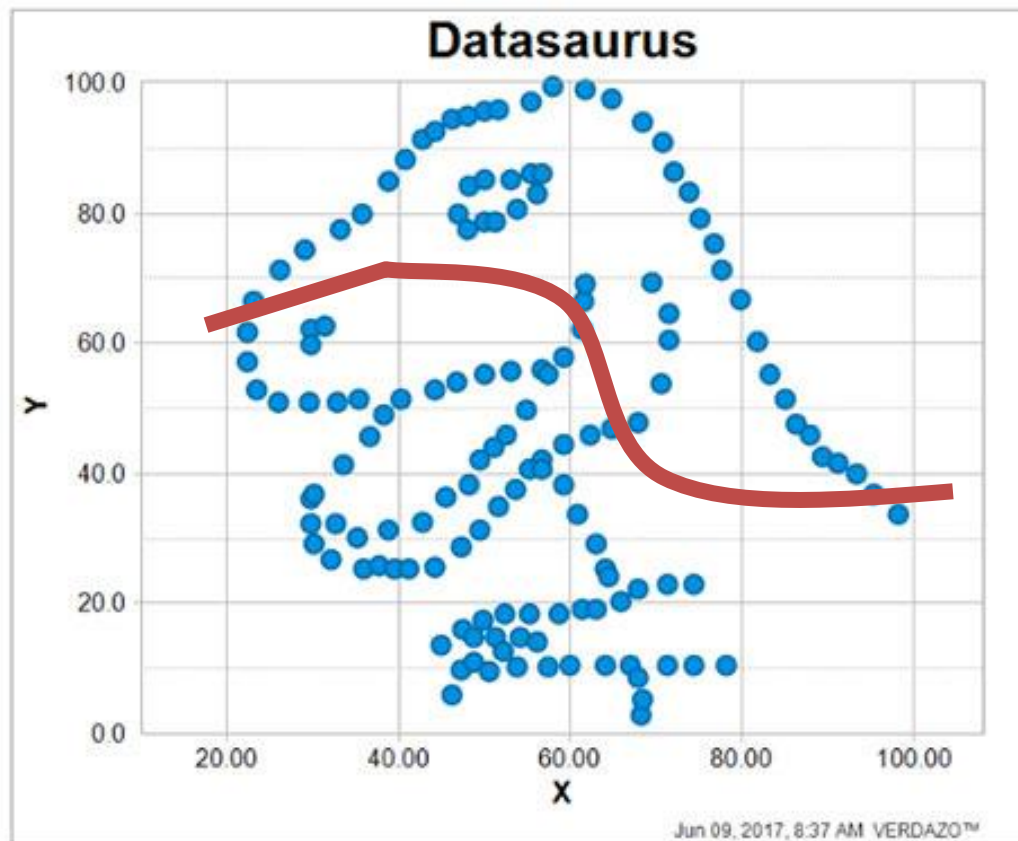
Algorithm: Logistic Regression

→ 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
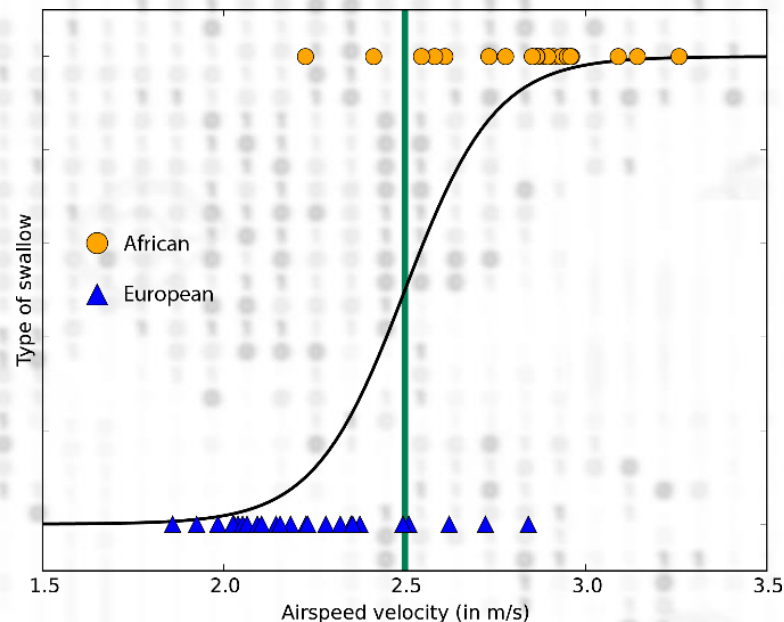
If you torture the data long enough, it will confess.
Ronald Coase

Algorithm: Clustering

- → Multiple algorithms for clustering
  - → K-means, etc.
- → Find grouping patterns in data
  - → Shopping trends
  - → Social network analysis
  - → Crime analysis (fraud detection)
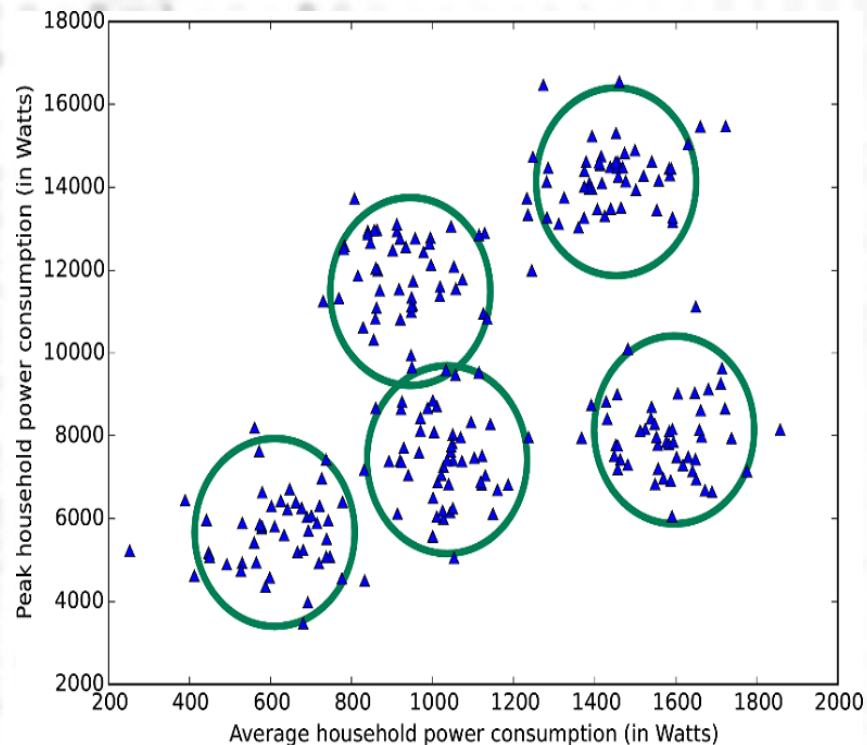
→ https://en.wikipedia.org/wiki/K-means_clustering

techforum technical community

Algorithm: Clustering with K-means
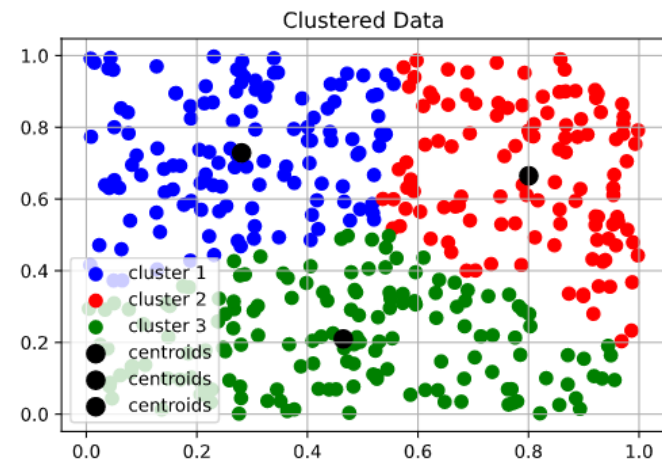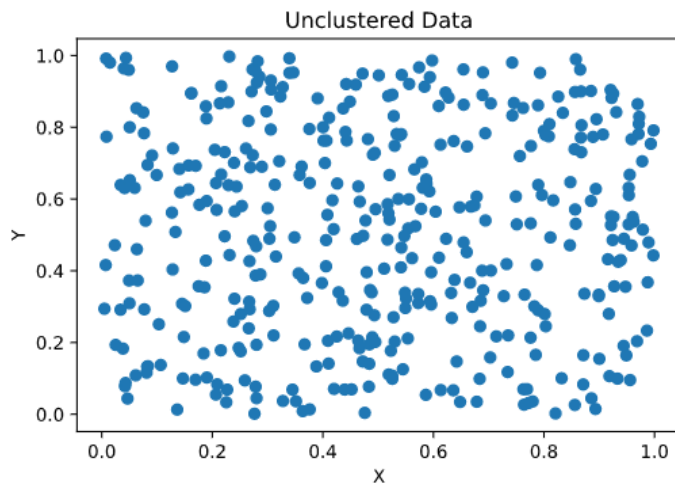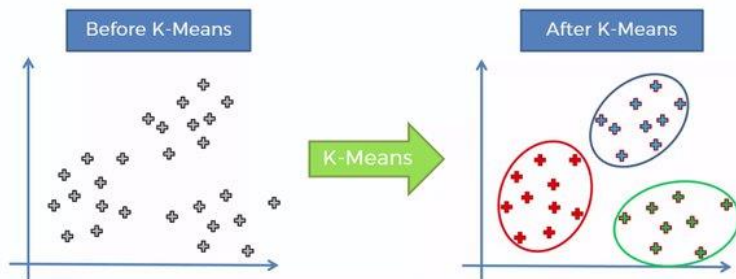
Algorithm: Clustering in N-dimensions
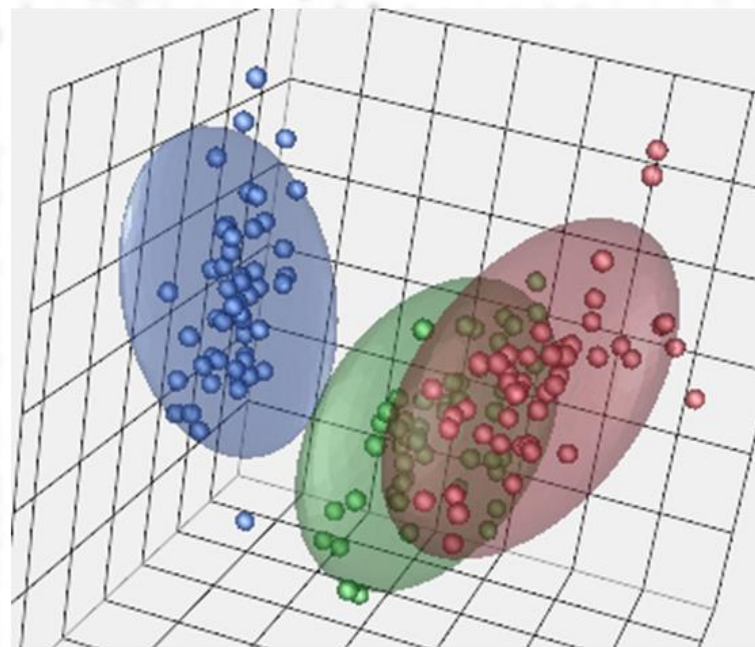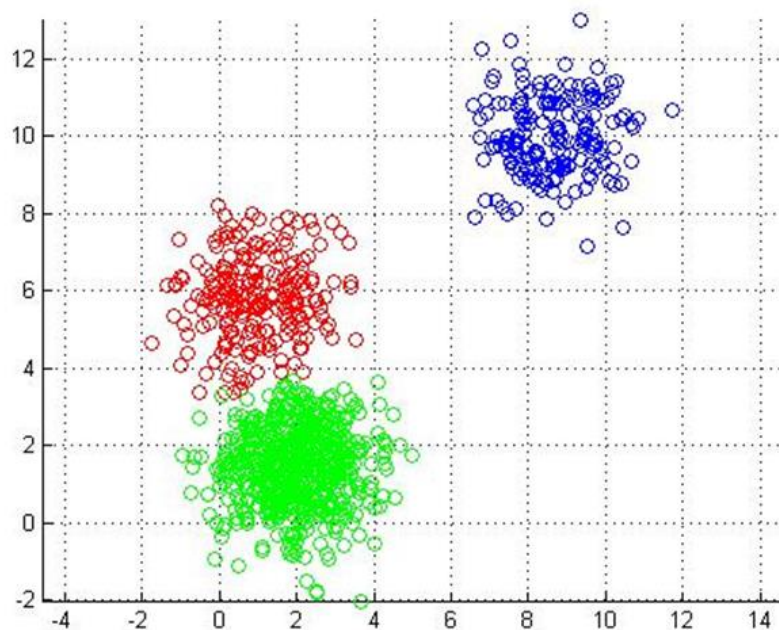
Explanation + code : https://stanford.edu/~cpiech/cs221/handouts/kmeans.html

STATISTICS
THE DISCIPLINE THAT PROVES
THE AVERAGE HUMAN HAS
ONE TESTICLE
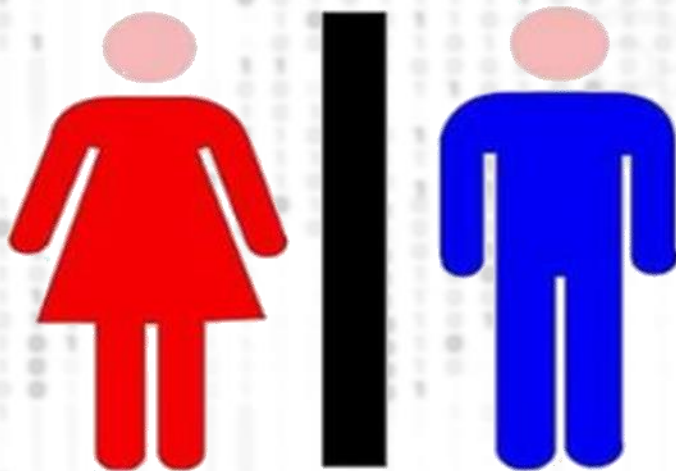
- → E.g.Multi-class model
- → Both for numerical and categorical data.
- → Conditions explained by simple Boolean logic.
- → Easy to expand the tree if more detail is needed

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

Is it a mammal?

yes → Do people commonly keep it as a pet?

no → Is it a kind of bird

Do people commonly keep it as a pet?
yes → a hamster
no → Is it a farm animal?

Is it a kind of bird
yes → a penguin
no → Does it live in the ocean?

Is it a farm animal?
yes → a goat
no → an elephant

Does it live in the ocean?
yes → a squid
no → a spider
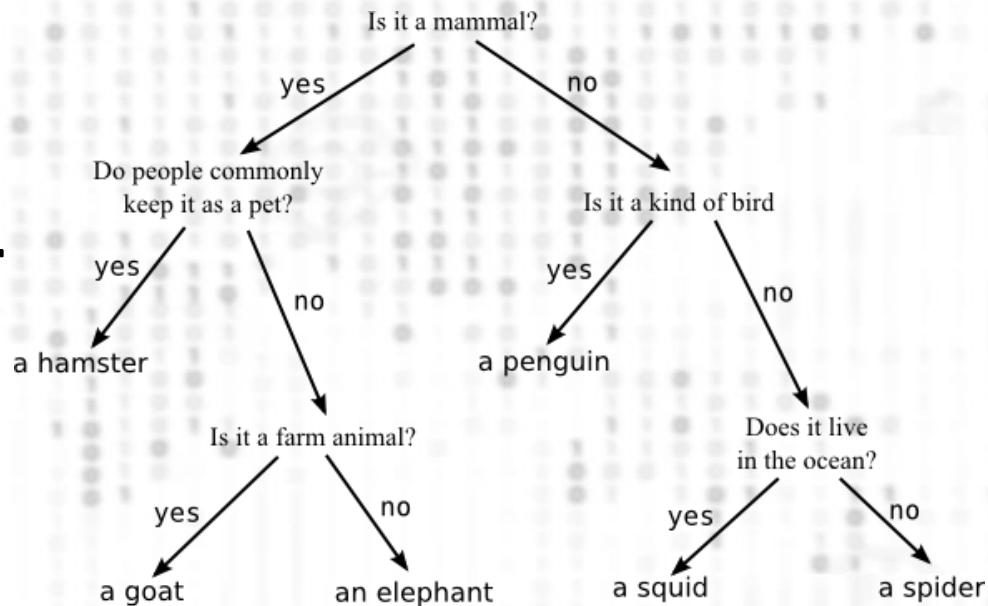
Decision tree in the wild (since the 1990s)

→ http://20Q.NET

20Q ·the neural-net on the internet.

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.

? Play 20Q
About Us
Products
More ...

393
Like
Share

"The 20Q is so good at guessing, it's almost scary."
Stephen Cass
"IEEE Spectrum",

Q20. I am guessing that it is a kitty cat?
Right, Wrong, Close

19. Is it striped? Sometimes.
18. Is it brown? Sometimes.
17. Is it colourful? Sometimes.
16. Can it climb? Yes.
15. Can it growl? Yes.
14. Is it considered valuable? Irrelevant.
13. Would you give it as a gift? Yes.
12. Is it black? Sometimes.
11. Does it live in a burrow? No.
10. Does it dig holes? Rarely.
9. Is it killed for its fur? No.
8. Is it white? Sometimes.
7. Does it eat cheese? No.
6. Is it used with animals? Irrelevant.
5. Can you see it in a zoo? Doubtful.
4. Is it a wild animal? Sometimes.
3. Does it have teeth? Yes.
2. Is it small? 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

Simple Neural Network

Deep Learning Neural Network

🔴 Input Layer   🟠 Hidden Layer   🔵 Output Layer

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

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

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
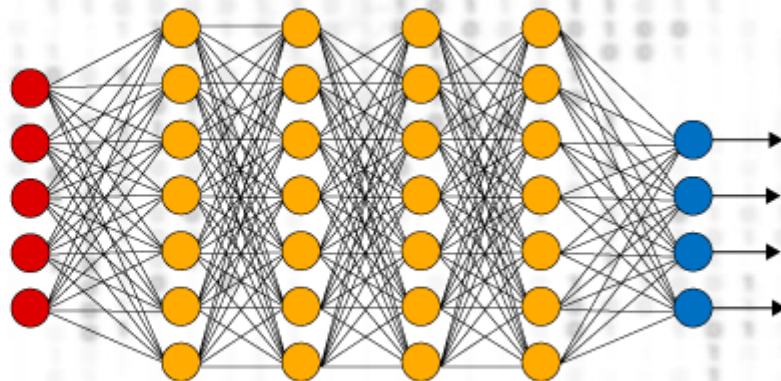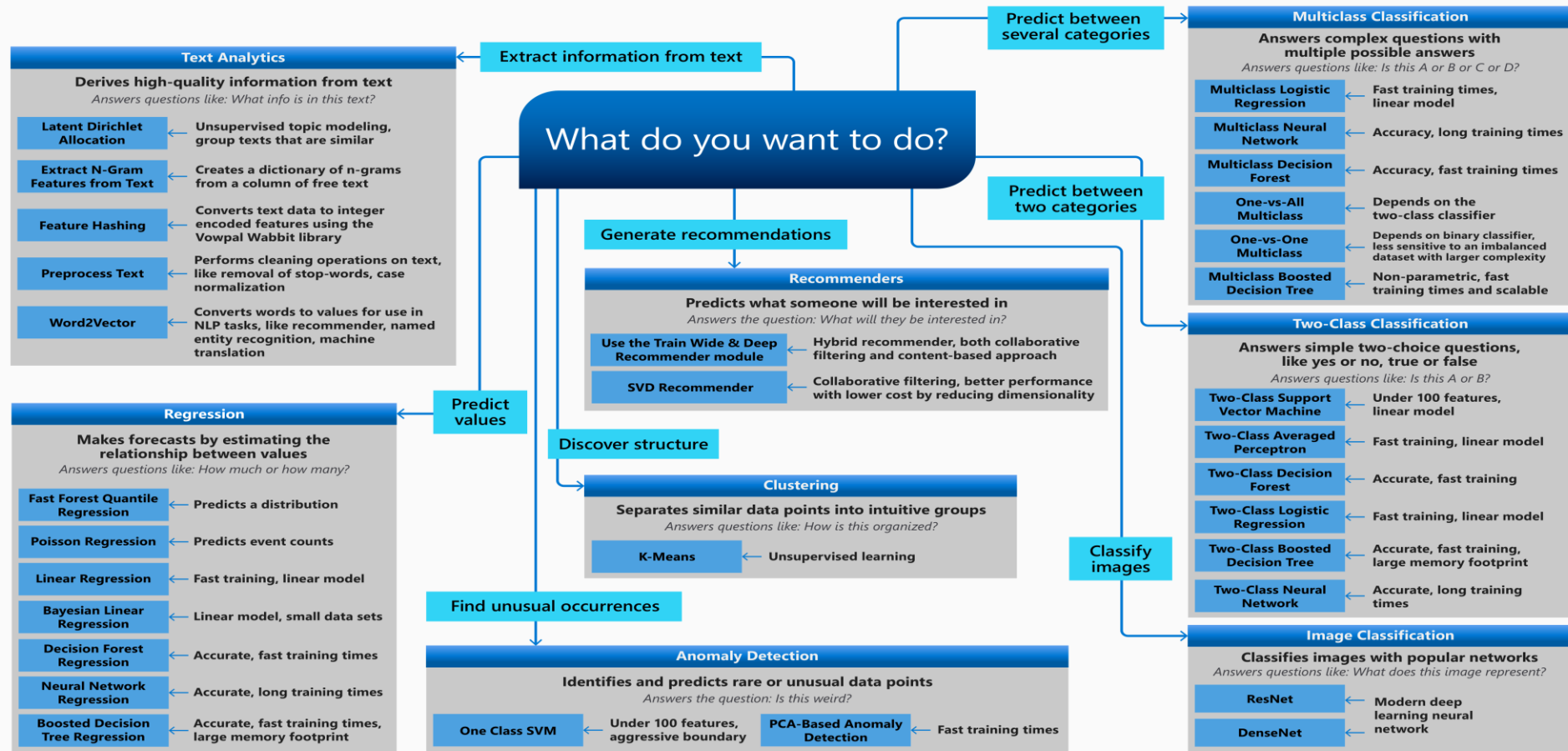- **Two-Class Neural Network** ← Accurate, long training times

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

Classify images

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

Image Classification

Classifies images with popular networks
Answers questions like: What does this image represent?

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

Machines Are Learning, You Should Learn ML Too: An Azure Story *by* André Melancia

techforum technical community

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, DBC, etc.) and you can use your tools like Azure Data Studio, etc.

Magic is in the libraries, not the languages!
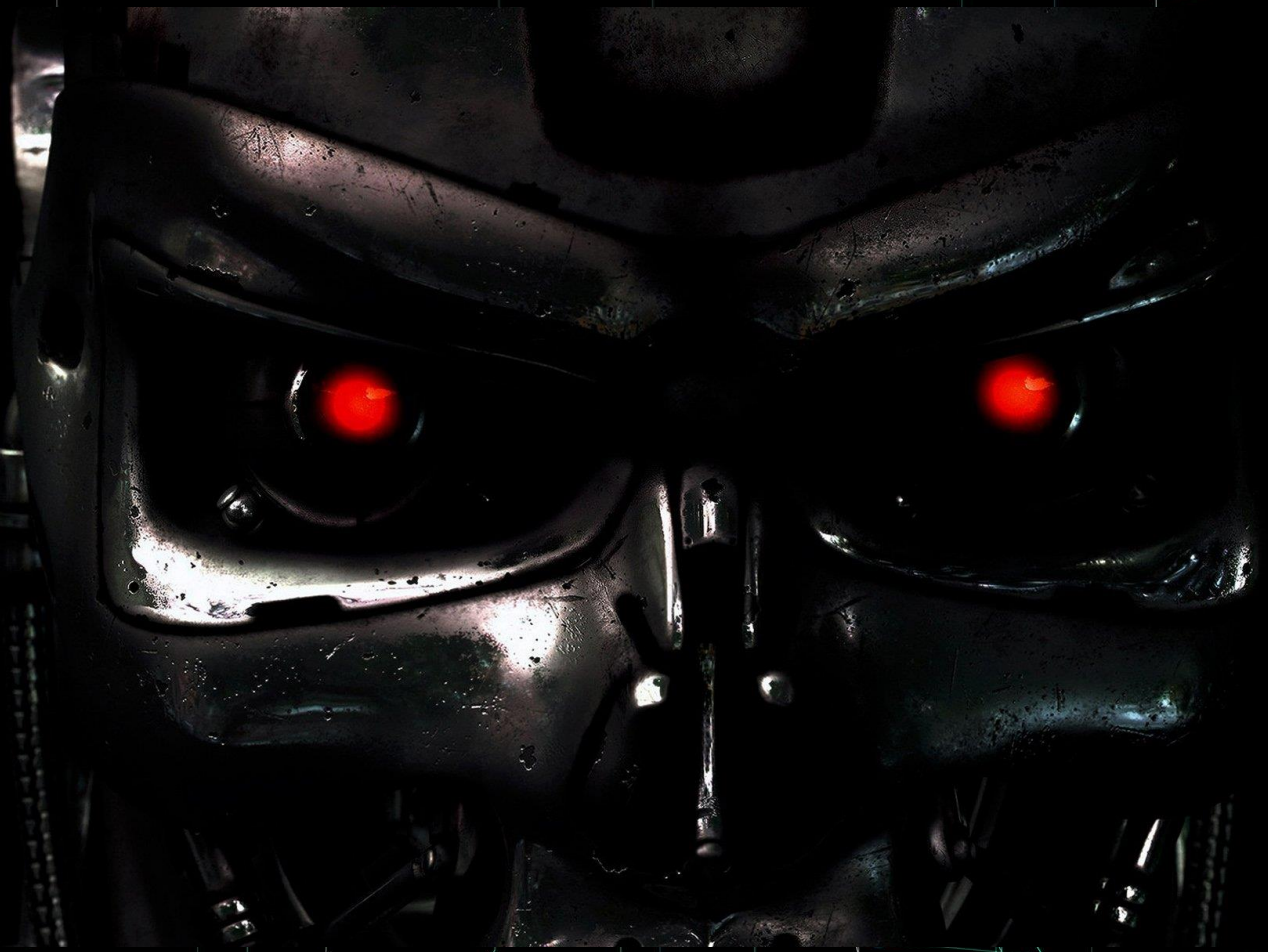
"Come with me if you want to DEMO..."

Questions, Humans?

ONE DOES NOT SIMPLY UNPLUG SKYNET

Machines Are Learning, You Should Learn ML Too: An Azure Story by André Melancia

"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.**"

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

André Melancia
Andy.PT
LunarCat.PT

techforum technical community