INTEL INDUSTRIAL IOT DEVELOPER WEBINAR SERIES

WEBINAR 3:
IOT ANALYTICS & COMPUTER VISION IN INDUSTRY 4.0

The webinar will begin momentarily, please wait a few moments.

If you are not yet a member, sign up to the Intel Industrial IoT Developer Community and make sure you don't miss out on future upcoming events and exclusive information:
https://intel.ly/2IYrK3E

PREDICTIVE ANALYTICS & COMPUTER VISION
INDUSTRIAL IOT

Core and Visual Computing Group
Developer Relations, Intel

LEGAL NOTICES AND DISCLAIMERS

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Any forecasts of goods and services needed for Intel's operations are provided for discussion purposes only. Intel will have no liability to make any purchase in connection with forecasts published in this document.

ARDUINO 101 and the ARDUINO infinity logo are trademarks or registered trademarks of Arduino, LLC.

Intel, the Intel logo, Intel Inside, the Intel Inside logo, OpenVINO, Intel Atom, Celeron, Intel Core, and Intel Movidius Myriad 2 are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

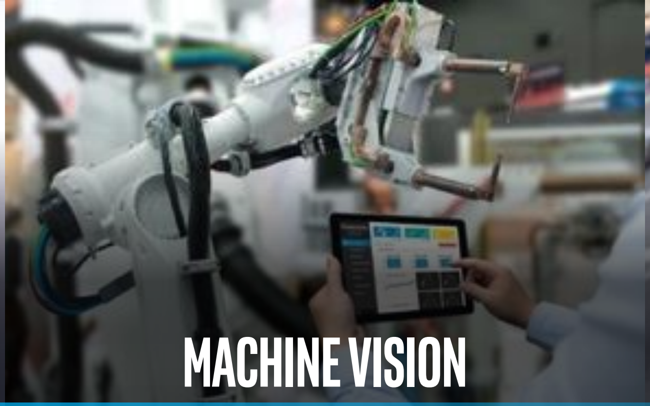
DEFINING PREDICTIVE ANALYTICS

- Predictive analytics is the **science of classifying** new observations into a set of categories on the basis of a training set of data that has observations with known category memberships.

- Predictive analytics is **not about predicting the future**, instead it's about constructing predictive models that compute a numerical score, on the probability that a particular observation will (or will not) occur in the future with an acceptable level of reliability

- Predictive analytics can be used to prepare **what-if scenarios, risk assessment** and determine observations that may be considered to be **outliers**.

- How can Predictive Analytics and Computer Vision be brought together to enable new business and scientific possibilities?



PREDICTIVE MODELING

CLASSIFYING AND LEARNING BUILDING BETTER SOLUTIONS

PREDICTIVE ANALYTICS USE IN INDUSTRIAL MANUFACTURING

Fujitsu Group and Intel jointly developed a solution using sensors and video to reduce the error rate of products manufactured in their business notebook and table factories

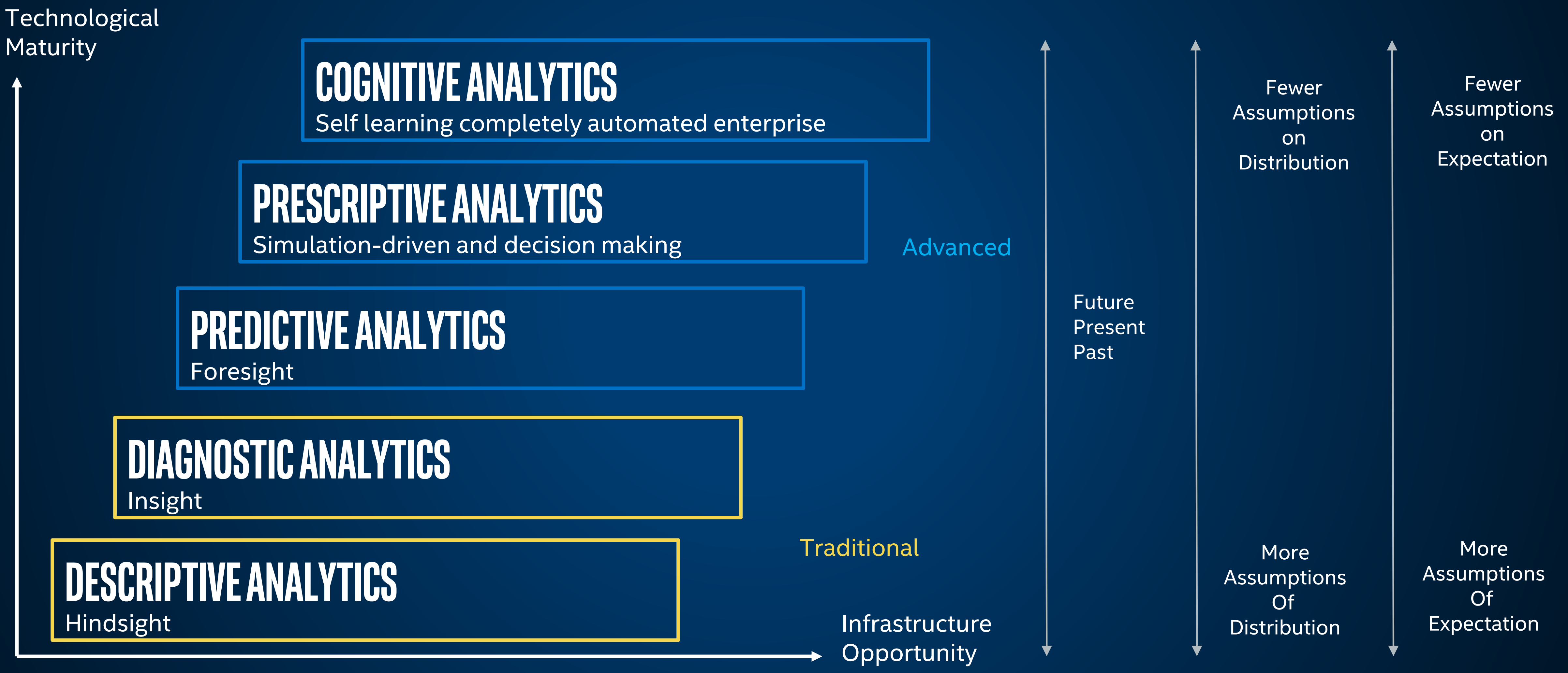
https://www.intel.com/content/dam/www/public/us/en/documents/case-studies/real-time-iot-tracking-manufacturing-case-study.pdf

PREDICTIVE ANALYTICS USED IN HEALTHCARE

Predictive Analytics will allow doctors, hospital administrators, researchers, and insurers to use and understand the relationships in larger data sets.

Predictors and Indicators drive nearly all decision making in this complex industry

ANALYTICS MATURITY CURVE



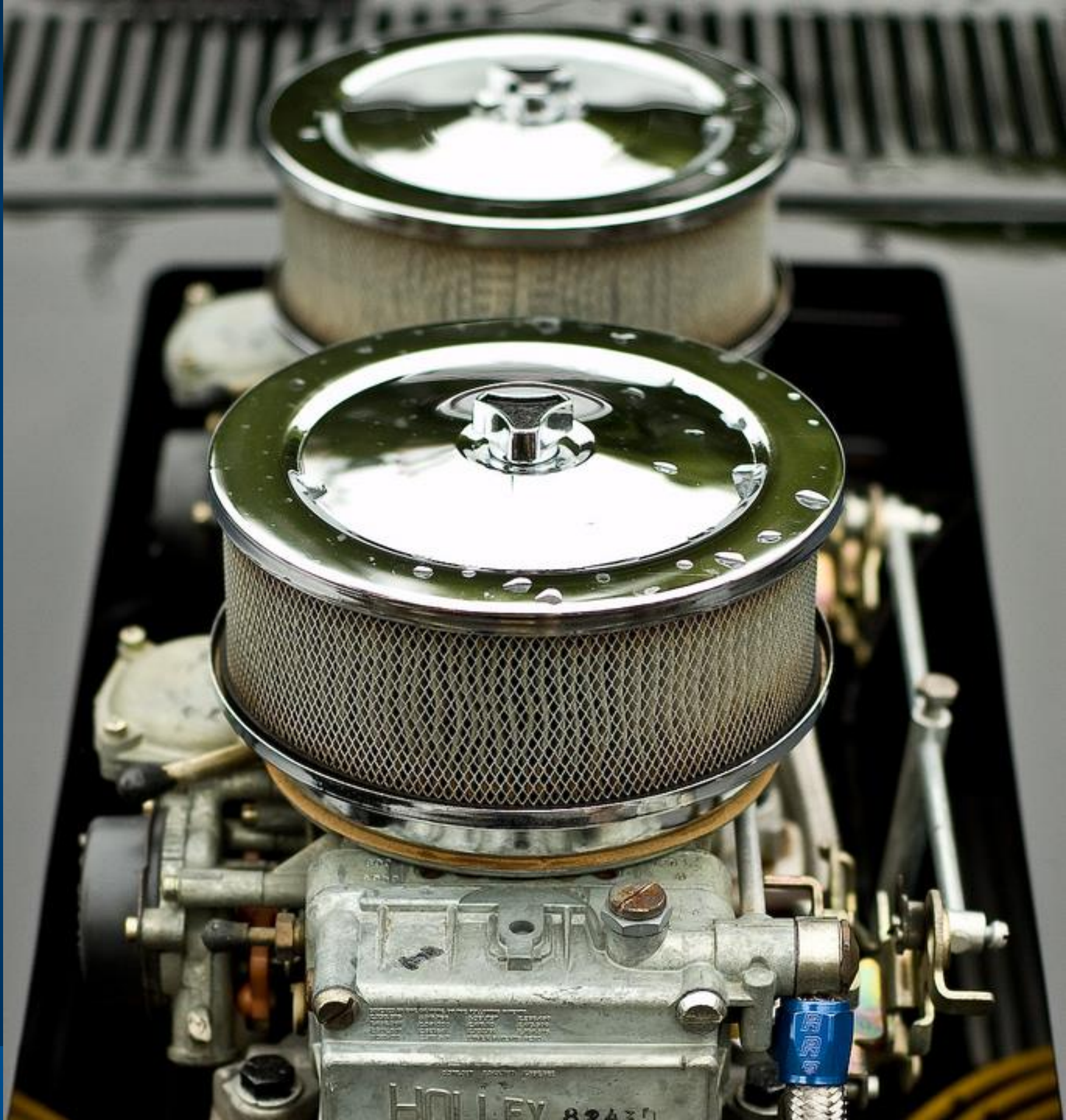
VIBRATION ON A MOTOR

The vibration of mechanical parts within industrial equipment can be analyzed by decomposing their waveform into component waveforms using the **Fast Fourier Transform**.

As the bearing wears the ball begins to slip rather than roll. The waveform then is non-periodic but becomes cyclostationary.

The vibration status is often a predictor of the quality and so is used for estimating the lifespan of the ball bearing.

The FFT can breakdown the waveform and be used to detect when cyclic slippage begins to occur on the bearings.

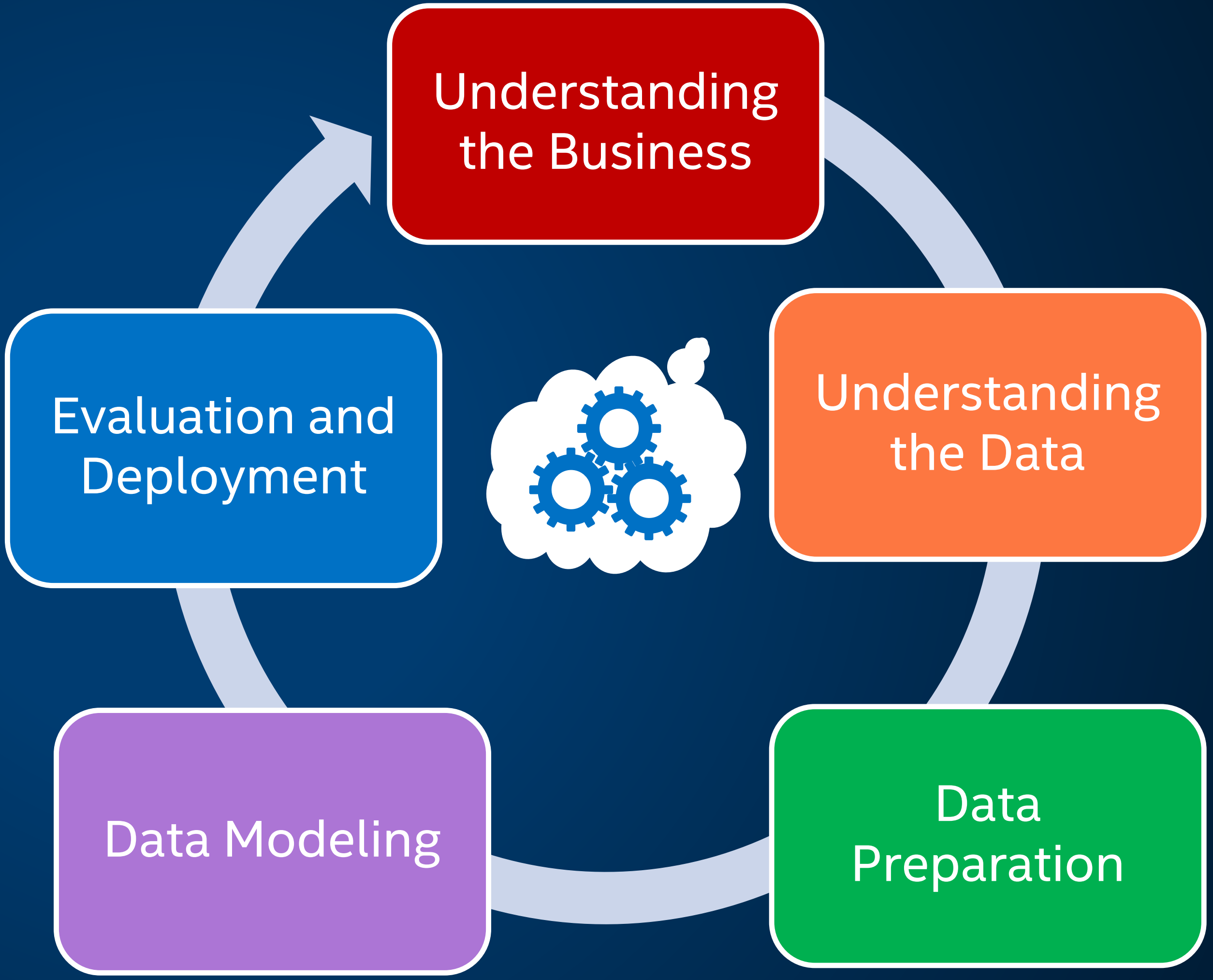


CRISP-DM: AN APPROACH TO UNDERSTANDING DATA

CRISP-DM stands for **Cross-industry standard process for data mining**, and it is a standard based upon a wide poll of industry data scientists.

CRISP-DM break the process of data mining into several distinct steps.

CRISP-DM notes that the data mining is a process of continuous improvement and that is continues even after a solution is deployed



REMINDERS FOR DATA SCIENTISTS

- More data does not necessarily mean more insight

- Increased Insight does not always mean increased value

CLEARLY ACTIONABLE DATA

Representing data presents challenges because the data must be related to high level ideas that are actionable, timely and relevant.

ALGORITHMS IN THE PREDICTIVE MODELING TOOLKIT

WHAT IS A PREDICTIVE MODEL?

A predictive model is **formally** an **function** that **relates** a **set of inputs** (*X*s) to one or more **outcomes** (*Y*s) by **separating** the response variation into **signal** and **noise**.

Models are often divided into **parametric** and **non-parametric**.

Parametric models assumes the **distribution** of the function is **defined** by a bounded, or **finite set of parameters** (Θ). $P(x|\Theta, D) = P(x|\Theta)$

Non-parametric models assume that the data **distribution** is **not defined** in terms of such a finite set of parameters. They do But they can often be defined by assuming an infinite dimensional, Θ.

Wikipedia has a list of some types of models

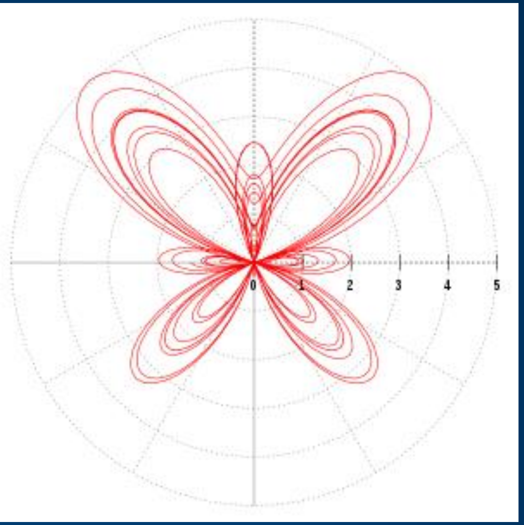
- Group method of data handling
- Naive Bayes
- *k*-nearest neighbor algorithm
- Majority classifier
- Support vector machines
- Random forests
- Boosted trees
- CART (Classification and Regression Trees)
- MARS
- Neural Networks
- ACE and AVAS
- Ordinary Least Squares
- Generalized Linear Models (GLM)
- Logistic regression
- Generalized additive models
- Robust regression
- Semiparametric regression

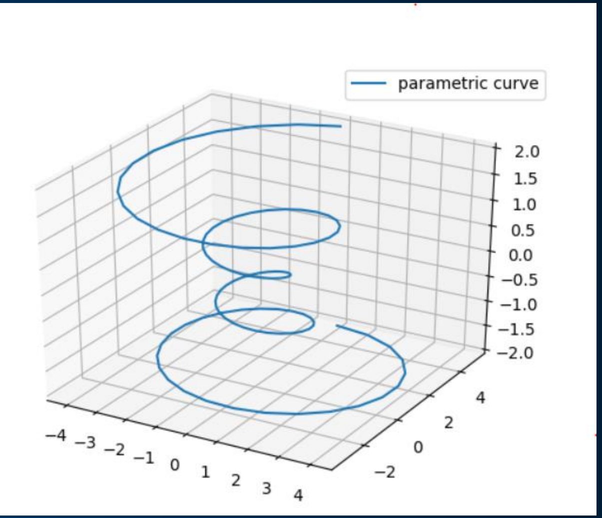
PARAMETRIC VS NON-PARAMETRIC

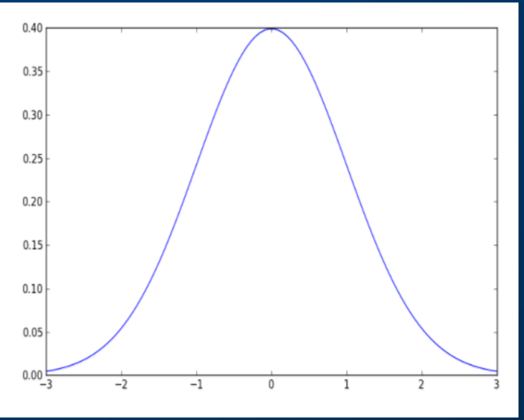
- **Parametric models** - assumes the **distribution** of the function is **defined** by a bounded, or **finite set of parameters** (Θ). P(x|Θ, D) = P(x|Θ)

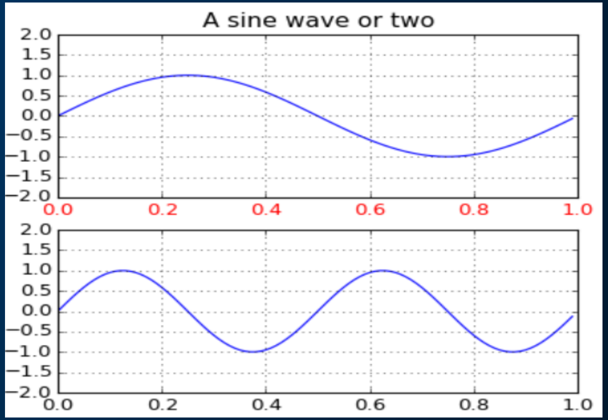
- **Non-parametric models** - assume that the data **distribution** is **not defined** in terms of such a finite set of parameters. They do But they can often be defined by assuming an infinite dimensional, Θ.

- LOWESS (locally weighted scatterplot smoothing)









THREE STYLES OF MACHINE LEARNING

Supervised Learning

Supervised Learning algorithms have an outcome that is predicted from a given set of predictors. A function is generated that map inputs to desired outputs and training process continues until the model achieves a desired level of accuracy on the training data.

Unsupervised Learning

In this model there is no labelled target data. The algorithm clusters the data around pre-existing or attributes that are defined during the training.

Reinforcement Learning

Training is done to learn specific target outputs and then algorithm is exposed to an environment where it trains itself continually using trial and error. The algorithm evaluates its performance and modifies its behavior to make accurate decisions.

REGRESSION ALGORITHMS

Regression analysis is a statistical processes for estimating the relationships between a dependent variable and one or more independent variables (or 'predictors').

Given a training set:
$Z - \{\{x_1, y_1\}, \ldots, \{x_n, y_n\}\}$ where xi are features and yi are targets.

The goal is to find f(x) using training so that:

$$\min \sum_{(x,y)} ((f(x) - y)^2$$

Find the function that minimizes the square of the error.

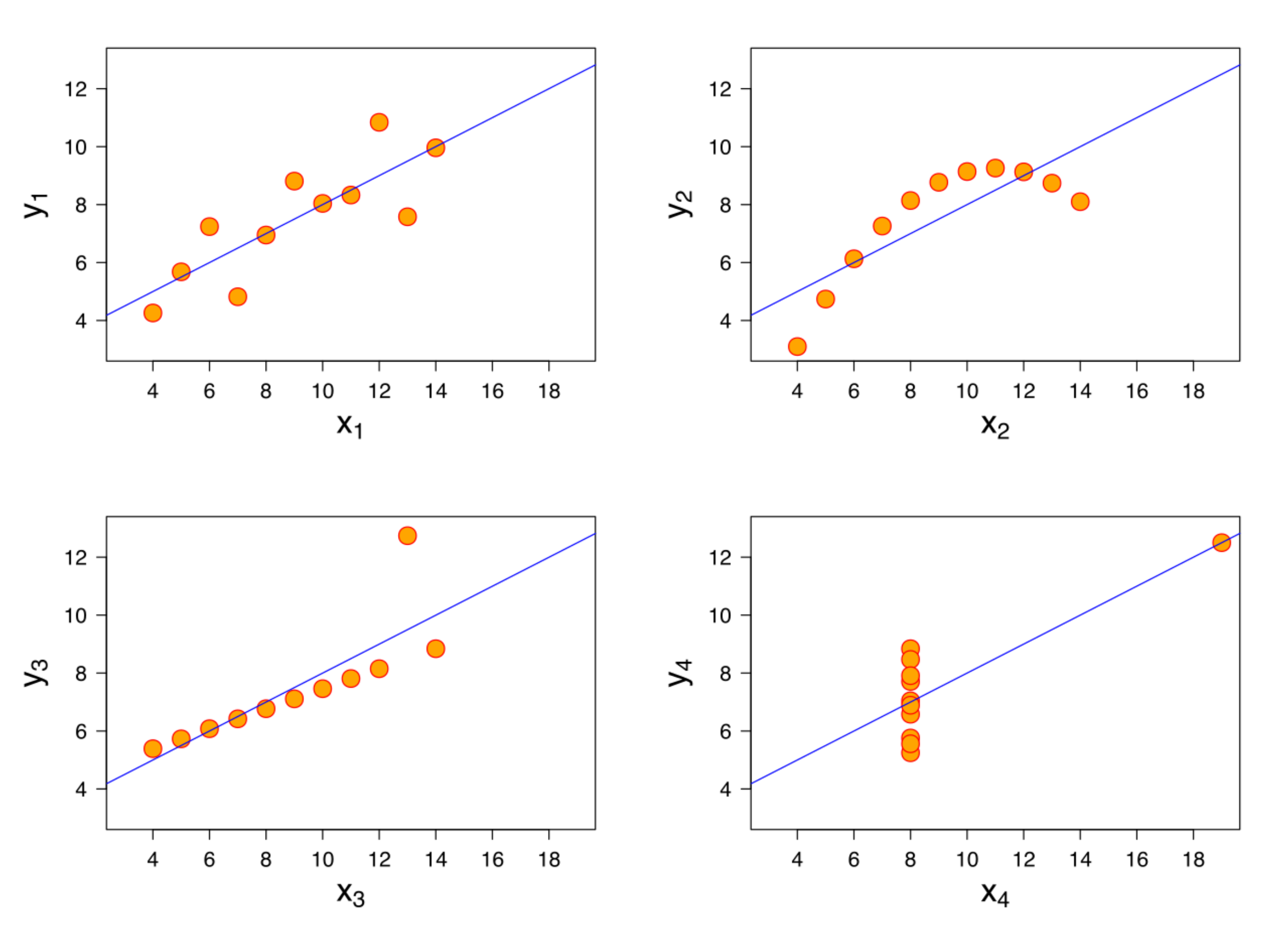


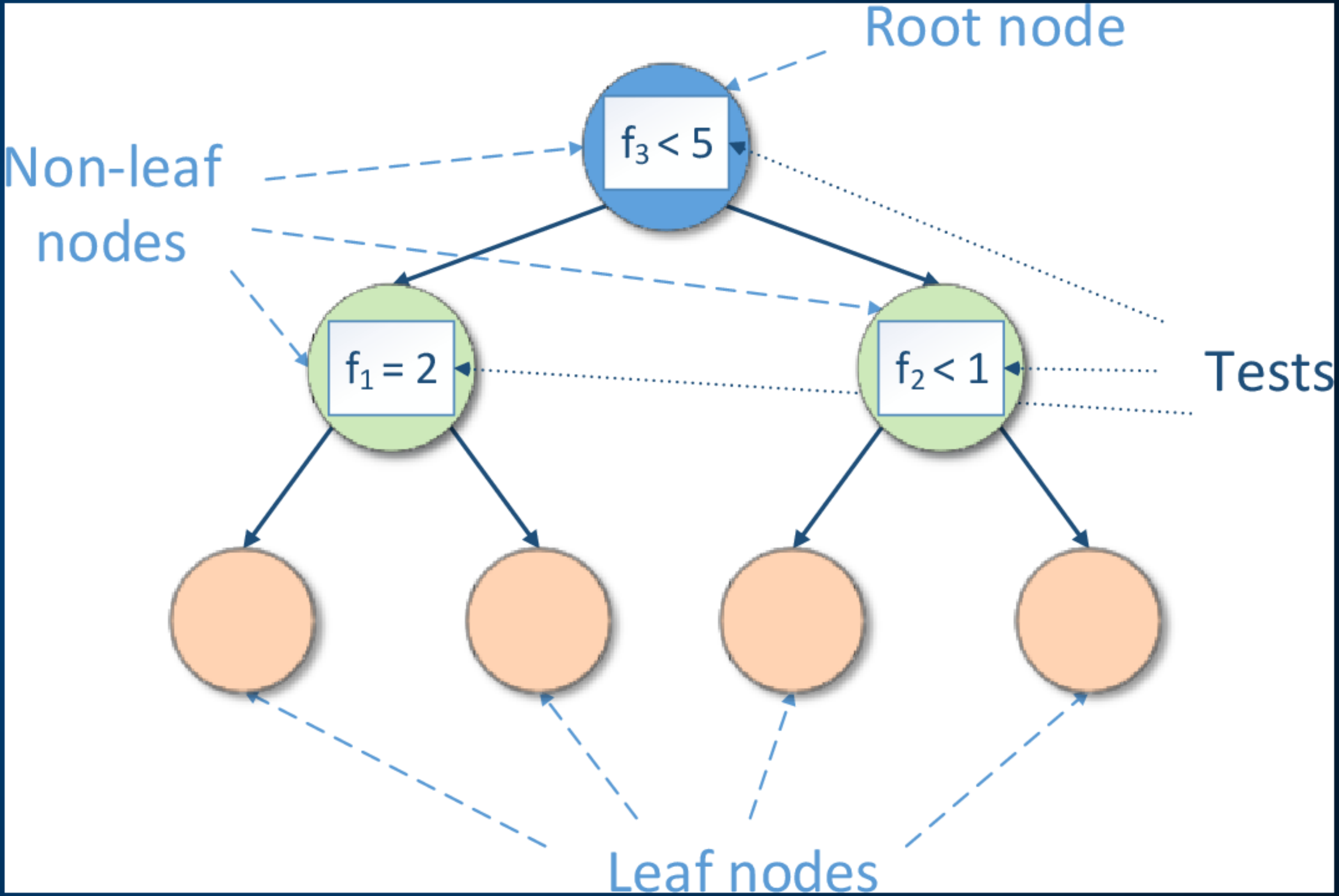
Image from https://en.wikipedia.org/wiki/Anscombe%27s_quartet [Creative Commons Attribution-ShareAlike License]

DECISION TREE LEARNING

The goal of decision tree learning is to construct a hierarchical set of rules that classify data into discreet labeled classes

It is a type of supervised learning algorithm that is mostly used for classification problems.

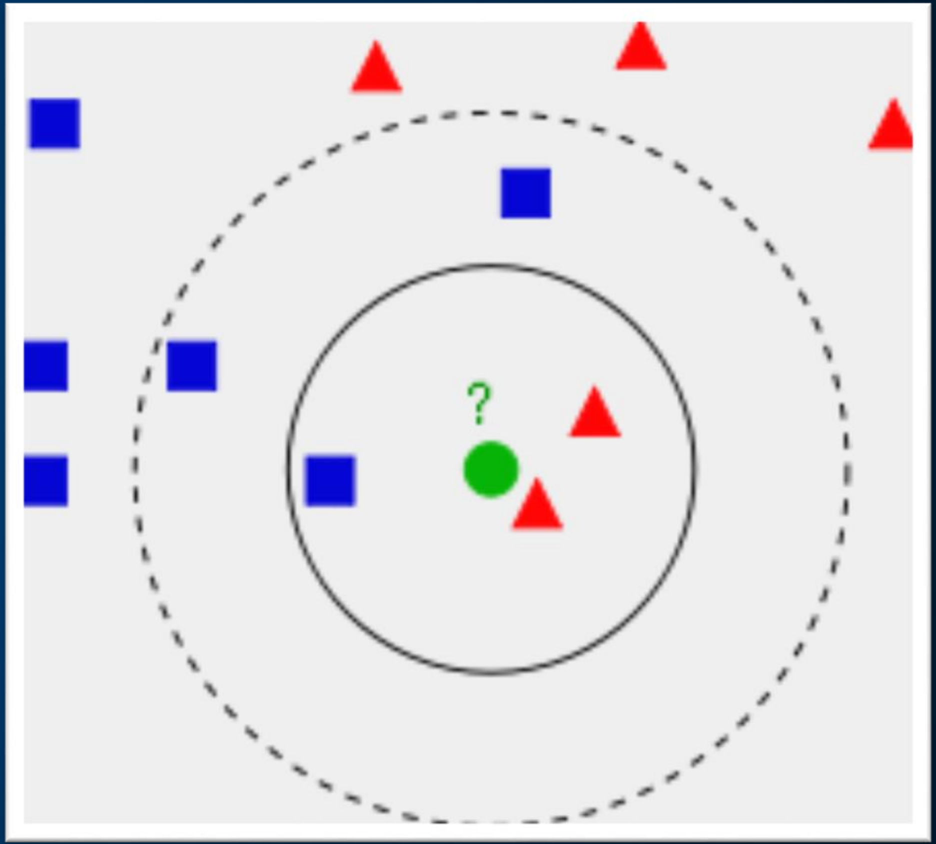
Often used in rule base systems and it works well on category based and continuously dependent variables



K - NEAREST NEIGHBOR ALGORITHM

In *k-NN classification*, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor.

K is the number of neighbors to consider when labelling.



NAÏVE BAYES

A Naive Bayes classifier uses the conditional probability assumes that the predictors used to classify incoming data points are independent of each other.

Performs well when the dimensionality of the inputs is high

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid \mathbf{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Humidity	Temperature	Wind Speed	Weather
Humid	Hot	Fast	Sunny
Humid	Hot	Fast	Sunny
Humid	Hot	Slow	Sunny
Not Humid	Cold	Fast	Sunny
Not Humid	Hot	Slow	Rainy
Not Humid	Cold	Fast	Rainy
Humid	Hot	Slow	Rainy
Humid	Cold	Slow	Rainy

Humidity %	Temp (C)	Wind	Weather
Humid	Cold	Fast	?

Recommendation System: Naive Bayes classifiers are used in various inferencing systems for making certain recommendations to users out of a list of possible options.
Real-Time Prediction: Naive Bayes is a fast algorithm, which makes it an ideal fit for making predictions in real time.
Multiclass Prediction: This algorithm is also well-known for its multiclass prediction feature. Here, we can predict the probability of multiple classes of the target variable.
Sentiment Analysis: Naive Bayes is used in sentiment analysis on social networking datasets like Twitter* and Facebook* to identify positive and negative customer sentiments.
Text Classification: Naive Bayes classifiers are frequently used in text classification and provide a high success rate, as compared to other algorithms.
Spam Filtering: Naive Bayes is widely used in spam filtering for identifying spam email.

GRADIENT BOOSTING ALGORITHMS

Boosting algorithms are used to combine a series of weak classifiers into a more robust classifer model.

Input: training set $\{\{x_1, y_1\}, \ldots, \{x_n, y_n\}\}$

M: The number of iterations

1. $f_0(x) = \frac{1}{n}\sum_{i=1}^{n} y_i$
2. Iterate loop m=1 to M
3. $y1_i = y_i - f_{m-1}(x_i)$ # Calculate the residual
4. Fit a decision tree $h_m(x)$ to the targets
5. $fm(x) = f_{m-1}(x) + h_m(x)$
6. Return $f_m(x)$

DIMENSIONALITY REDUCTION ALGORITHMS

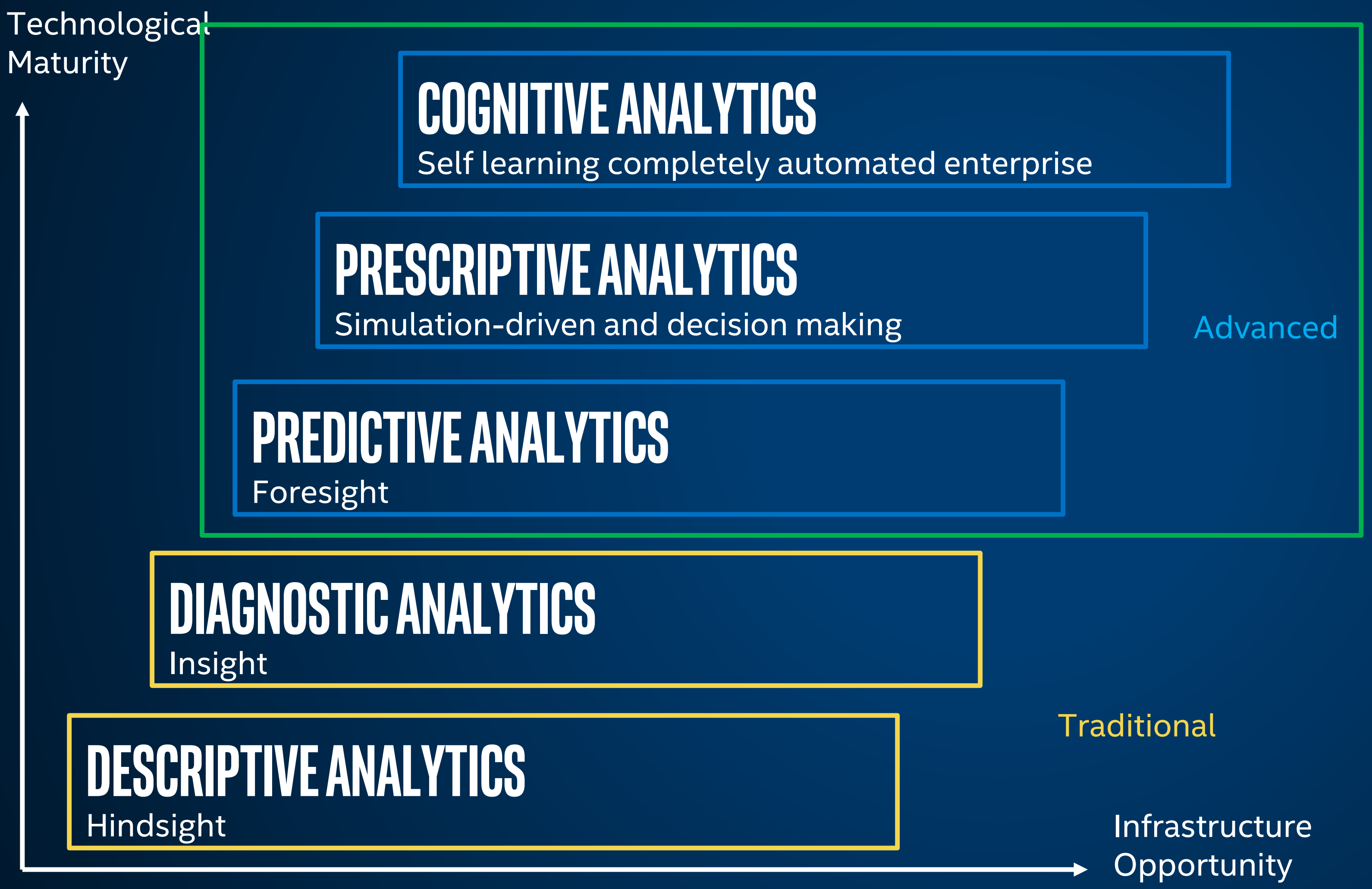
Benefits of Reducing the Number of Dimension or Features in a Data Set

- Removal of multi-collinearity improves the performance of the machine learning model.
- Reduction of time and storage space requirements
- Improves the ability to create datasets from existing data
- It becomes easier to visualize the data when reduced to very low dimensions such as 2D or 3D.

Methods of Reducing

- Principal component analysis (PCA)
- Low Variance
- High Correlation

ANALYTICS MATURITY CURVE: NEURAL NETWORKS



Neural Networks are increasing used higher on the Analytics Maturity Curve

COMPUTER VISION AT THE EDGE

- Sensing and interacting in real-time with the world around you.

- Becoming increasingly mobile

- Computing at the edge of the network enables new
- possibilities

MULTIPLE USAGES

CAPTURE

- Acquire Data from Sensors
- Imaging pipeline
- Initial processing
- Encoding

"MEDIA"

- Decoding
- Aggregation
- Muxing

COMPUTER-VISION

- "Traditional' Visual Understanding
- Pre-Processing for Deep Learning

DEEP LEARNING

- Inference using Deep Learning models
- Different models, tasks, for various models/topologies

"OUTPUT"

- Generate Insights
- Render to Screen
- Alert
- Store

CLEAR TRENDS

CAPTURE
- Acquire Data from Sensors
- Imaging pipeline
- Initial processing
- Encoding

"MEDIA"
- Decoding
- Aggregation
- Muxing

COMPUTER-VISION
- "Traditional' Visual Understanding
- Pre-Processing for Deep Learning

DEEP LEARNING
- Inference using Deep Learning models
- Different models, tasks, for various models/topologies

"OUTPUT"
- Generate Insights
- Render to Screen
- Alert
- Store

LESS STORAGE REQUIRED →

← FASTER RESPOND TIME, MORE CONTROLLABILITY ON THE EDGE

LESS BAND-WIDTH →

← MORE ANALYTICS TO THE EDGE

APPLICATIONS OF COMPUTER VISION

- Stitching/Registration
- Filtering
- Thresholding
- Quality Assurance
- Segmentation
- Edge detection
- Color Analysis
- Blob detection and extraction
- Pattern recognition
- Barcode and QRC reading
- Optical character recognition
- Gauging/Metrology

List of algorithms is from https://en.wikipedia.org/wiki/Machine_vision

COMPUTER VISION ALGORITHMS

Feature Extraction
- Active Contour Modeling
- Blob detection
- Canny edge detector

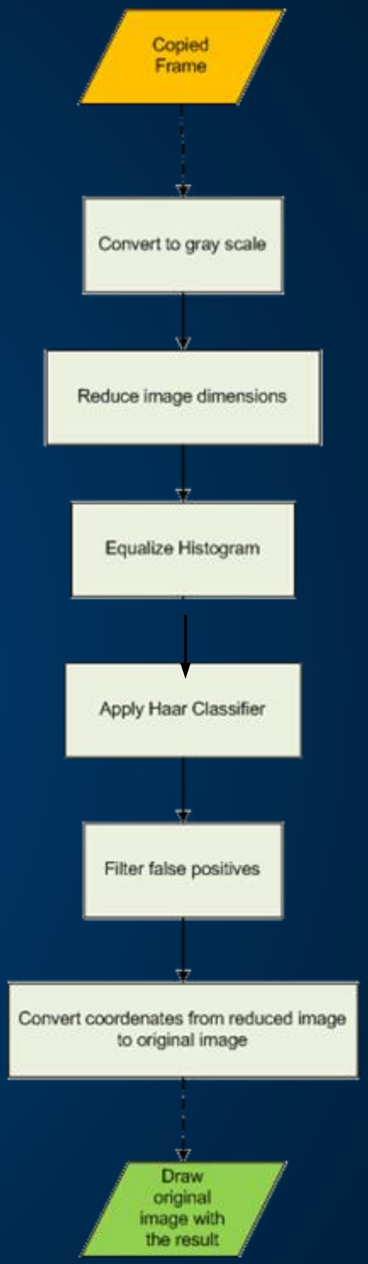
Visual Recognition
- Object Recognition
- Scale-invariant feature transform (SIFT)
- Gesture recognition
- Viola-Jones Algorithm
- Kalman Filters

Edge and Corner Detection
- Canny Edge Detector
- Harris Corner Detector
- Sobel

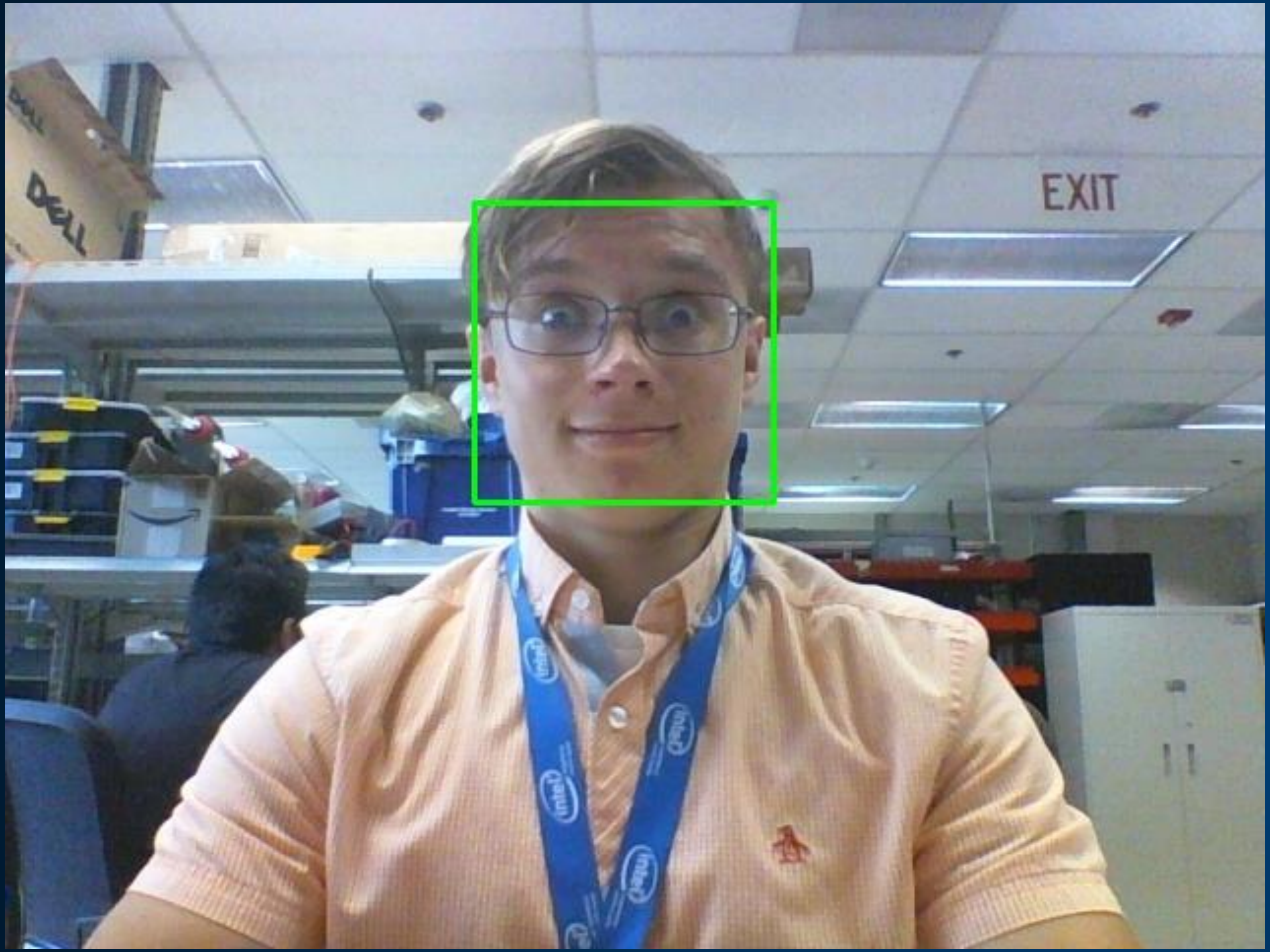
OPENCV* LIBRARY

```
import cv2
import numpy as np
```

- Open Source library for computer vision
- Written in C++
- Bindings for most popular languages
- Block-like programming structure
- Image data is sequentially passed through functions
- Fundamental concepts allow for powerful image processing tools
- https://opencv.org/



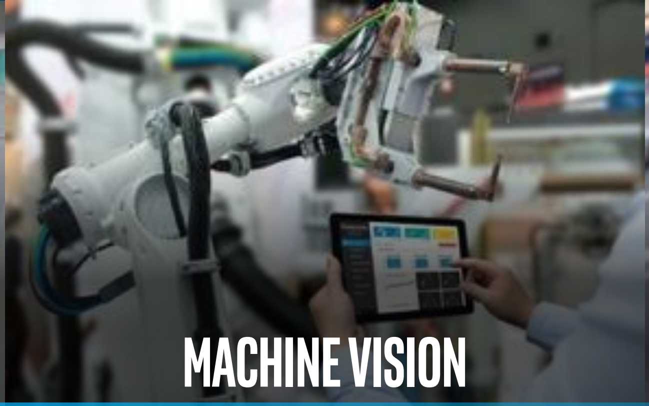
HAAR CASCADES IN OPENCV*




```
gray_img = cv2.cvtColor( img,
cv2.COLOR_BGR2GRAY )
faces = cascade.detectMultiScale(
            gray_img,
            scaleFactor = 1.25,
            minNeighbors = 5,
            minSize = ( 30, 30 ),
)

for (x, y, w, h) in faces:
            cv2.rectangle( img, x, y, (x+w, y+h),
(0, 255, 0), 2)
```

```
cascade = cv2.CascadeClassifier( 'haarcascade_frontalface_default.xml' )
```



VIDEO: THE "EYE OF IOT"

USE OF VIDEO, COMPUTER VISION AND DEEP LEARNING IS GROWING RAPIDLY

ARTIFICIAL INTELLIGENCE

is playing a large role in emerging predictive systems

PREDICTIVE ANALYTICS AND COMPUTER VISION NEED COMPUTE AT THE EDGE

Motivating Example: Alice (2004-2007)

Alice
- 300+ miles of fully autonomous driving
- 8 cameras, 8 LADAR, 2 RADAR
- 12 Core 2 Duo CPUs + Quad Core
- 3 Gb/s data network
- ~75 person team over 18 months (x 2)

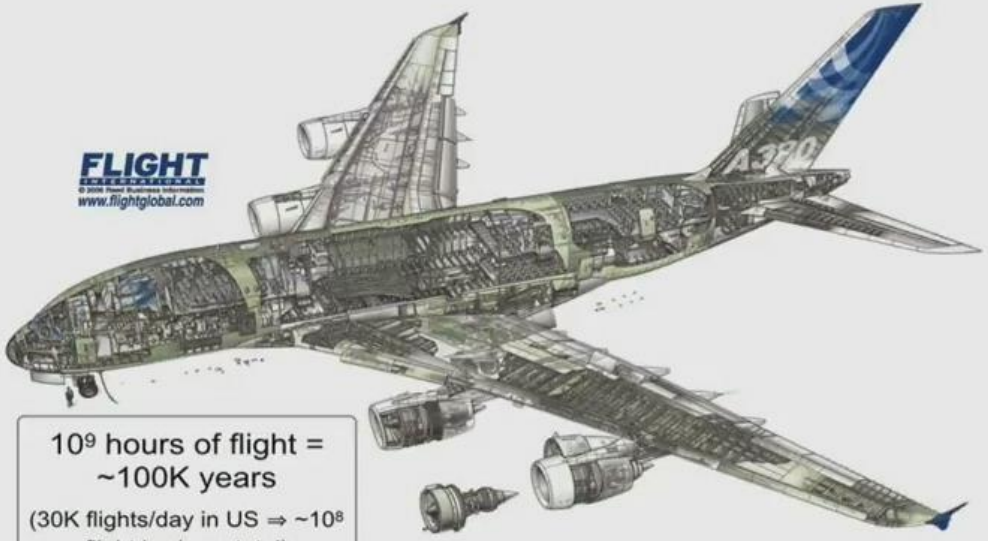
Software
- 25 programs with ~200 exec threads
- 237,467 lines of executable code

Safety Critical Autonomous Systems

Question: How safe do autonomous vehicles need to be?
- As safe as human-driven cars (7 deaths every 10^9 miles)
- As safe as buses and trains (0.1-0.4 deaths every 10^9 miles)
- As safe as airplanes (0.07 deaths every 10^9 miles)

I. Savage, "Comparing the fatality risks in United States transportation across modes and over time", *Research in Transportation Economics*, 43:9-22, 2013.

Safety-Critical Systems: Commercial Air



10^9 hours of flight = ~100K years

(30K flights/day in US ⇒ ~10^8 flight hrs/year total)

Hazard Classification	Development Assurance Level	Maximum Probability per Flight Hour
Catastrophic	A	10^{-9}
Hazardous	B	10^{-7}
Major	C	10^{-5}
Minor	D	–
No Effect	E	–

- 3.22 trillion miles (US, 2016)
- 40,200 fatalities (US, 2016) – roughly 100 people each day
- 1 fatality per 80 million miles
- 1 in 625 chance of dying in car crash (in your lifetime)


LOCAL AND DISTRIBUTED COMPUTATIONAL POWER (COMPUTE AT THE EDGE) FUELED BY INTEL PROCESSORS ALLOW NEW APPLICATIONS AND NEW POSSIBILITIES

THE NEW INTEL® XEON® SCALABLE PROCESSOR IS OPTIMIZED FOR ADVANCED ANALYTICS

The latest generation Intel® Xeon® Scalable processor is workload-optimized to support high-demand applications and designed to drive actionable insight.

Discover the new Intel® Xeon® Scalable processor ›

Learn about Advanced Analytics using the Intel® Xeon® Scalable processor › intel®

INDUSTRIES BENEFITING FROM PREDICTIVE ANALYTICS

INTRODUCTION

1. Introduction to Intel and the IIoT
2. Formalized Structure to IIoT

Each Module contains a lecture and a hands-on lab exercise that builds towards an model of an IIoT infrastructure based on a formalized architecture.

CONTROL

3. Physical Sensors and Actuators
4. Communications and Protocols

OPERATIONS

5. Automated Control Systems
6. Security and IIoT

INFORMATION

7. Smart Video Systems
8. Machine Learning

APPLICATION

9. Predictive Analytics
10. Business Analytics

IMPORTANCE OF DOMAIN KNOWLEDGE

Depending on the available data and feature types, the performance of your predictive model can vacillate dramatically. Therefore, selecting the right features is one of the most important steps before the inferencing takes place. This is called feature engineering, which can be defined as follows:

TIP

Feature engineering

In this process, domain knowledge about the data is used to create only selective or useful features that help prepare the feature vectors to be used so that a machine learning algorithm works.

STATISTICS, PROBABILITY, AND INFORMATION THEORY FOR PREDICTIVE MODELING

random sampling, hypothesis testing, chi-square test, correlation, expectation, variance, covariance and Bayes' rule

EXPECTATION

The expectation of a discrete random variable is the weighted average of all values, where each value is weighted by its probability of being selected. The expected value of a random variable X is given by:

CENTRAL LIMIT THEOREM

The **central limit theorem (CLT)** is one of the fundamental theorems in statistics. It states that the average of all samples of a sufficiently large sample size with independent and identically distributed variables is approximately equal to the mean of the sample space, regardless of the size of the distribution with a finite variance level.