Introduction to AI/Machine Learning

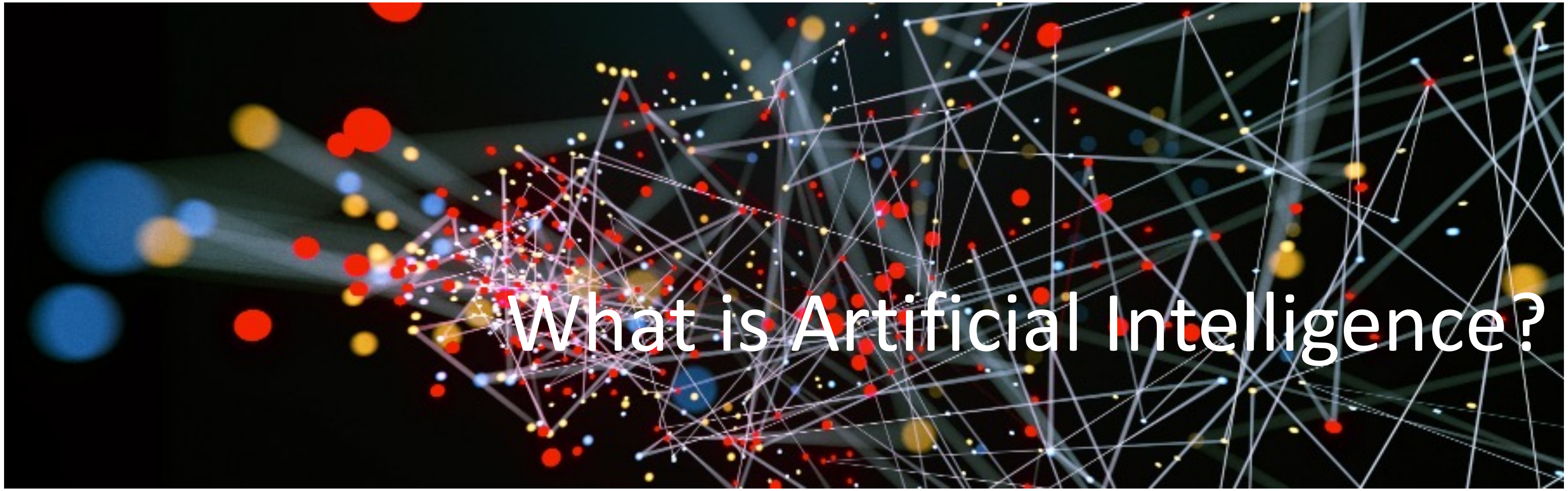
Ishwar K Sethi

Dept. Computer Science & Engineering

Oakland University

Agenda

- What is AI/Machine Learning?
- Machine learning terminology
- Overview of machine learning methods
- Machine learning to deep learning
- Summary and Q & A



iksinc@yahoo.com

What is AI?

- The term *artificial intelligence* was coined by John McCarthy circa 1956. He defined it as “the science and engineering of making intelligent machines”

Artificial intelligence is technology that appears to emulate human performance typically by learning, coming to its own conclusions, appearing to understand complex content, engaging in natural dialogs with people, enhancing human cognitive performance (also known as cognitive computing) or replacing people on execution of non-routine tasks.

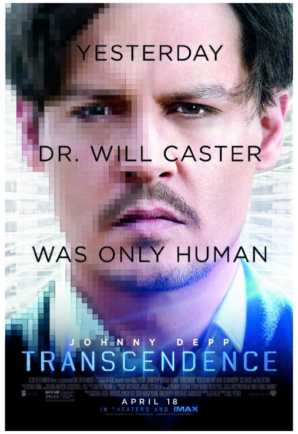
Gartner Definition

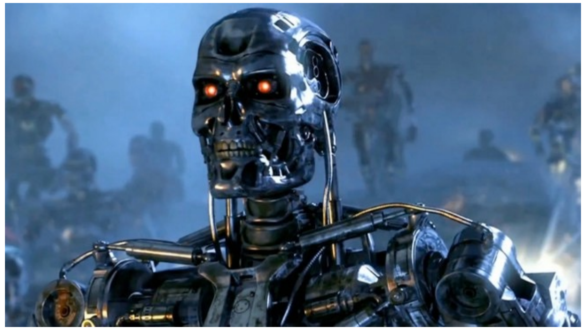
Weak AI

- Also known as Narrow AI
  - a descriptive term used for AI that can demonstrate human like intelligence, but only for a specific task or tasks. Majority of today's AI systems fall in this category.

Artificial General Intelligence (AGI)

- Also known as Strong AI
  - a term used to describe a certain mind-set of artificial intelligence development. Strong AI's goal is to develop artificial intelligence to the point where the machine's intellectual capability is functionally equal to humans.





Artificial Super Intelligence (ASI)

- A term used for AI of the future. It will be be superior to any level of human intelligence and will (potentially), if allowed, be in complete control of its own decision making.

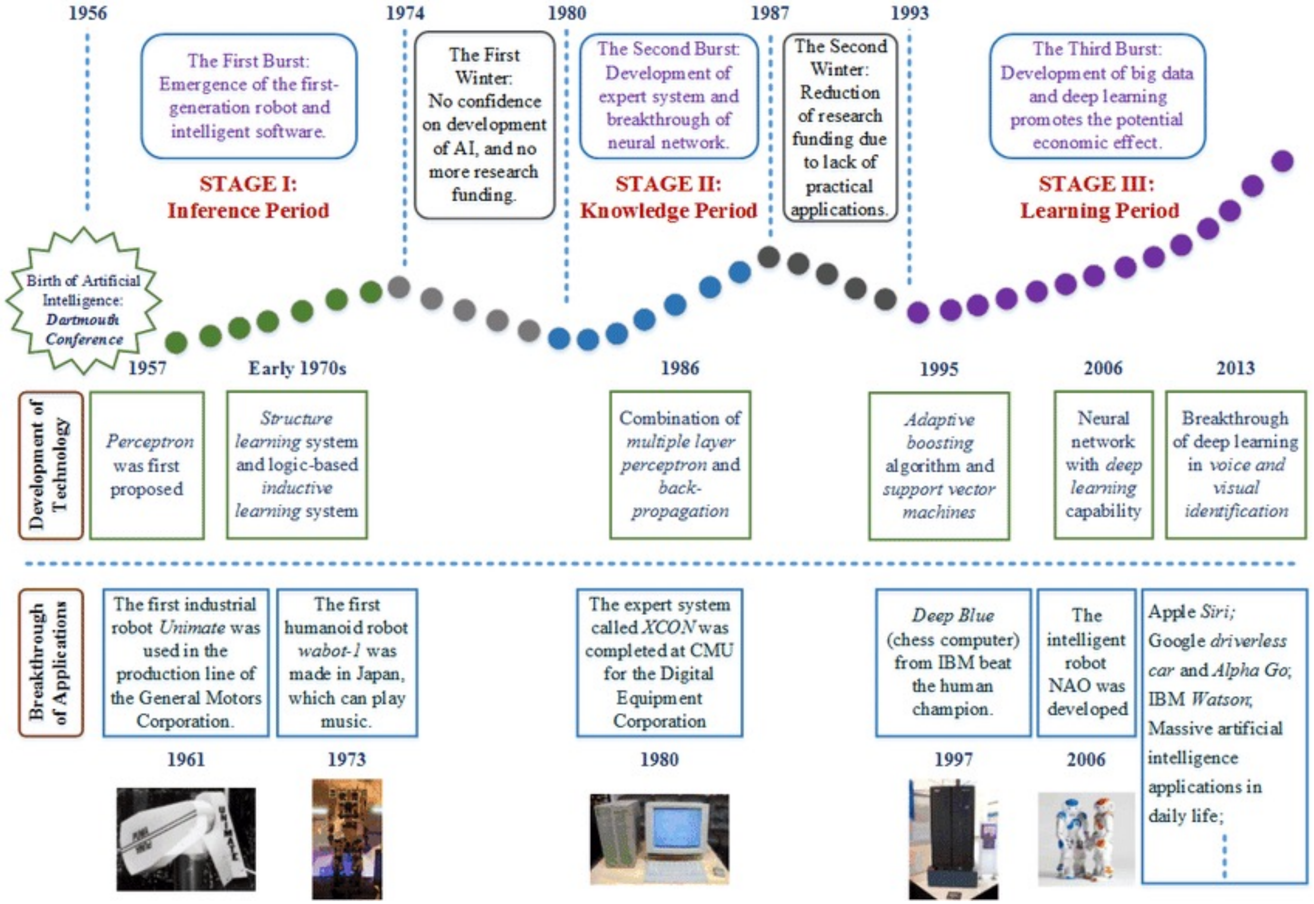
> "It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers... They would be able to converse with each other to sharpen their wits. At some stage therefore, we should have to expect the machines to take control."

Alan Turing, the 'godfather of AI'

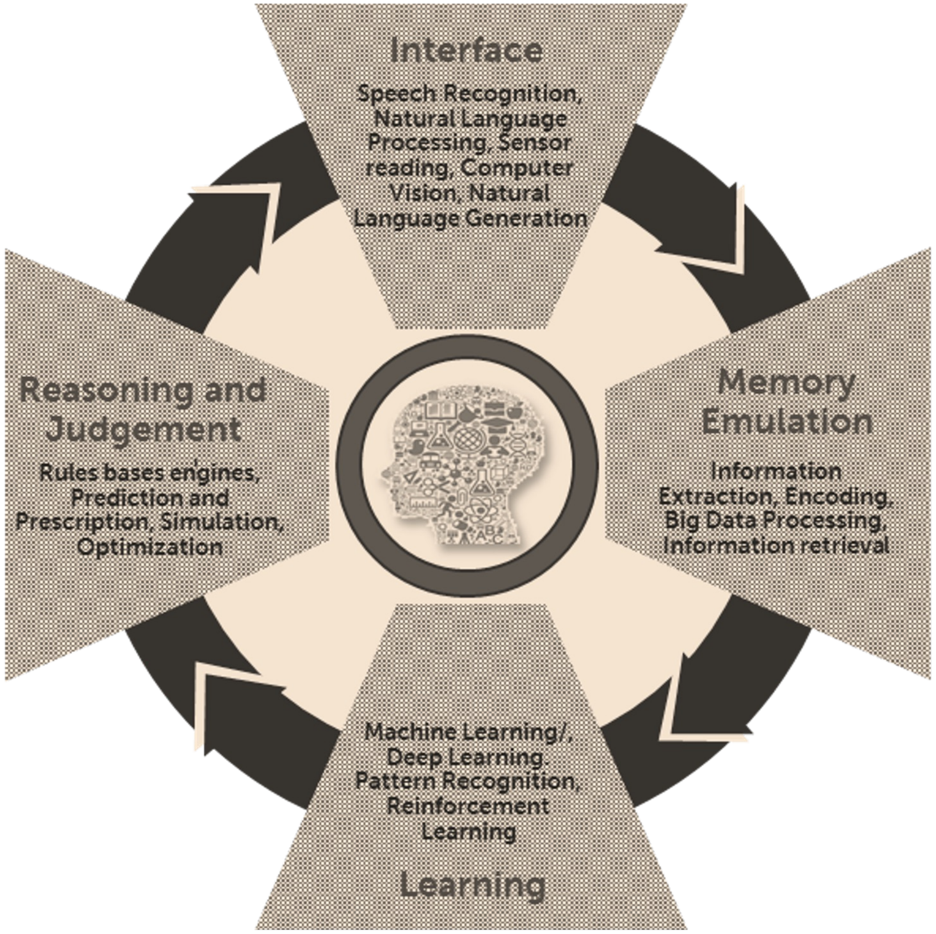
> "First the machines will do a lot of jobs for us and not be super intelligent. A few decades after that though, the intelligence is strong enough to be a concern. I agree with Elon Musk and some others on this and don't understand why some people are not concerned"

Bill Gates
Co-founder, Microsoft

Evolution of AI Since 1950



AI Technologies

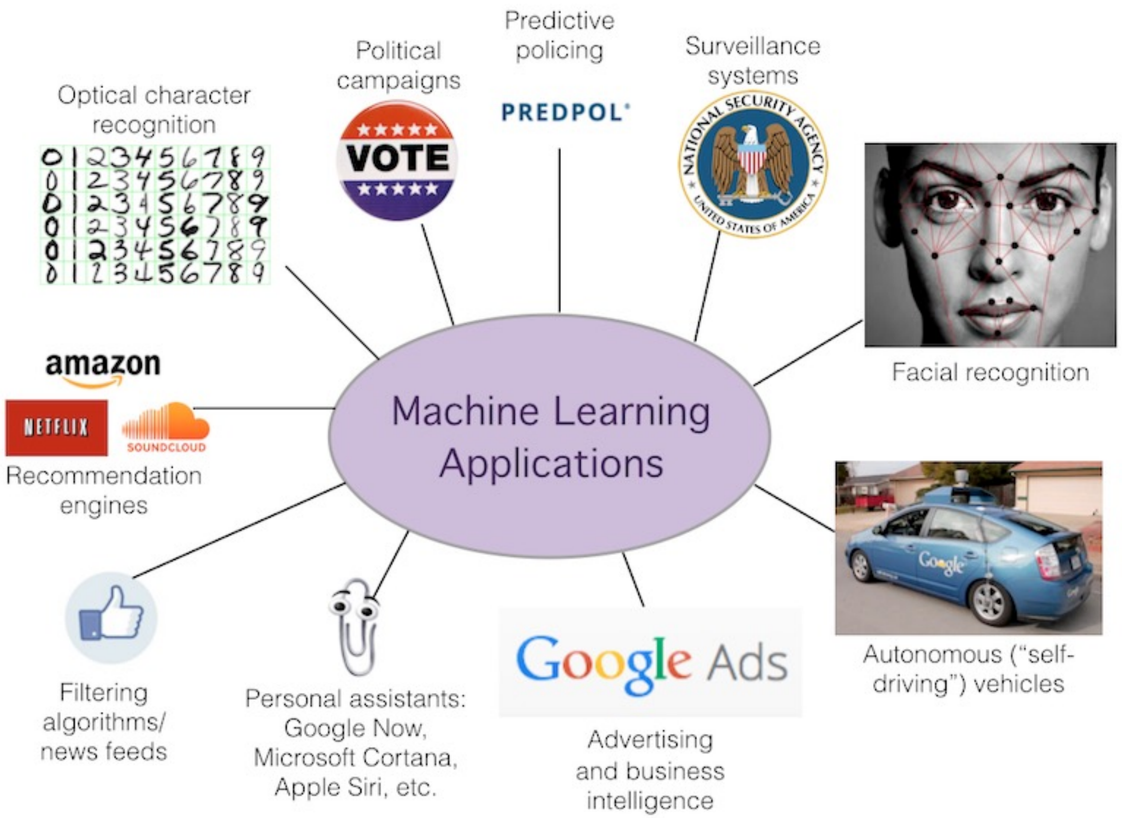


What is Machine Learning?

- Machine learning deals with making computers learn to make predictions/decisions without explicitly programming them. Rather a large number of examples of the underlying task are shown to optimize a performance criterion to achieve learning.

http://dilbert.com/strips/comic/2013-02-02/

Why Machine Learning?



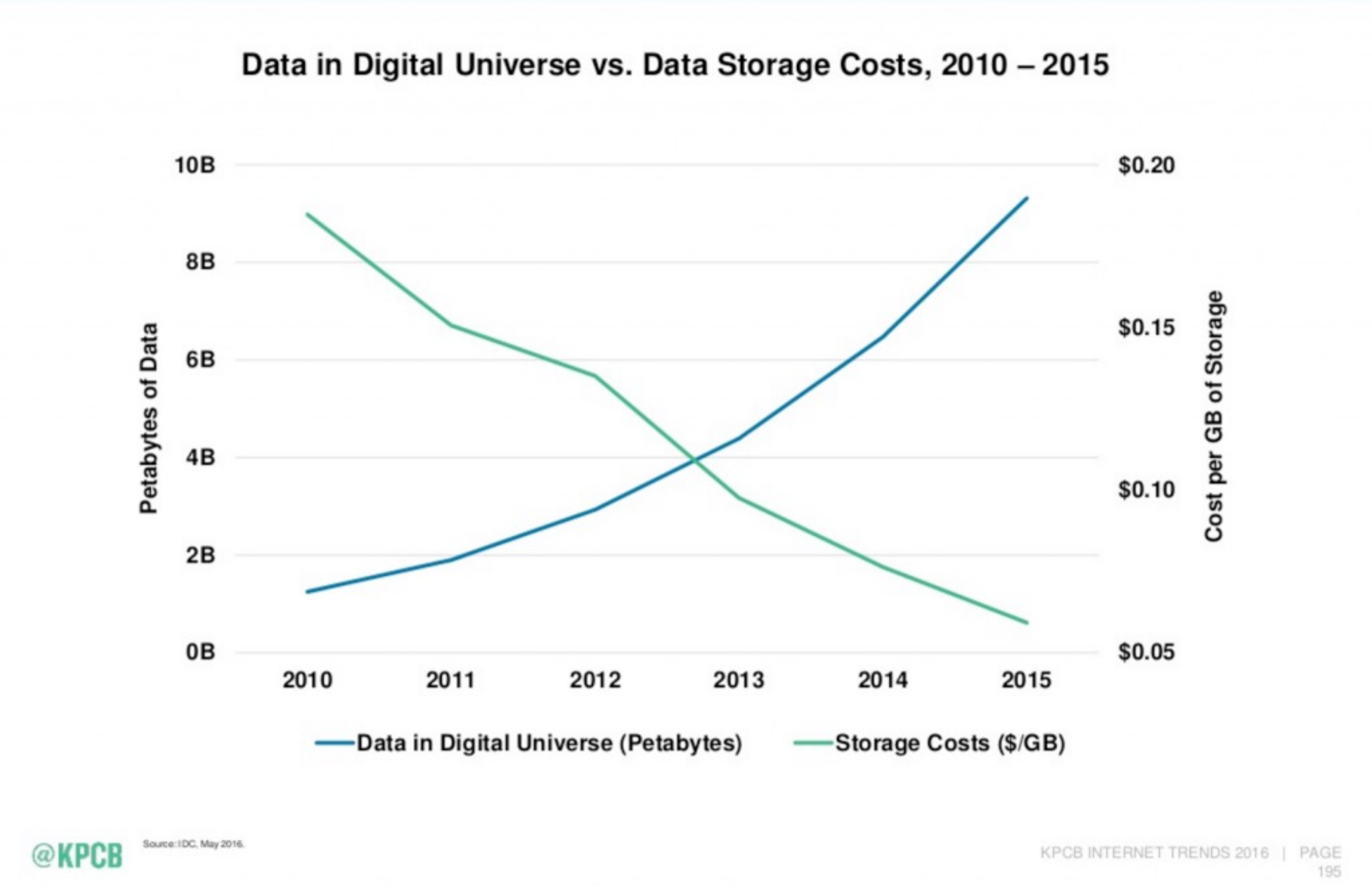
Buzz about Machine Learning

Three factors have contributed to making machine learning hot. These are cheap data, algorithmic economy, and cloud-based solutions.

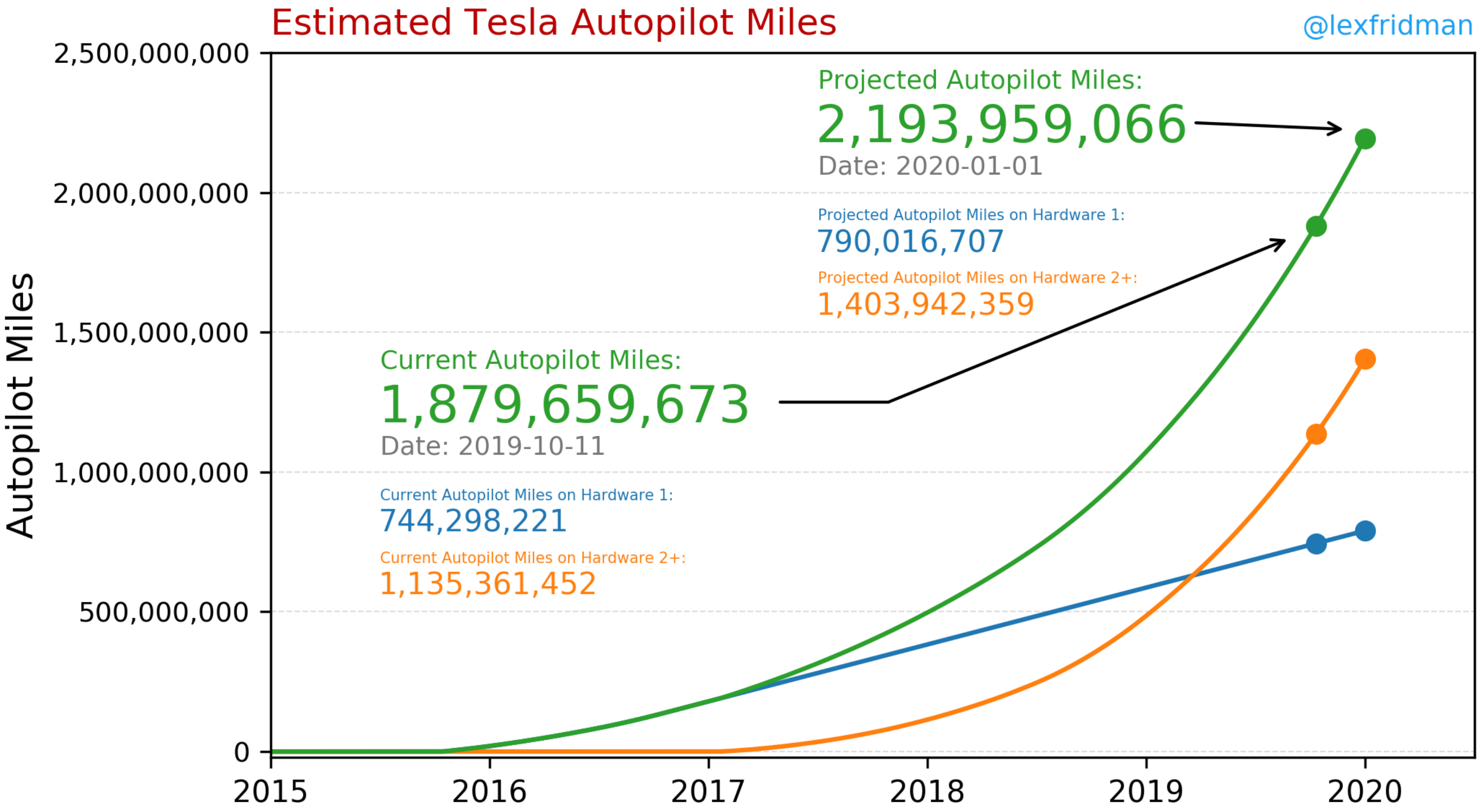
"Every company is now a data company, capable of using machine learning in the cloud to deploy intelligent apps at scale, thanks to three machine learning trends: data flywheels, the algorithm economy, and cloud-hosted intelligence."

MADRONA
VENTURE GROUP

Data is Getting Cheaper



Data Accumulation at Tesla



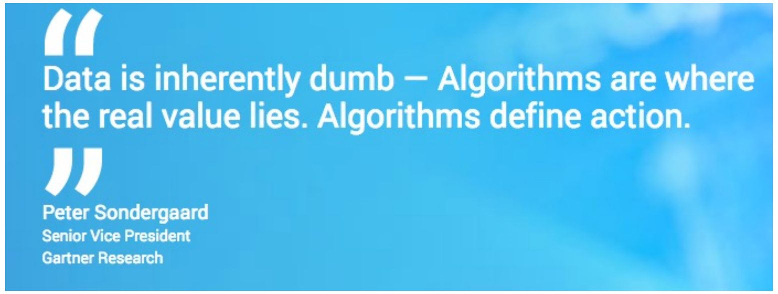
Algorithmic Economy

Host algorithms
Anyone can turn their algorithms into scalable/shareable, production ready web services
Typical users: scientists, academics, domain experts

Make algorithms discoverable
Anyone can use and integrate these algorithms into their solutions
Typical users: businesses, data scientists, app developers, IoT makers

Are monetizable
Align incentives between algorithm creators and consumers
Typical scenarios: heavy-load use cases with large user base

Are modular
Algorithms can be stacked or piped together
Typical scenarios: interpretation of unstructured data



Algorithm Economy Players in ML

Metamind
Indico
Clarifai
Microsoft Translator
MonkeyLearn
Image Processing
Text Utilities

HP Idol OnDemand
Robust Links
Lateral.io
Aylien
Algorithmia
Alchemy API
BigML

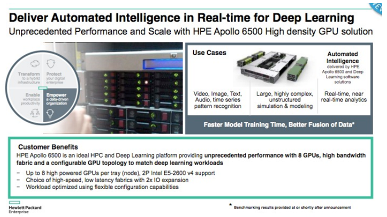
Cloud-Based Intelligence

Emerging machine intelligence platforms hosting **pre-trained machine learning models-as-a-service** are making it easy for companies to get started with ML, allowing them to rapidly take their applications from prototype to production.

Amazon Machine Learning

CLOUD SPEECH API BETA
Speech to text conversion powered by machine learning

DataRobot



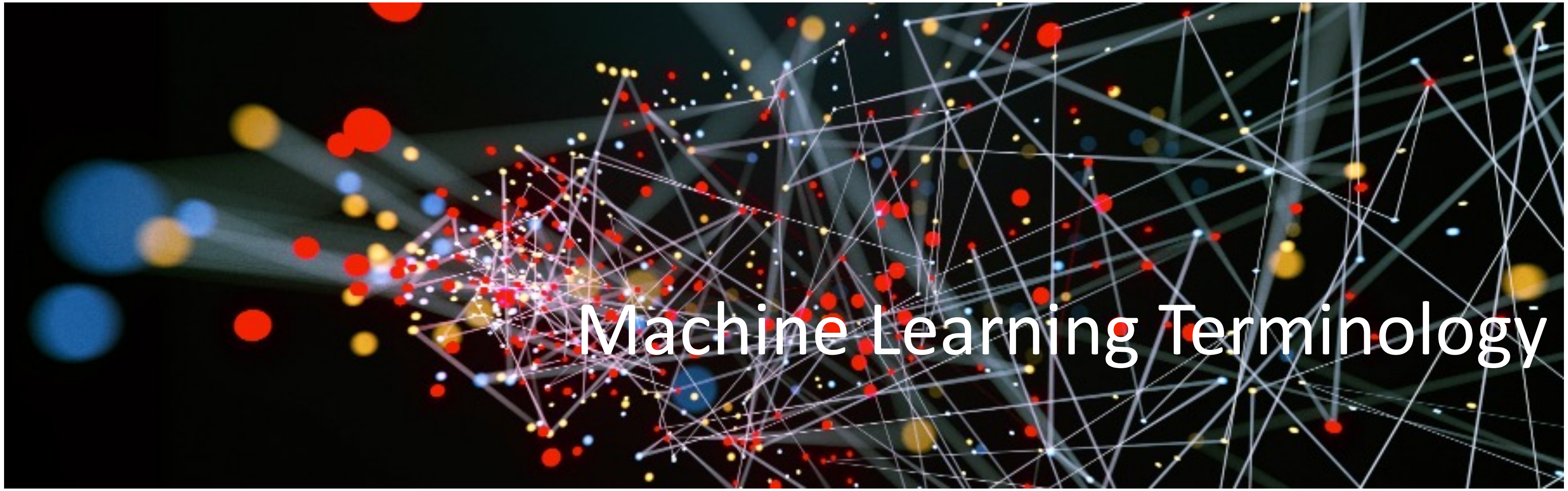
ersatz labs

Microsoft Azure

databricks

Many open source machine learning and deep learning frameworks running in the cloud allow easy leveraging of pre-trained, hosted models to tag images, recommend products, and do general natural language processing tasks.

iksinc@yahoo.com



Machine Learning Terminology

iksinc@yahoo.com

Feature Vectors in ML

- A machine learning system builds models using properties of objects being modeled. These properties are called *features* or *attributes* and the process of measuring/obtaining such properties is called *feature extraction*. It is common to represent the properties of objects as feature vectors.

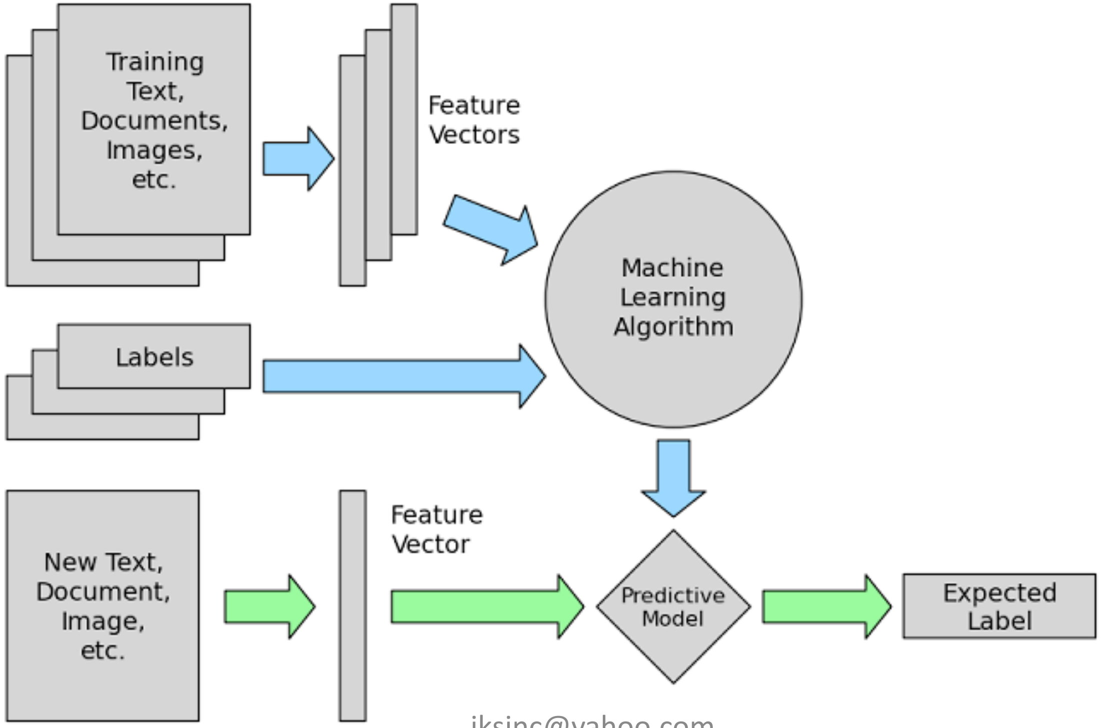
$$\longrightarrow \mathbf{x} = \begin{array}{cl} \left[\begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{matrix}\right] & \begin{matrix} \text{Sepal width} \\ \text{Sepal length} \\ \text{Petal width} \\ \text{Petal length} \end{matrix} \end{array}$$

Size (feet²) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4
2104	5	1	45
1416	3	2	40
1534	3	2	30
852	2	1	36

iksinc@yahoo.com

Learning Styles

- Supervised Learning
  - Training data comes with answers, called *labels*
  - The goal is to produce labels for new data



Supervised Learning Models

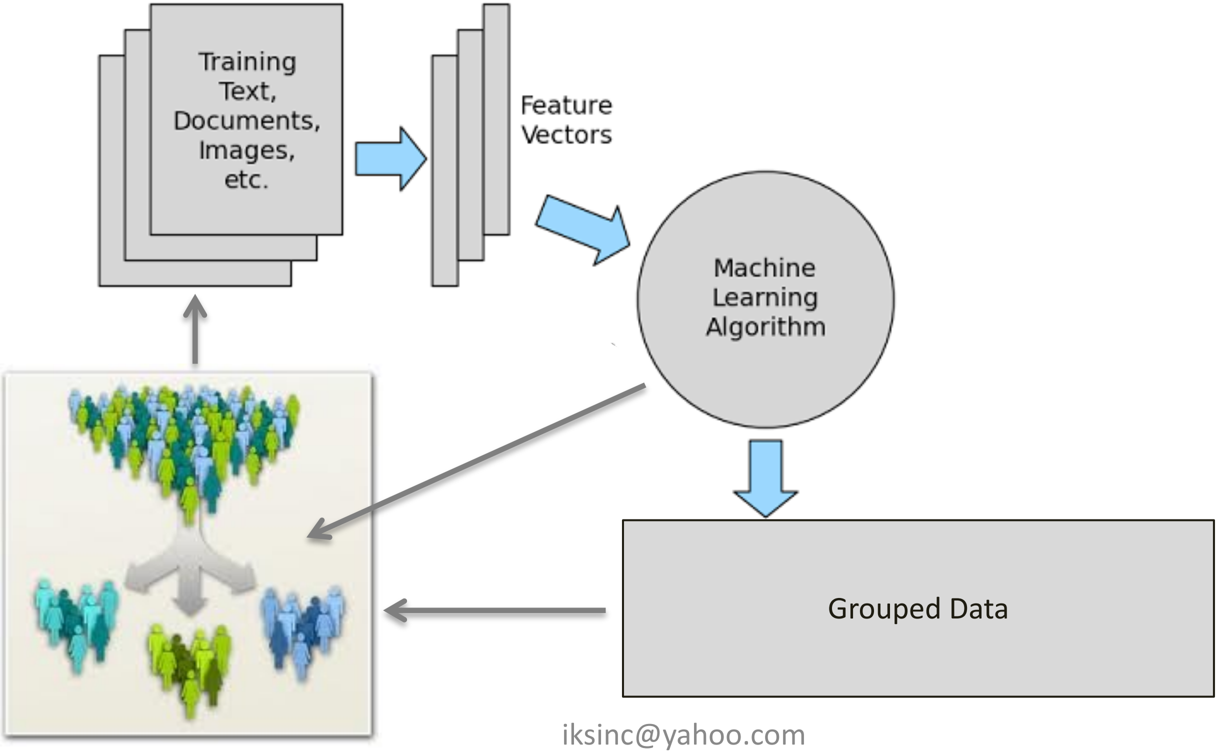
- Classification models
  - Predict customer churning
  - Tag objects in a given image
  - Determine whether an incoming email is spam or not

Supervised Learning Models

- Regression models
  - Predict credit card balance of customers
  - Predict the number of 'likes' for a posting
  - Predict peak load for a utility given weather information

Learning Styles

- Unsupervised Learning
  - Training data comes without labels
  - The goal is to group data into different categories based on similarities

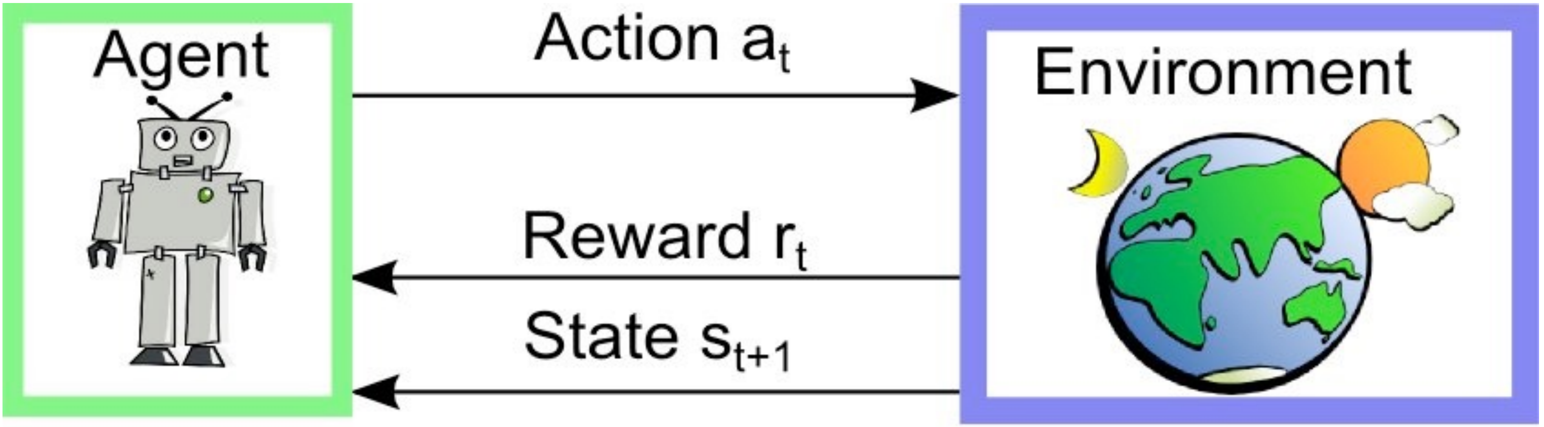


Unsupervised Learning Models

- Segment/ cluster customers into different groups
- Organize a collection of documents based on their content
- Make product Recommendations

Learning Styles

- Reinforcement Learning
  - Training data comes without labels
  - The learning system receives feedback from its operating environment to know how well it is doing
  - The goal is to perform better



Reinforcement Learning Setup

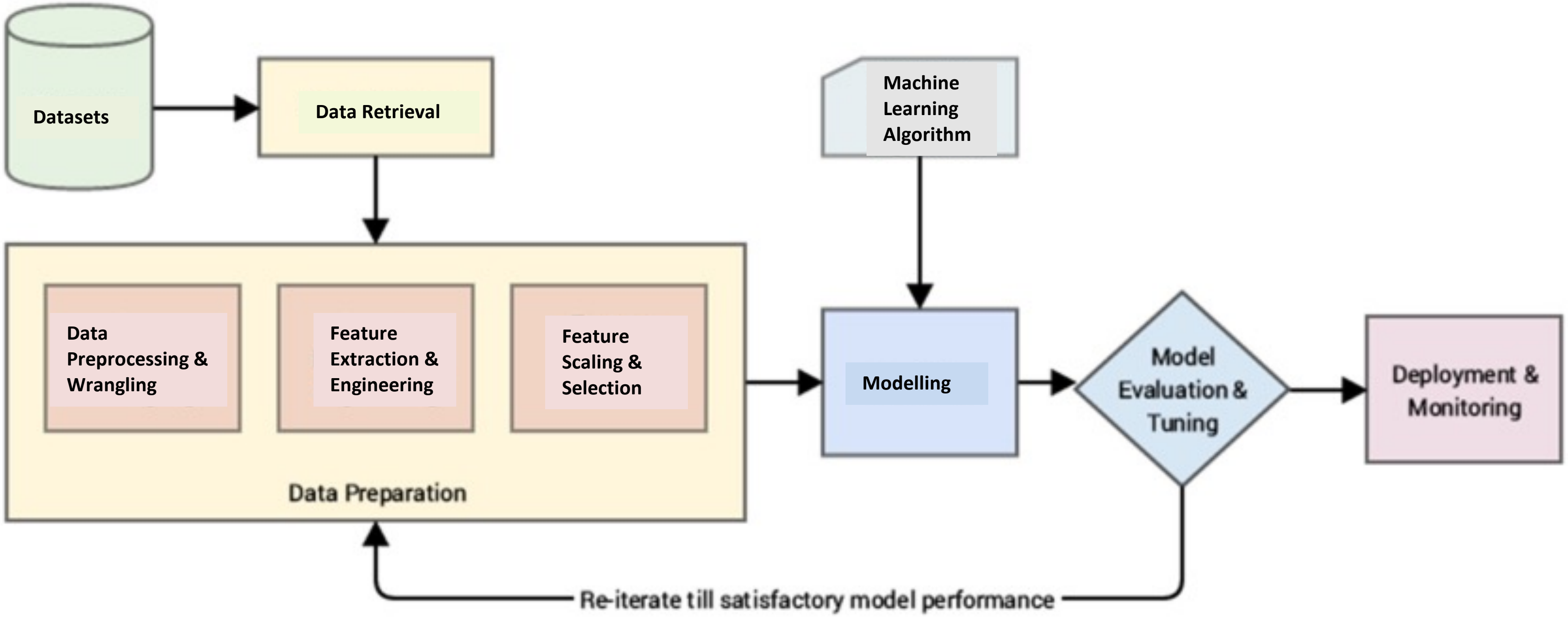
Reinforcement Learning

- Making a robot walk
- Portfolio optimisation
- Playing games better than humans
- Helicopter stunt manoeuvres
- Optimal communication protocols in radio networks
- Display ads
- Search engines
- ...

Machine Learning Workflow

iksinc@yahoo.com

Machine Learning Workflow



A Walk-Through Example: Flower Classification

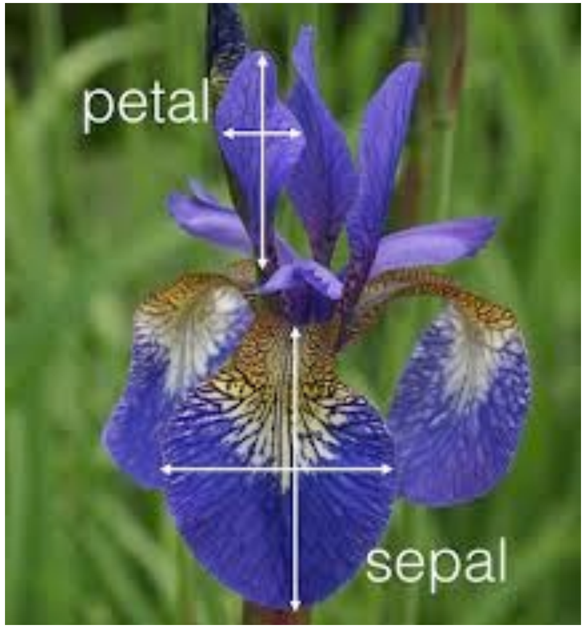
- Build a classification model to differentiate between two classes of flowers

Iris Virginica

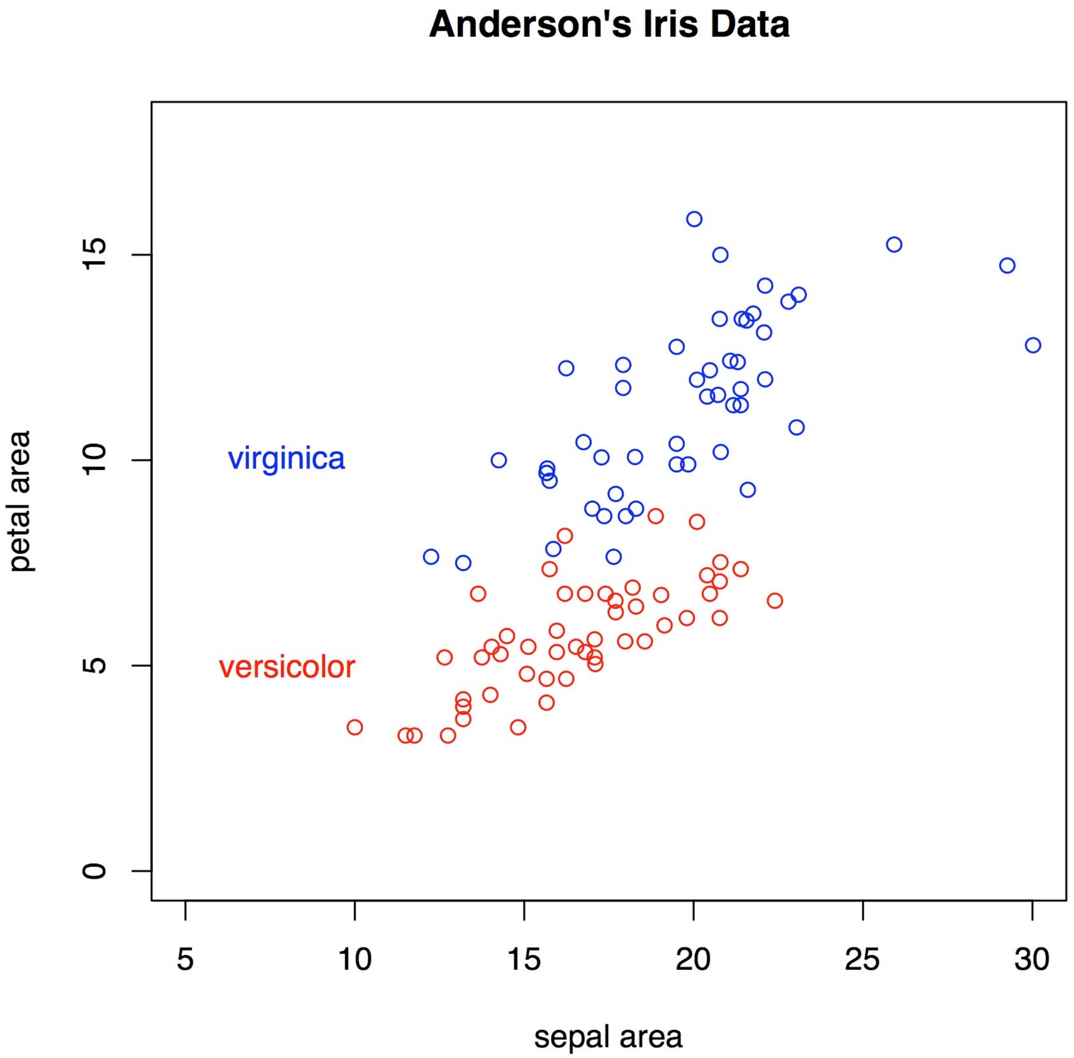
Iris Versicolor

How Do We Go About It?

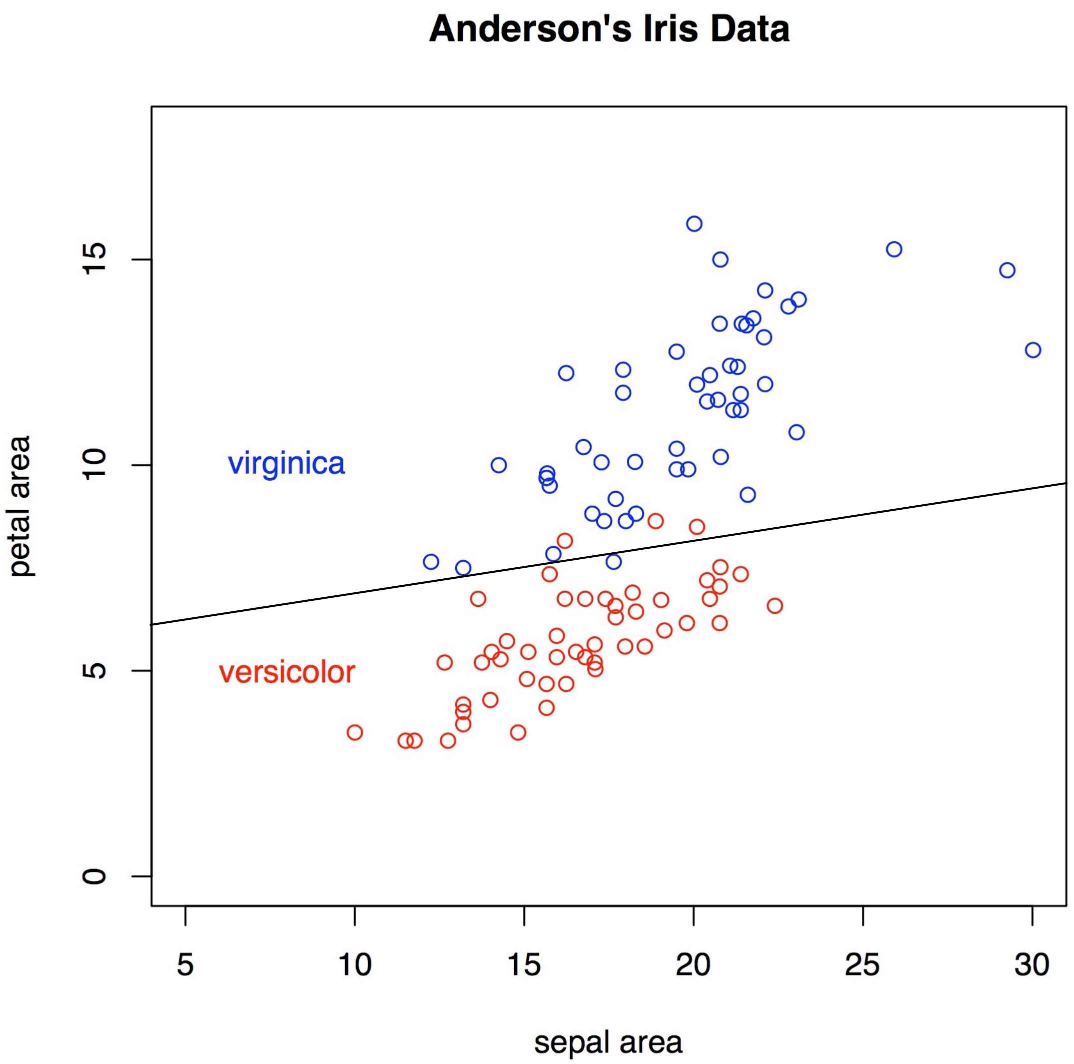
- Data Collection Step: Collect many flowers of both types with the help of an expert (Expert labeling)
- Feature Design and Extraction Step: Measure some attributes that can help differentiate between the two types of flowers. Let those attributes be petal area and sepal area.



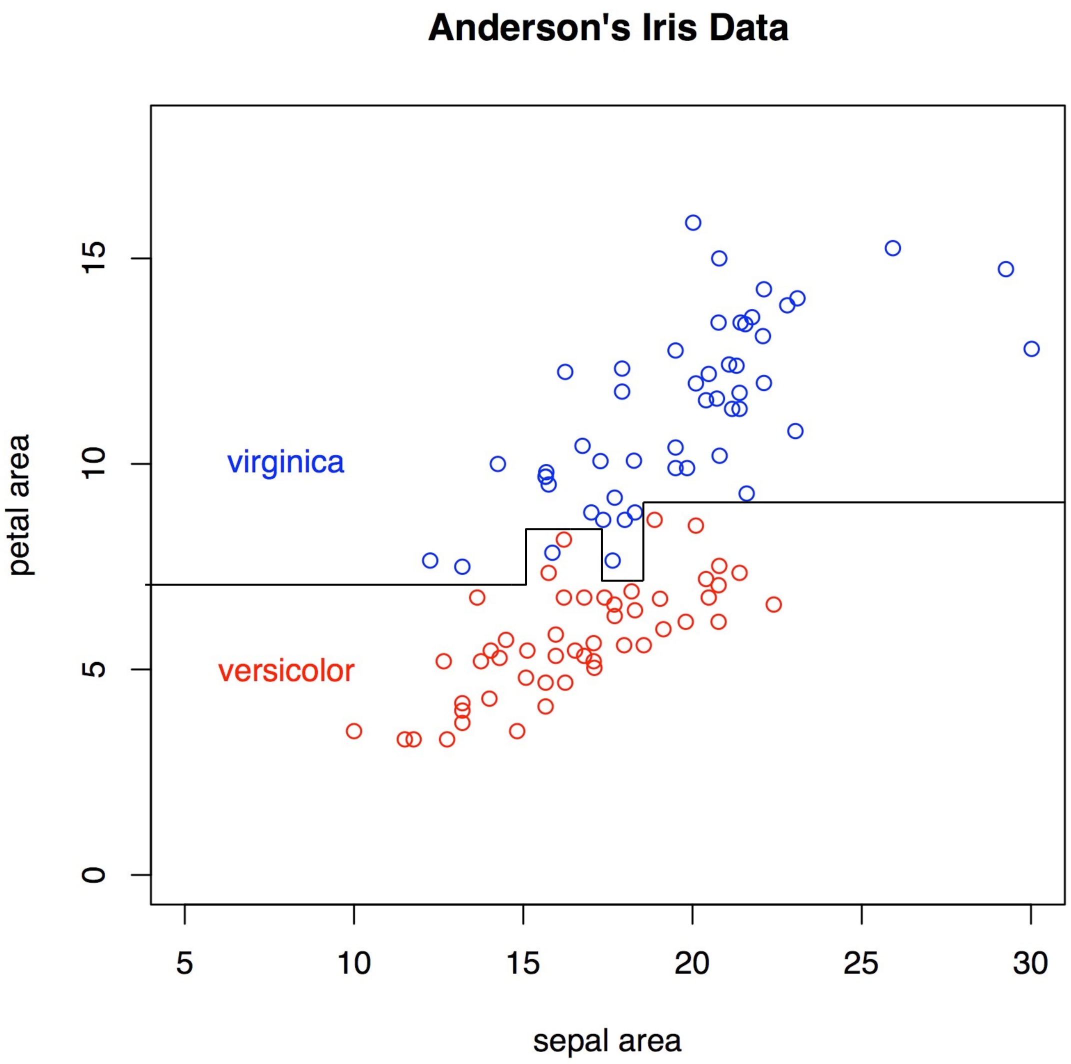
Scatter plot of 100 examples of flowers



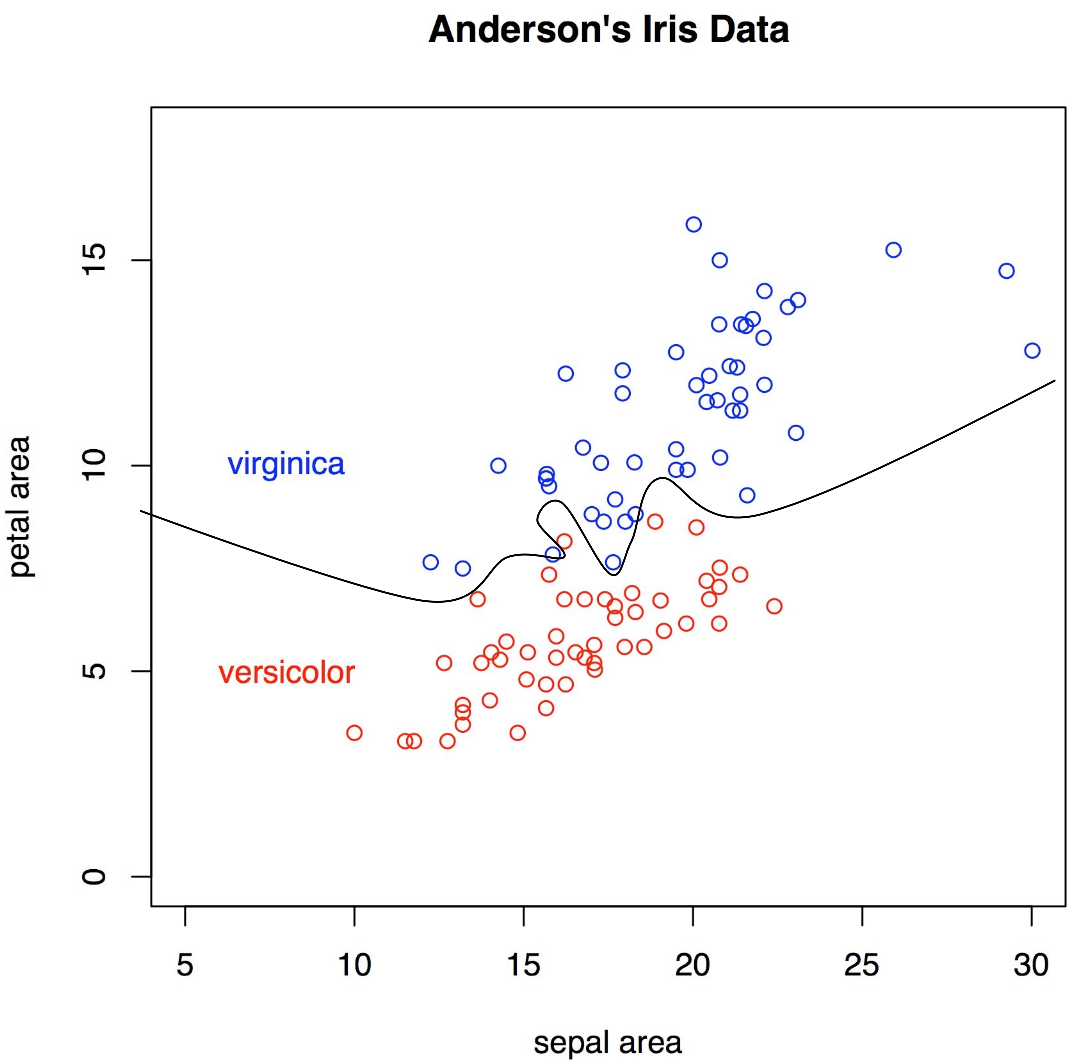
Visualization to understand the problem



We can separate the flower types using the linear boundary shown next. The parameters of the line represent the learned classification model.



Another possible boundary. This boundary cannot be expressed via an equation. However, a tree structure can be used to express this boundary. Note, this boundary does better prediction of the collected data.



Yet another possible boundary. This boundary does prediction without any error. Is this a better boundary?

Model Complexity

- There are tradeoffs between the complexity of models and their performance in the field. A good design (model choice) weighs these tradeoffs.
- A good design should avoid overfitting. How?
  - Divide the entire data into three sets
    - Training set (about 70% of the total data). Use this set to build the model
    - Test set (about 20% of the total data). Use this set to estimate the model accuracy after deployment
    - Validation set (remaining 10% of the total data). Use this set to determine the appropriate settings for free parameters of the model. May not be required in some cases.

Measuring Model Performance

- True Positive: Correctly identified as relevant
- True Negative: Correctly identified as not relevant
- False Positive: Incorrectly labeled as relevant
- False Negative: Incorrectly labeled as not relevant

Cat vs. No Cat

Image:

True Positive | True Negative | False Negative | False Positive

Precision, Recall, and Accuracy

- Precision
  - Percentage of positive labels that are correct
  - Precision = (# true positives) / (# true positives + # false positives)
- Recall
  - Percentage of positive examples that are correctly labeled
  - Recall = (# true positives) / (# true positives + # false negatives)
- Accuracy
  - Percentage of correct labels
  - Accuracy = (# true positives + # true negatives) / (# of samples)

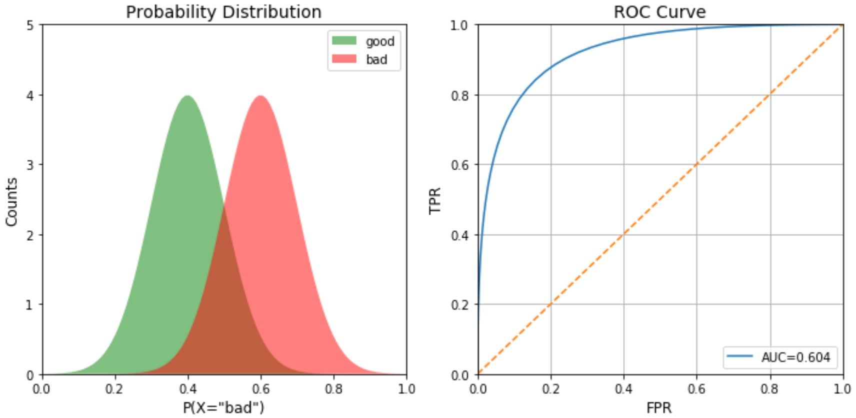
iksinc@yahoo.com

Accuracy?

- The previous definition of accuracy can be misleading when you have uneven representation from two classes. For example, consider a 2-class problem with 90 examples from class 1 and 10 from class 2.
  - Class predictive model 1 classifies 75 of class 1 examples and 3 of class 2 examples correctly
  - Another model classifies 60 of class 1 examples and 7 of class 2 examples correctly
  - Which model is a better predictor?
- What happens when different mistakes do not cost the same?

ROC (Receiver Operating Characteristics) Curve
AUC (Area under the ROC Curve)

- True Positive Rate (TPR) = TP/(TP+FN) = a/(a+c)
- False Positive Rate (FPR) = FP/(FP+TN) = b/(b+d)
- A ROC curve plots TPR vs. FPR at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives.



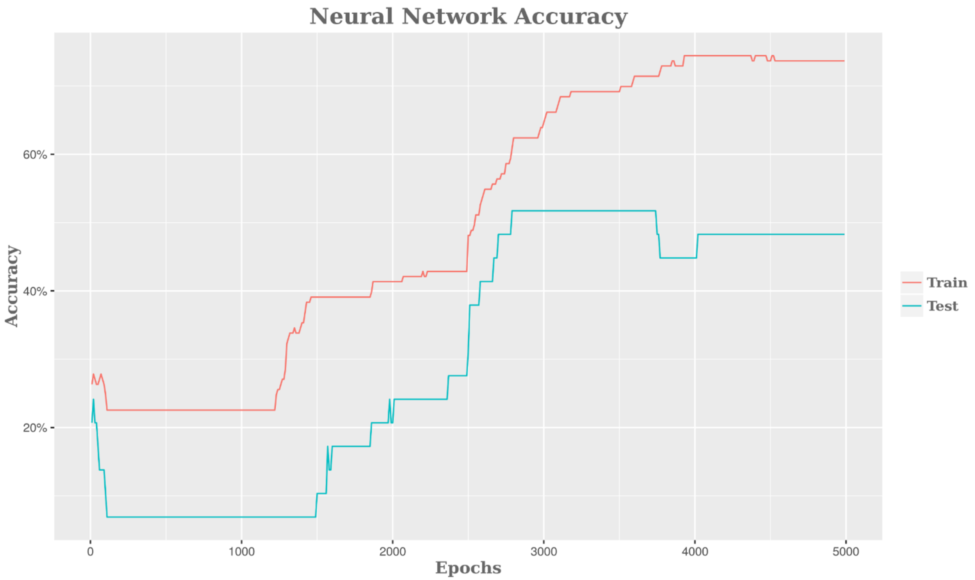
Confusion Matrix for M Classes

Predicted class (columns); Actual class (rows)

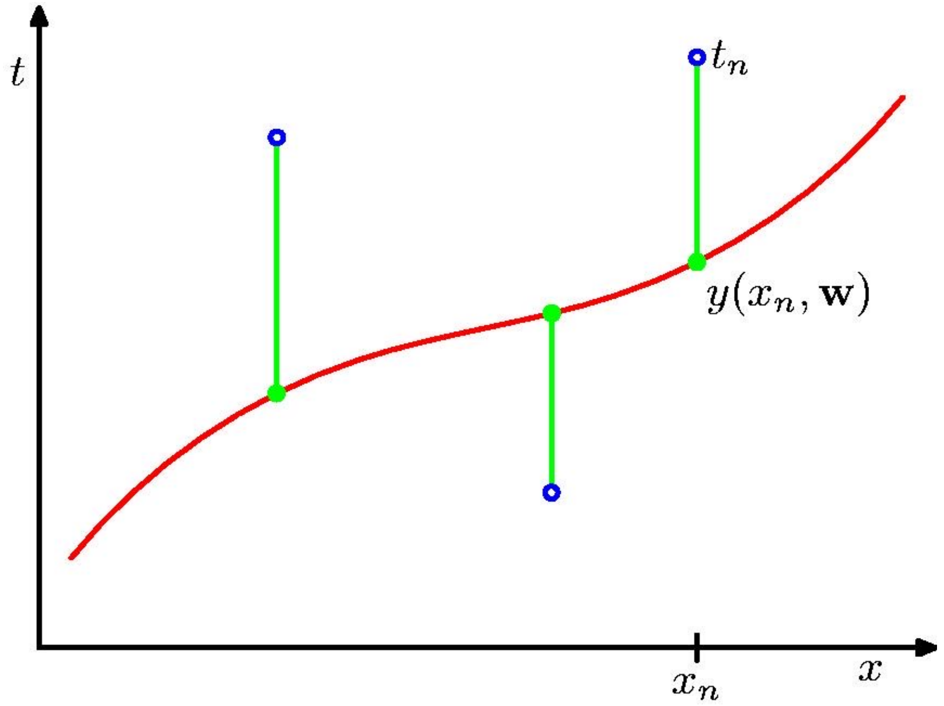
	Sit	Stand	Walk Jog	Ascend	Descend	Cycle	**Class-specific recall**
Sit	*3202*	*2*	*0*	*0*	*0*	*14*	**0.99**
Stand	*7*	*3191*	*2*	*7*	*0*	*0*	**0.99**
Walk	*0*	*0*	*10647*	*74*	*0*	*0*	**0.99**
Ascend	*0*	*0*	*34*	*500*	*15*	*1*	**0.90**
Descend	*0*	*0*	*41*	*60*	*405*	*0*	**0.80**
Cycle	*146*	*3*	*0*	*0*	*0*	*2539*	**0.94**
Class-specific precision	**0.95**	**1.00**	**0.99**	**0.78**	**0.96**	**0.99**	**0.98**

Overfitting

- Model overfitting causes poor performance on test data while giving an excellent performance on training data

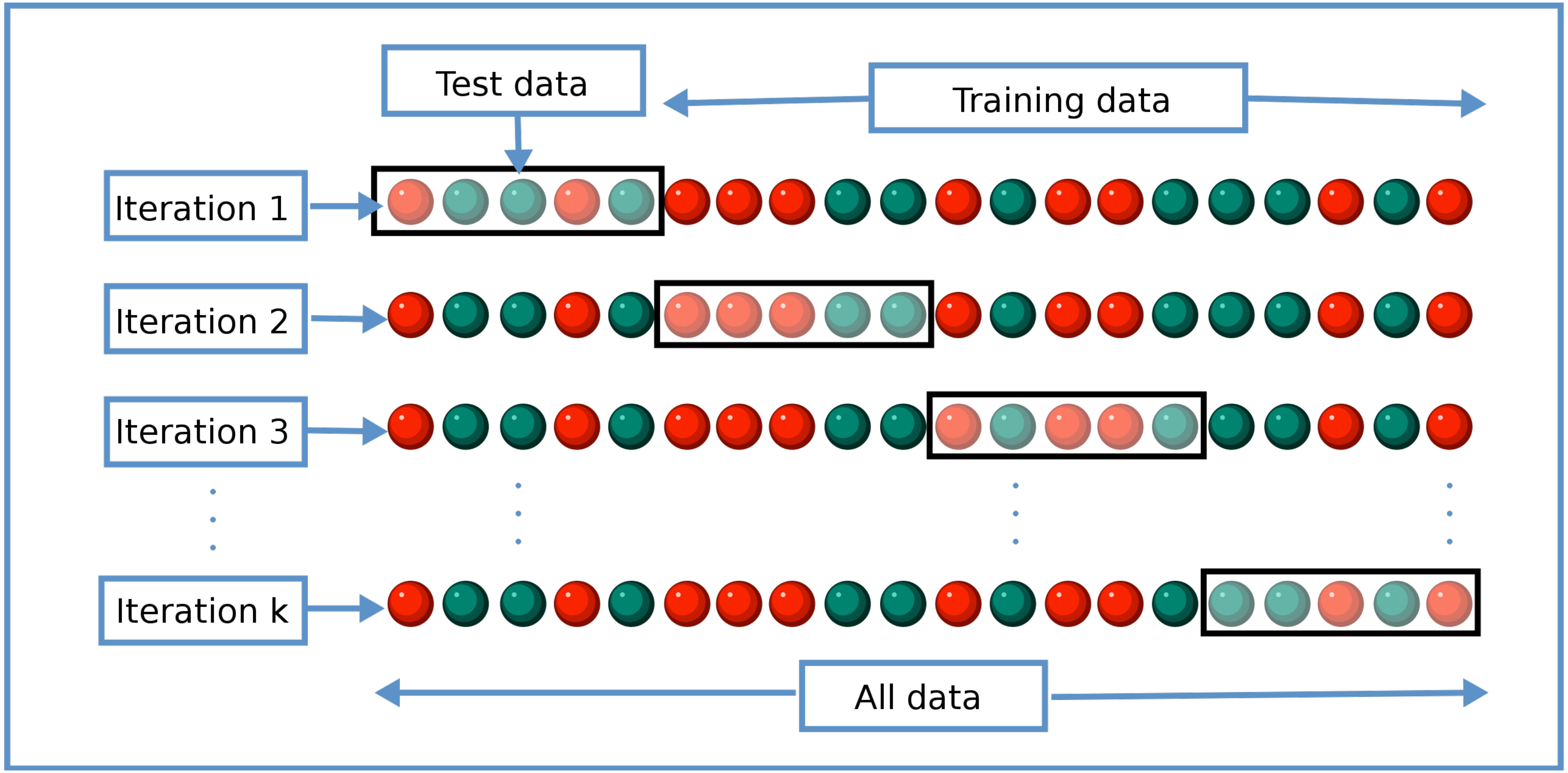


Sum-of-Squares Error for Regression Models



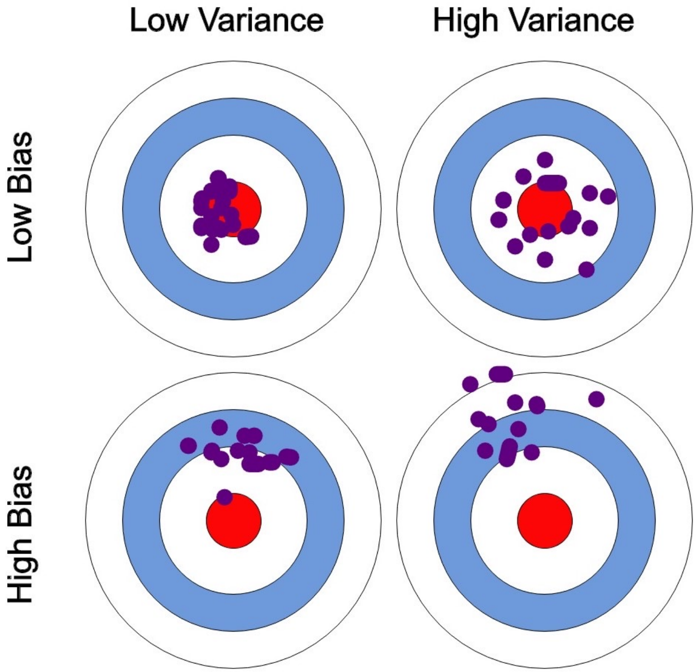
For regression model, the error is measured by taking the square of the difference between the predicted output value and the target value for each training (test) example and adding this number over all examples as shown

Cross Validation

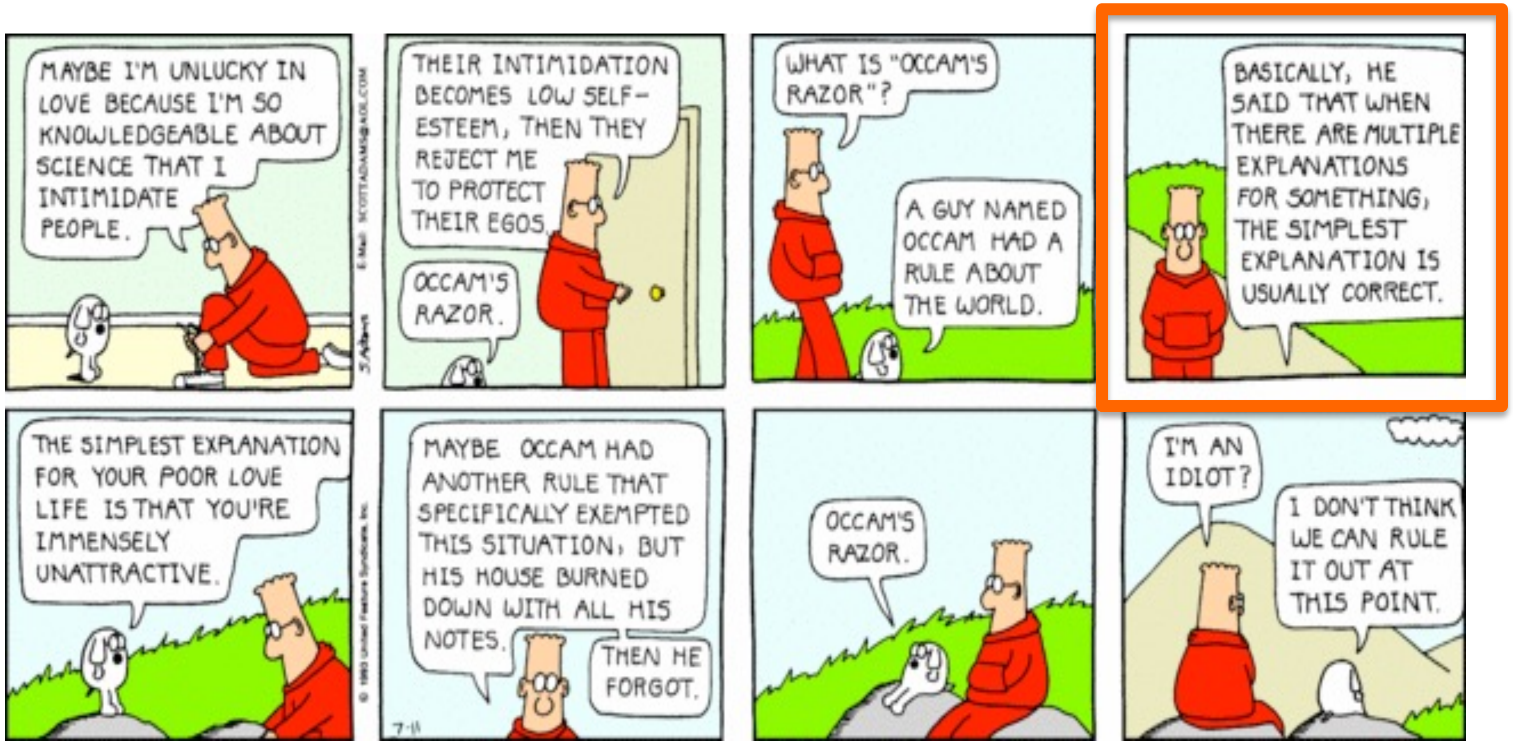


Bias and Variance

- Bias: expected difference between model's prediction and truth
- Variance: how much the model differs among training sets
- Model Scenarios
  - High Bias: Model makes inaccurate predictions on training data
  - High Variance: Model does not generalize to new datasets
  - Low Bias: Model makes accurate predictions on training data
  - Low Variance: Model generalizes to new datasets



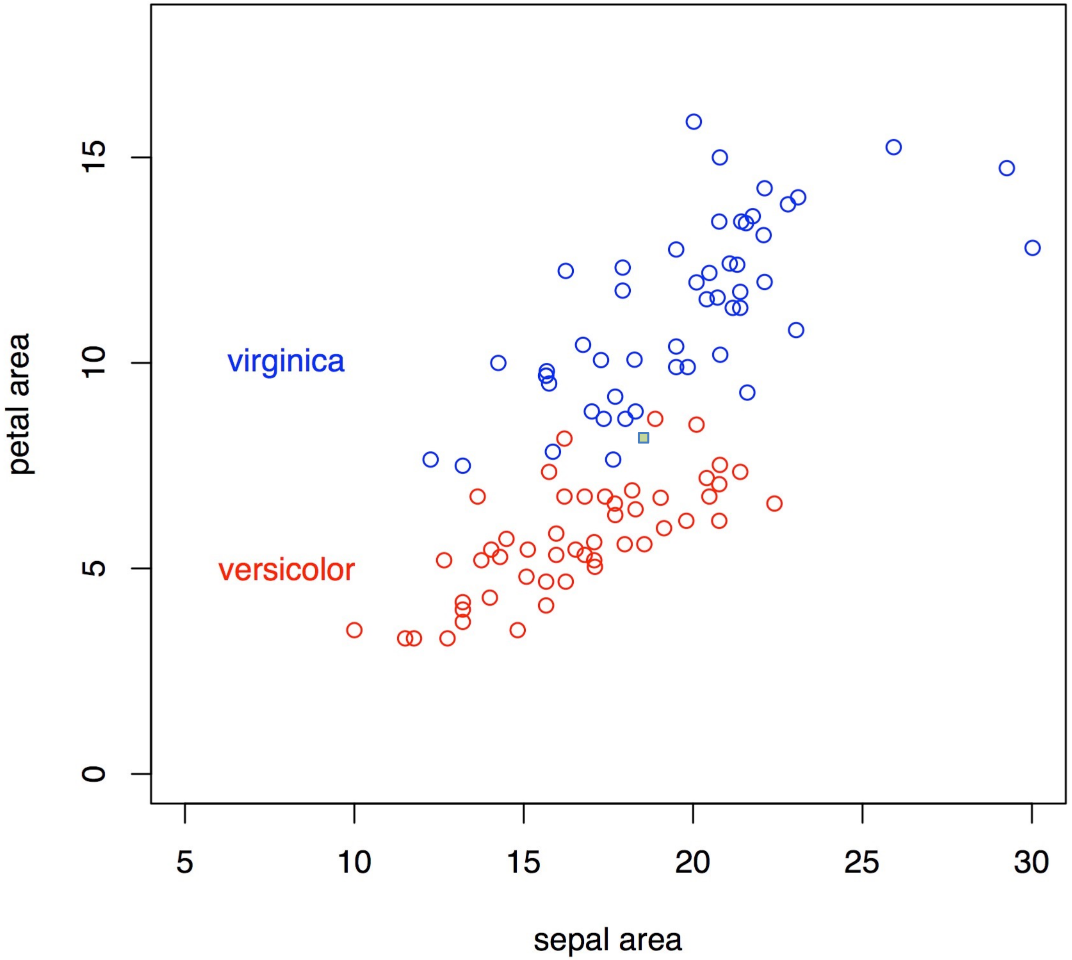
The Guiding Principle for Model Selection: Occam's Razor



Model Building Algorithms

- Supervised learning algorithms
  - Linear methods
  - k-NN classifier
  - Neural networks
  - Support vector machine
  - Decision trees
  - Ensemble methods (Random Forest, XGBT)

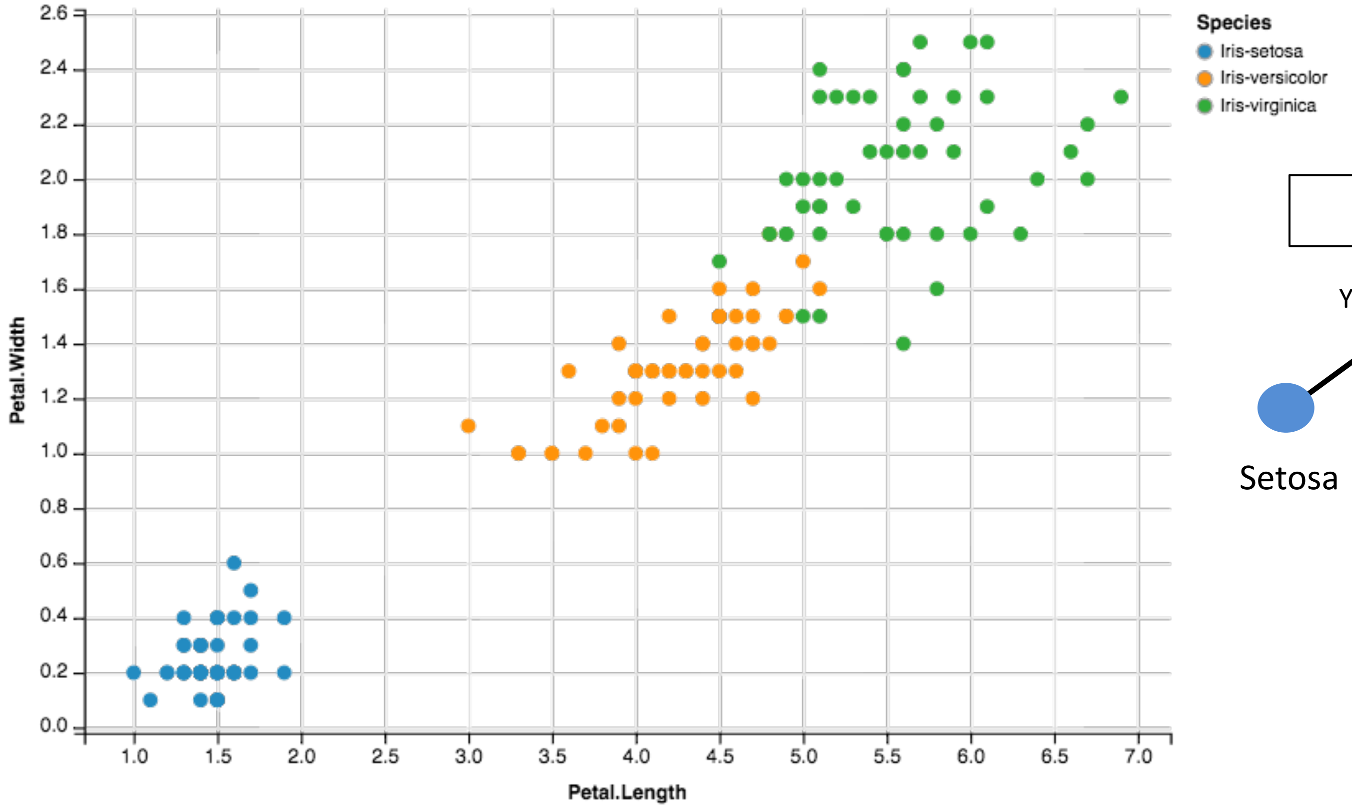
Illustration of k-NN Model

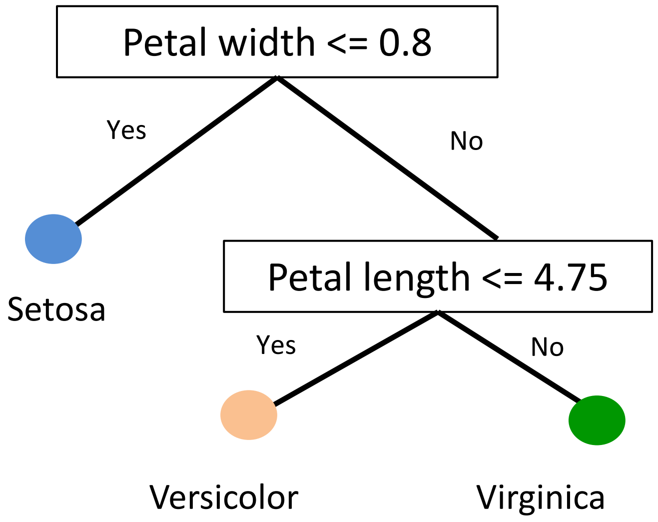


Test example

Predicted label of test example with 1-NN model : Versicolor
Predicted label of text example with 3-NN model: Virginica

Illustration of Decision Tree Model



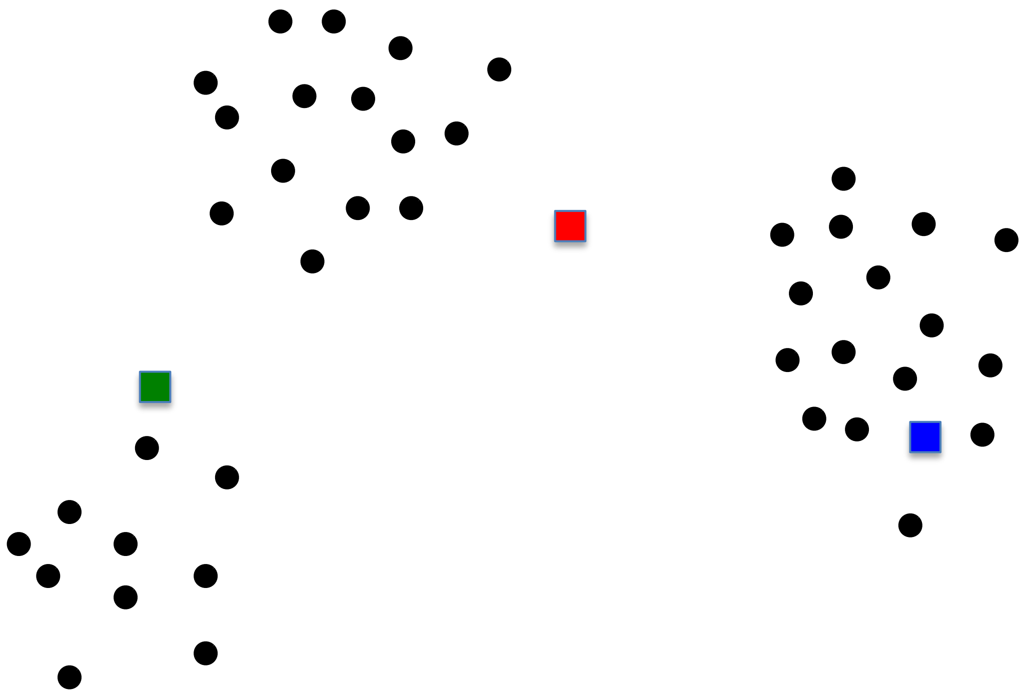


The decision tree is automatically generated by a machine learning algorithm.

Model Building Algorithms

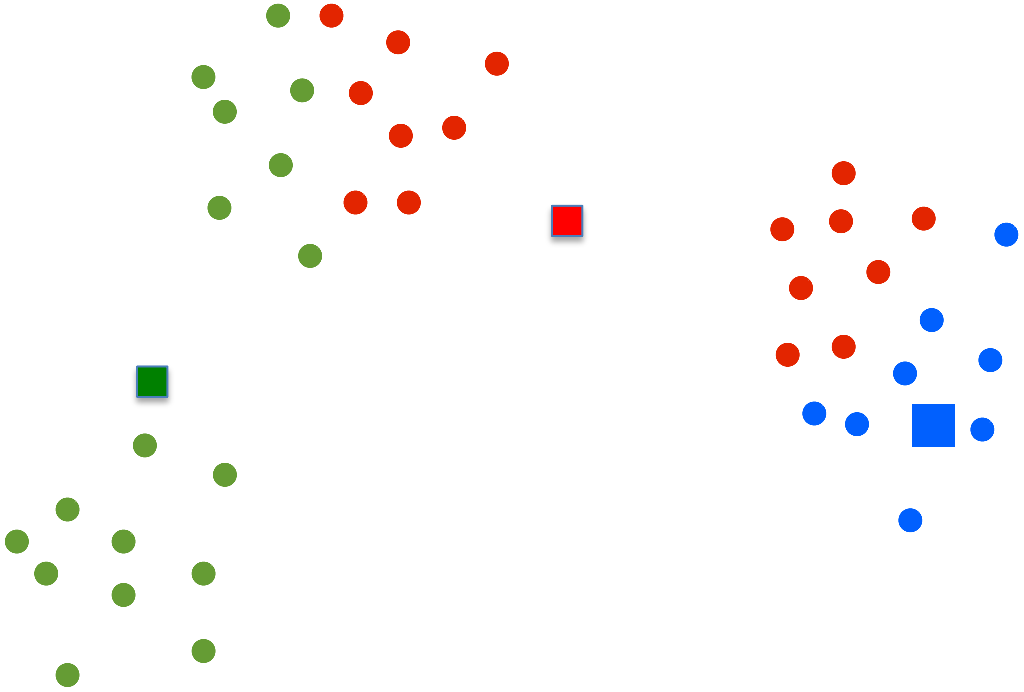
- Unsupervised learning
  - k-means clustering
  - Agglomerative clustering
  - Self organization feature maps
  - Recommendation system

K-means Clustering



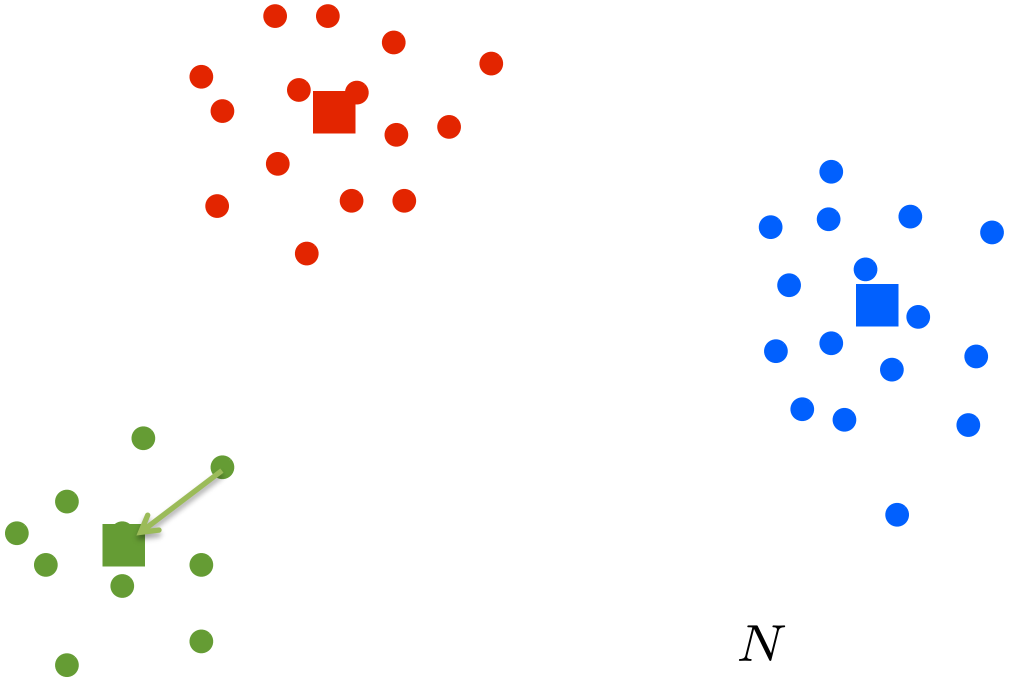
Choose the number of clusters, k, and initial cluster centers

K-means Clustering



Assign data points to clusters based on distance to cluster centers

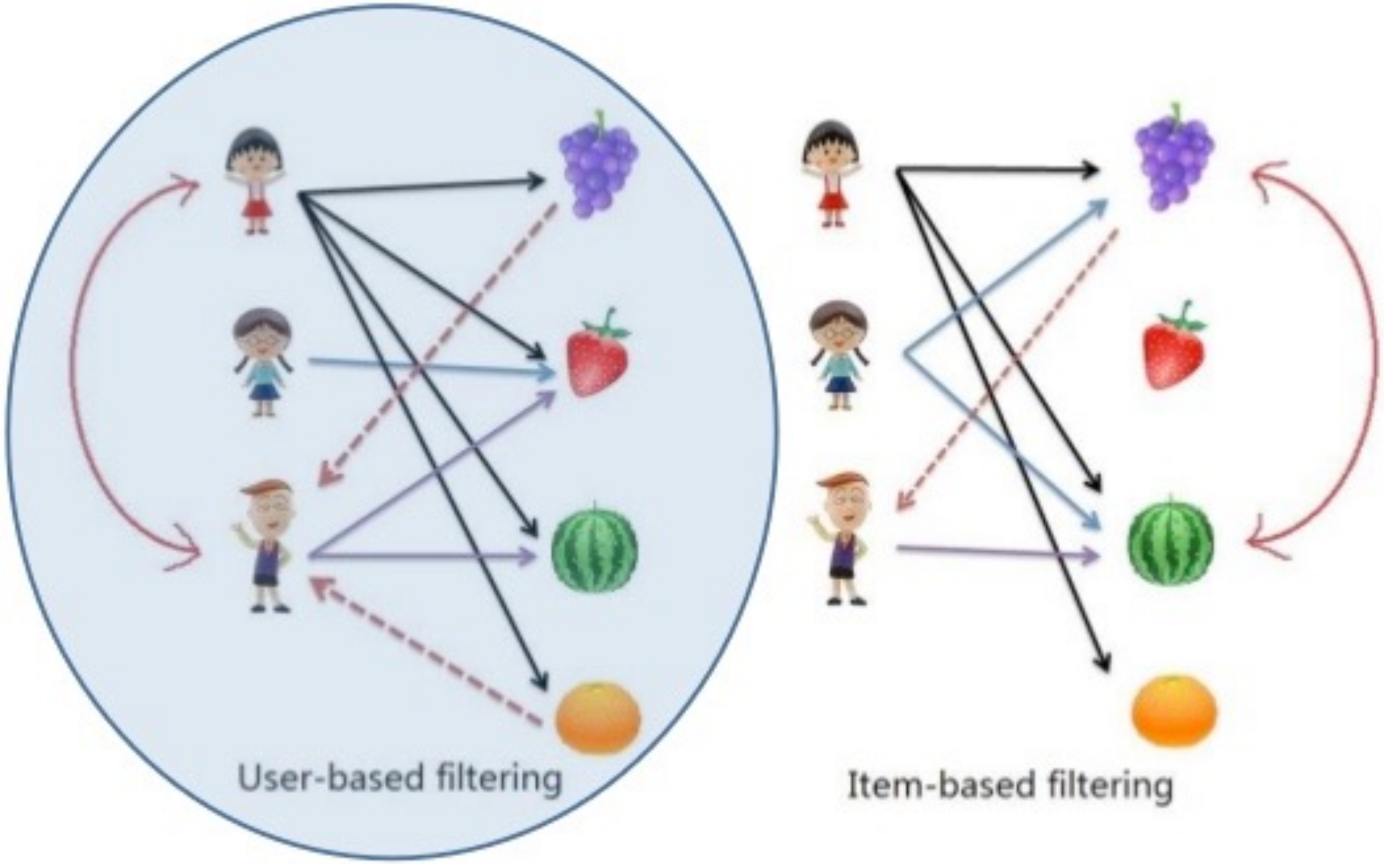
K-means Clustering



Update cluster centers and reassign data points.

$$\text{minimize} \sum_{n=1}^{N} \|x_n - \text{center}_n\|^2$$

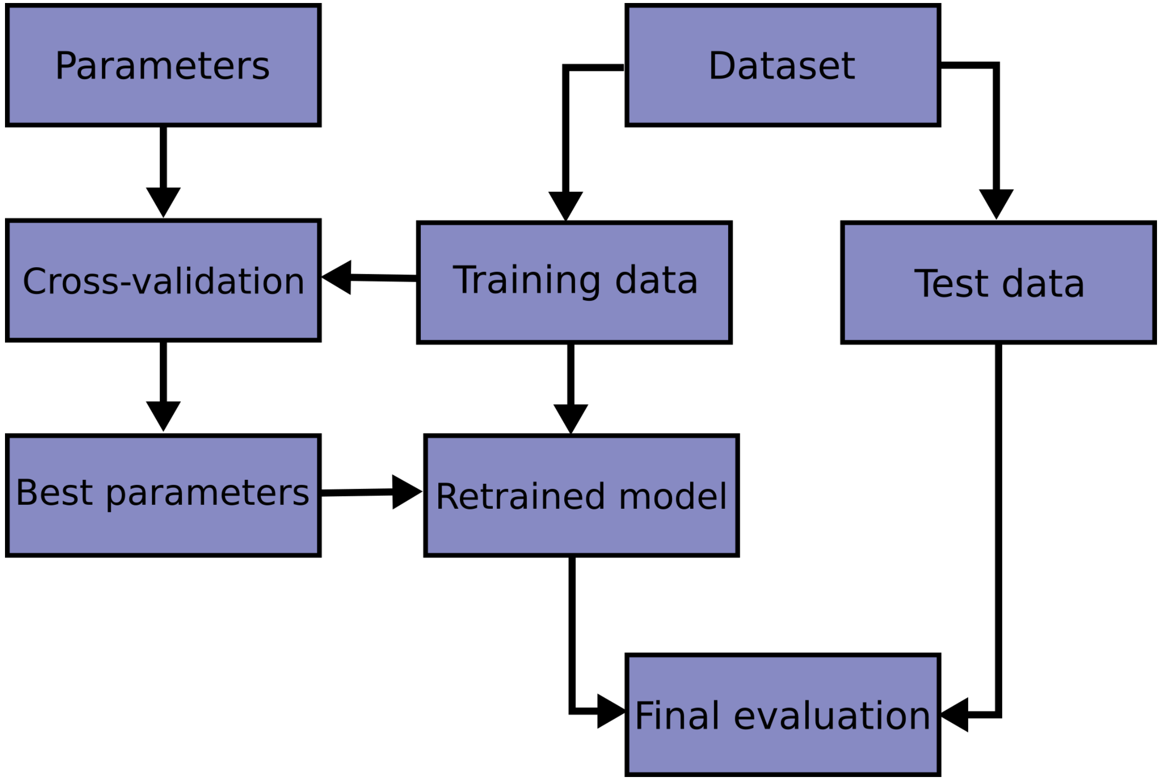
Illustration of Recommendation System




NO, I WILL NOT TAKE YOU TO A-1 STEAKHOUSE. I RECOMMEND YOU GO TO THE GYM FIRST, AND THEN TO V8 VEGETARIAN RESTAURANT.
KDNUGGETS.COM

Steps Towards a Machine Learning Project

- Have a business case
- Collect data
- Explore data via scatter plots, histograms. Remove duplicates and data records with missing values
- Check for dimensionality reduction
- Build model (iterative process)
- Transport/Integrate with an application



iksinc@yahoo.com

Machine Learning Limitation

- Machine learning methods operate on manually designed features or it relies on feature engineering.
- The design of such features for tasks involving computer vision, speech understanding, natural language processing is extremely difficult. This puts a limit on the performance of the system.

Processing Sensory Data is Hard

You see this:

But the camera sees this:

194	210	201	212	199	213	215	195	178	158	182	209
180	189	190	221	209	205	191	167	147	115	129	163
114	126	140	188	176	165	152	140	170	106	78	88
87	103	115	154	143	142	149	153	173	101	57	57
102	112	106	131	122	138	152	147	128	84	58	66
94	95	79	104	105	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	88	76	67
41	56	68	99	63	45	60	82	58	76	75	65
20	43	69	75	56	41	51	73	55	70	63	44
50	50	57	69	75	75	73	74	53	68	59	37
72	59	53	66	84	92	84	74	57	72	63	42
67	61	58	65	75	78	76	73	59	75	69	50

How do we bridge this gap between the pixels and meaning via machine learning?

Sensory Data Processing is Challenging

Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Sensory Data Processing is Challenging

Challenges: illumination

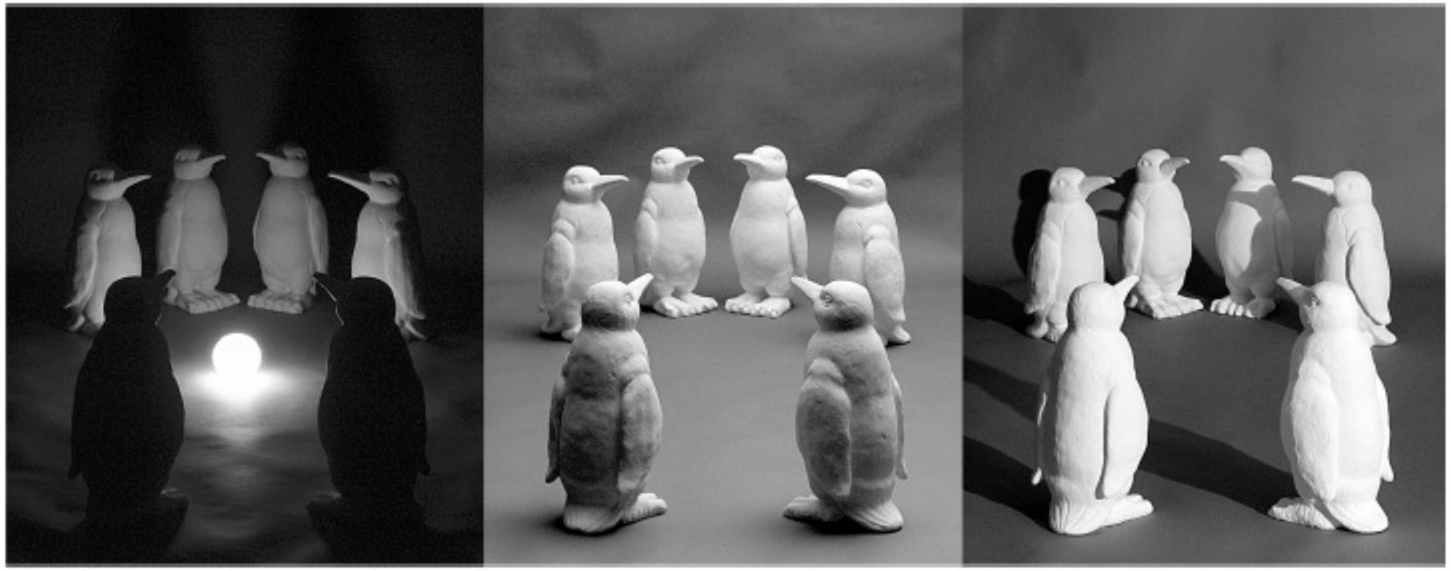


image credit: J. Koenderink

Sensory Data Processing is Challenging

Challenges: object intra-class variation

End-to-End Learning

Designing features is often difficult, time consuming, and requires expert knowledge.

So why not build integrated learning systems that perform end-to-end learning, i.e. learn the representation as well as classification from raw data without any engineered features.

Feature Learner

Trainable Classifier

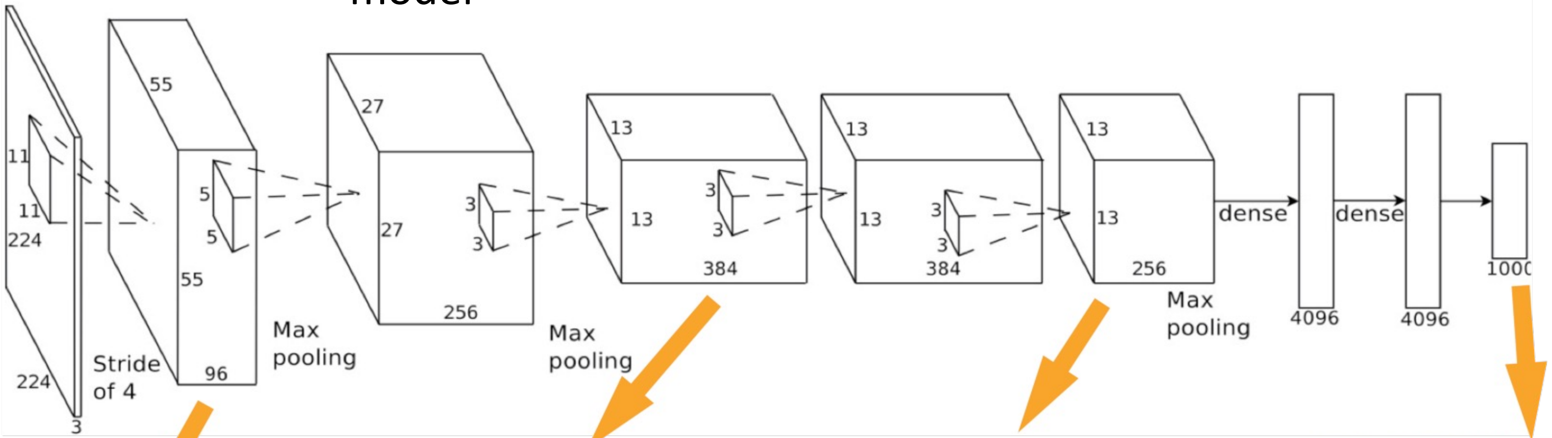
An approach performing end-to-end learning, typically performed through a series of successive abstractions, is in a nutshell deep learning

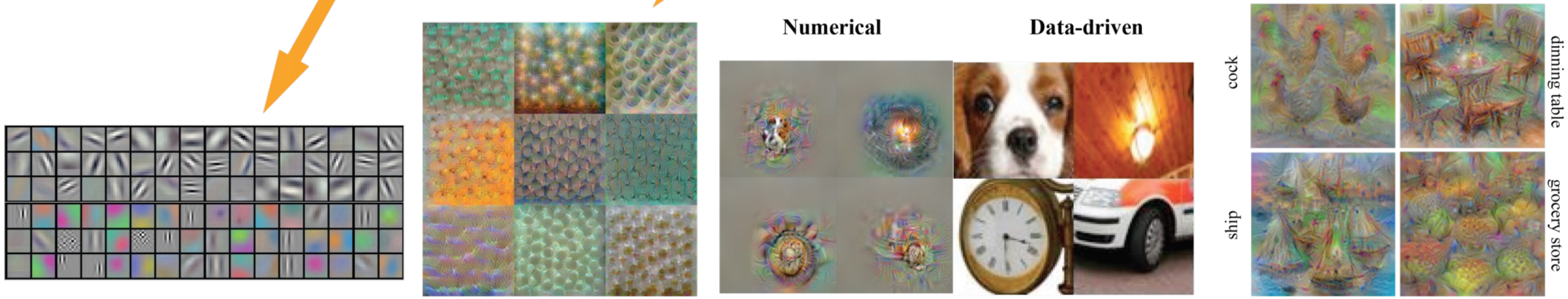
Deep Learning

- It’s a subfield of machine learning that has shown remarkable success in dealing with applications requiring processing of pictures, videos, speech, and text.
- Deep learning is characterized by:
  – Extremely large amount of data for training
  – Neural networks with exceedingly large number of layers
  – Training time running into weeks in many instances
  – End to end learning (No human designed rules/features are used)

Deep Learning Models

Convolutional Neural Networks: Most popular model





Conv 1: Edge+Blob

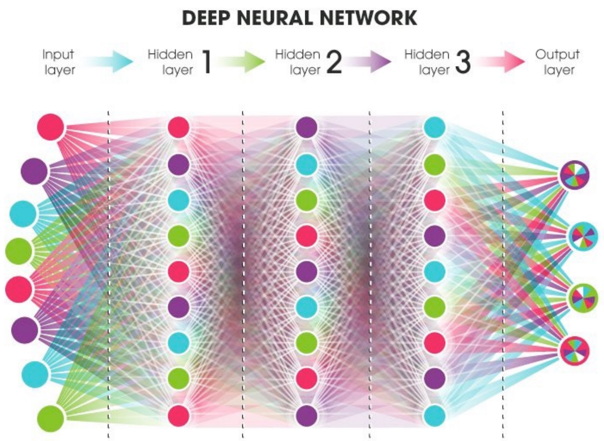
Conv 3: Texture

Conv 5: Object Parts

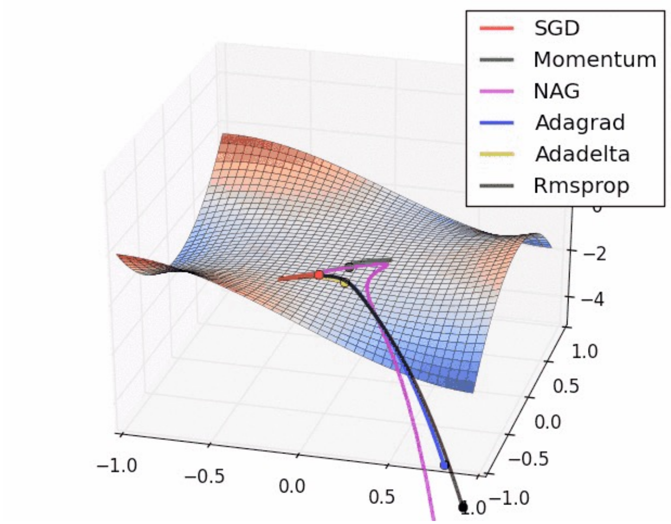
Fc8: Object Classes

Training Deep Neural Networks

- The work horse is the backpropagation algorithm based on chain-rule differentiation
- Training consists of optimizing a suitable loss function using the stochastic gradient descent (SGD) algorithm to adjust networks parameters typically running into millions.



=



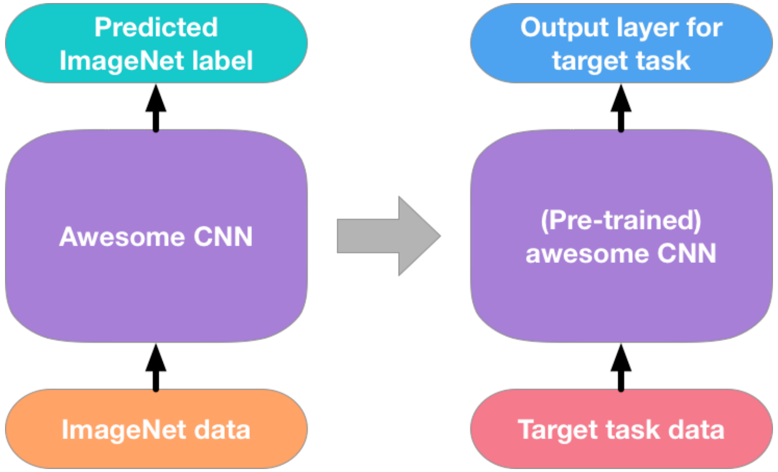
CNN Application Example:
Object Detection and Labeling

Other Deep Learning Models

- Recurrent neural networks (RNN)
- Generative Adversarial Networks (GANs)
- Autoencoders
- Transformers

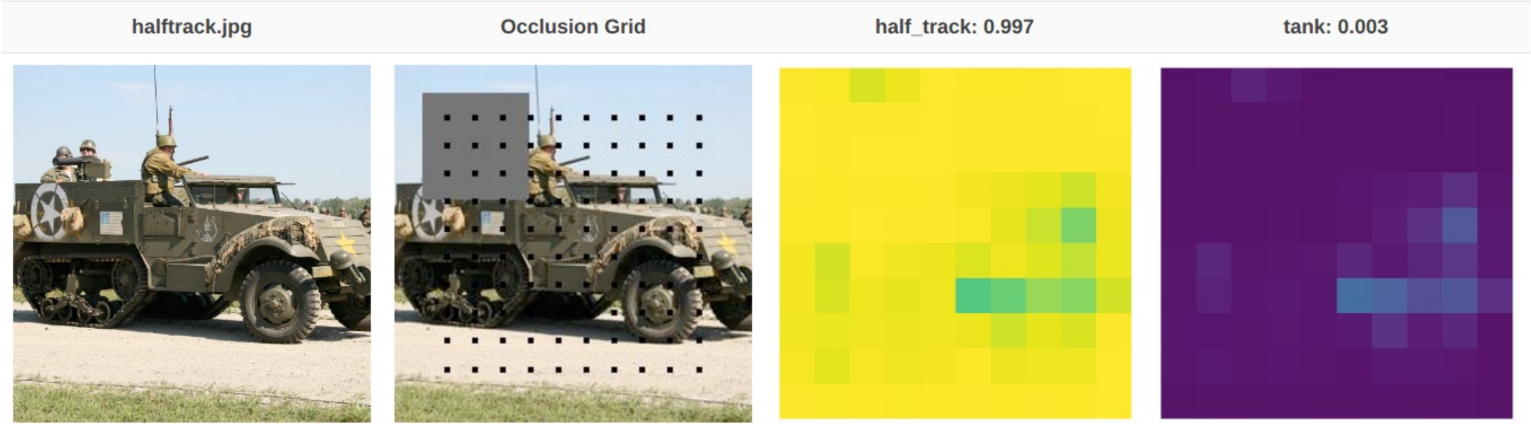
Transfer Learning

- Trained models for one task can be used for another task via transfer learning
  - Faster training
  - Data needs are low



Learning Bias

There is an urban legend that back in the 90's, the US government commissioned for a project to detect tanks in a picture. The researchers built a neural network and used it classify the images. Once the product was actually put to test, it did not perform at all. On further inspection they noticed that the model had learnt the weather patterns instead of the tanks. The trained images with tanks were taken on a cloudy day and images with no tanks were taken on a sunny day. This is a prime example of how we need to understand the learning by a neural net.



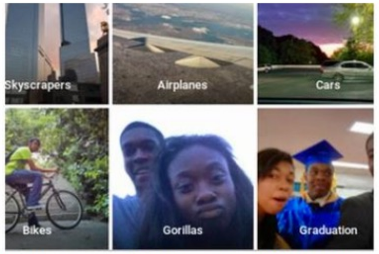
Learning Bias

Google Mistakenly Tags Black People as ‘Gorillas,’ Showing Limits of Algorithms

By Alistair Barr

Jul 1, 2015 3:40 pm ET

0 COMMENTS



Black programmer Jacky Alciné said on Twitter that the new Google Photos app had tagged photos of him and a friend as gorillas. *JACKY ALCINÉ AND TWITTER*

Microsoft silences its new A.I. bot Tay, after Twitter users teach...

Microsoft’s newly launched A.I.-powered bot called Tay, which was responding to tweets and chats on GroupMe and Kik, has already been s...

techcrunch.com

M Technology ▸ Apple

iPhone X racism row: Apple's Face ID fails to distinguish between Chinese users

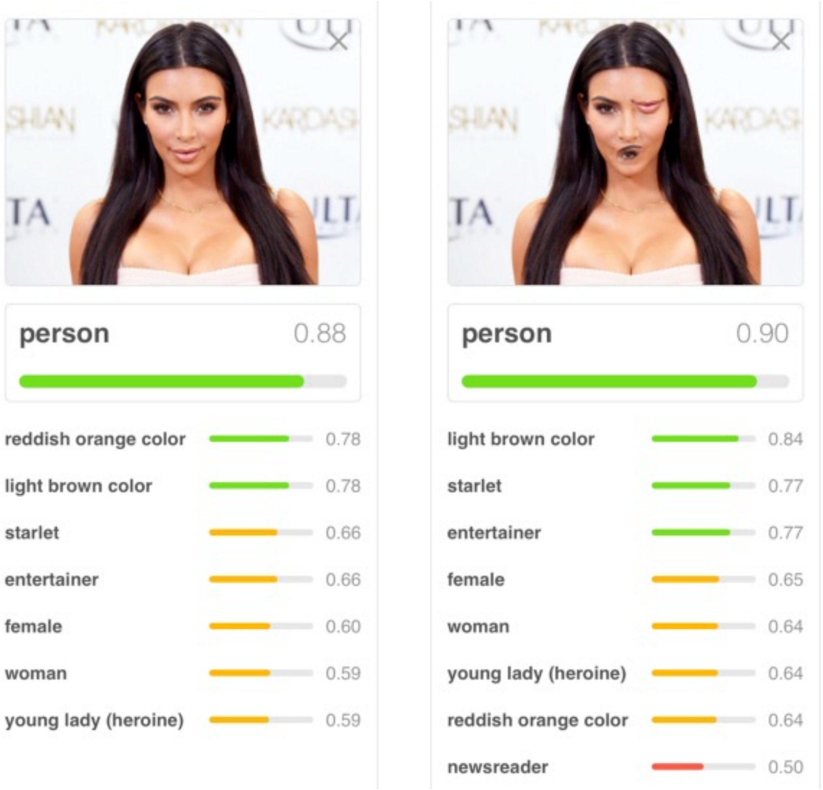
Facial Recognition Is Accurate, if You’re a White Guy

By STEVE LOHR FEB. 9, 2018

Architecture Deficiencies

(a) (b)

Fig. 1: Examples of the original image (left) and the negative image (right). a) An example from MNIST dataset, b) An example from GTSRB dataset. DNNs fail to recognize the negative images and classify them randomly into other classes.



An unmanned delivery vehicle drives into undried cement thinking the surface is wet.

Summary

- Machine learning and deep learning are growing in their usage
- Many of the ideas have been around for a long time but are being put into practice now because of technology progress
- Several opensource software resources (R, Rapid Miner, and Scikit-learn, PyTorch, TensorFlow etc.) to learn via experimentation
- Applications based on vision, speech, and natural language processing are excellent candidates for deep learning
- Need to filter hype from reality

"Machines will be capable, within twenty years, of doing any work that a man can do."

Herbert Simon, 1965

isethi@oakland.edu
https://iksinc.online/home/
iksinc@yahoo.com