MCSE 666:Pattern and Speech Recognition

Learning

(Biological, Machine Learning, Regression, and Classification)

Dr. Md. Aminul Haque Akhand

Dept. of CSE, SUB

Learning

Learning is the process of acquiring new or existing modifying knowledge, behaviors, skills, values, or preferences

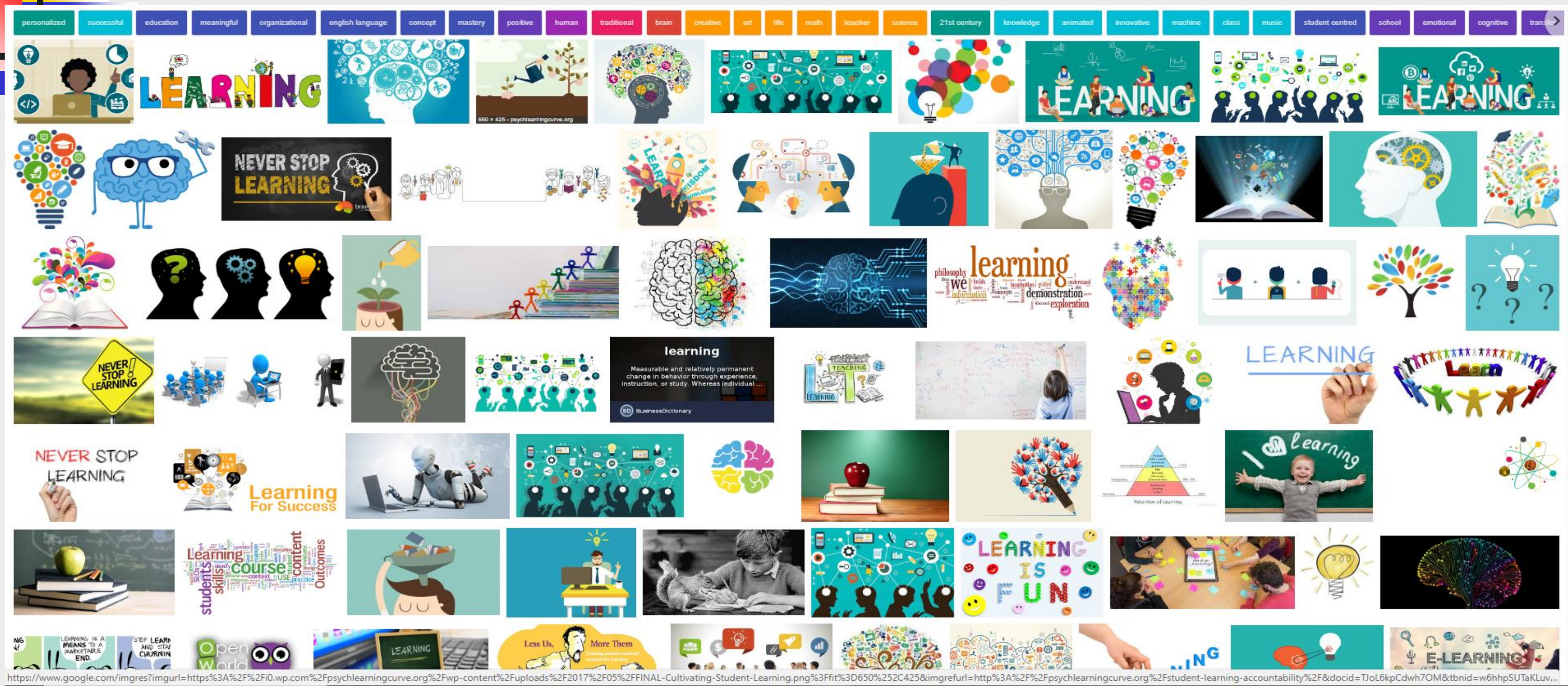
https://en.wikipedia.org/wiki/Learning

Learning is “a process that leads to change, which occurs as a result of experience and increases the potential for improved performance and future learning”
---(Ambrose et al, 2010, p.3).

“A change in human disposition or capability that persists over a period of time and is not simply ascribable to processes of growth.”

— ***From The Conditions of Learning by Robert Gagne***

Learning



The change in the learner may happen at the level of knowledge, attitude, or behaviour. As a result of learning, learners come to see concepts, ideas, and/or the world differently.

Learning is not something done to students, but rather something students themselves do. It is the direct result of how students interpret and respond to their experiences.

https://www.teachwithmrst.com/post/what-is-learning

Human brain is the main element of learning

Machine

A machine is a physical system using power to apply forces and control movement to perform an action. … Machines can be driven by animals and people, by natural forces such as wind and water, and by chemical, thermal, or electrical power ...They can also include computers and sensors that monitor performance and plan movement, often called mechanical systems. **https://en.wikipedia.org/wiki/Machine**

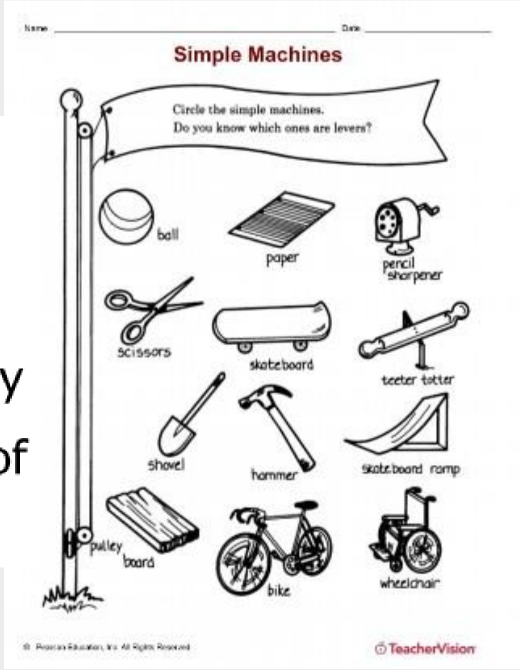
An apparatus using or applying mechanical power and having several parts, each with a definite function and together performing a particular task. *'a fax machine'*

Oxford Dictionary

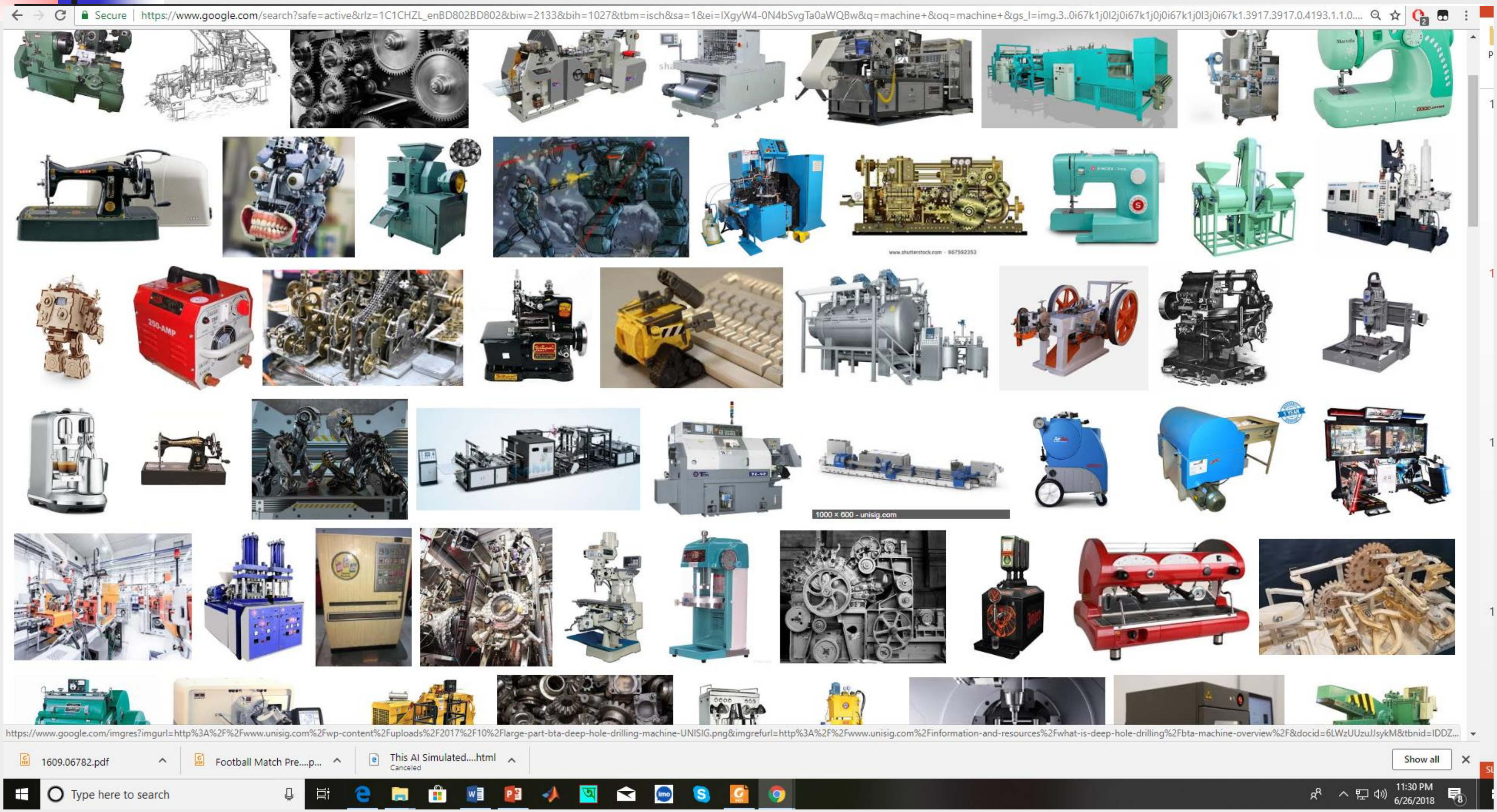
1.General: Semi or fully automated device that magnifies human physical and/or mental capabilities in performing one or more operations.

2.Mechanics: Device that makes mechanical work easier by overcoming a resistance (load) at one end by application of effort (force) at the other end.

BusinessDictionary.com



Machine



Machines include a system of mechanisms that shape the actuator input to achieve a specific application of output forces and movement.

How make decision to perform task(s)?

Machine Learning

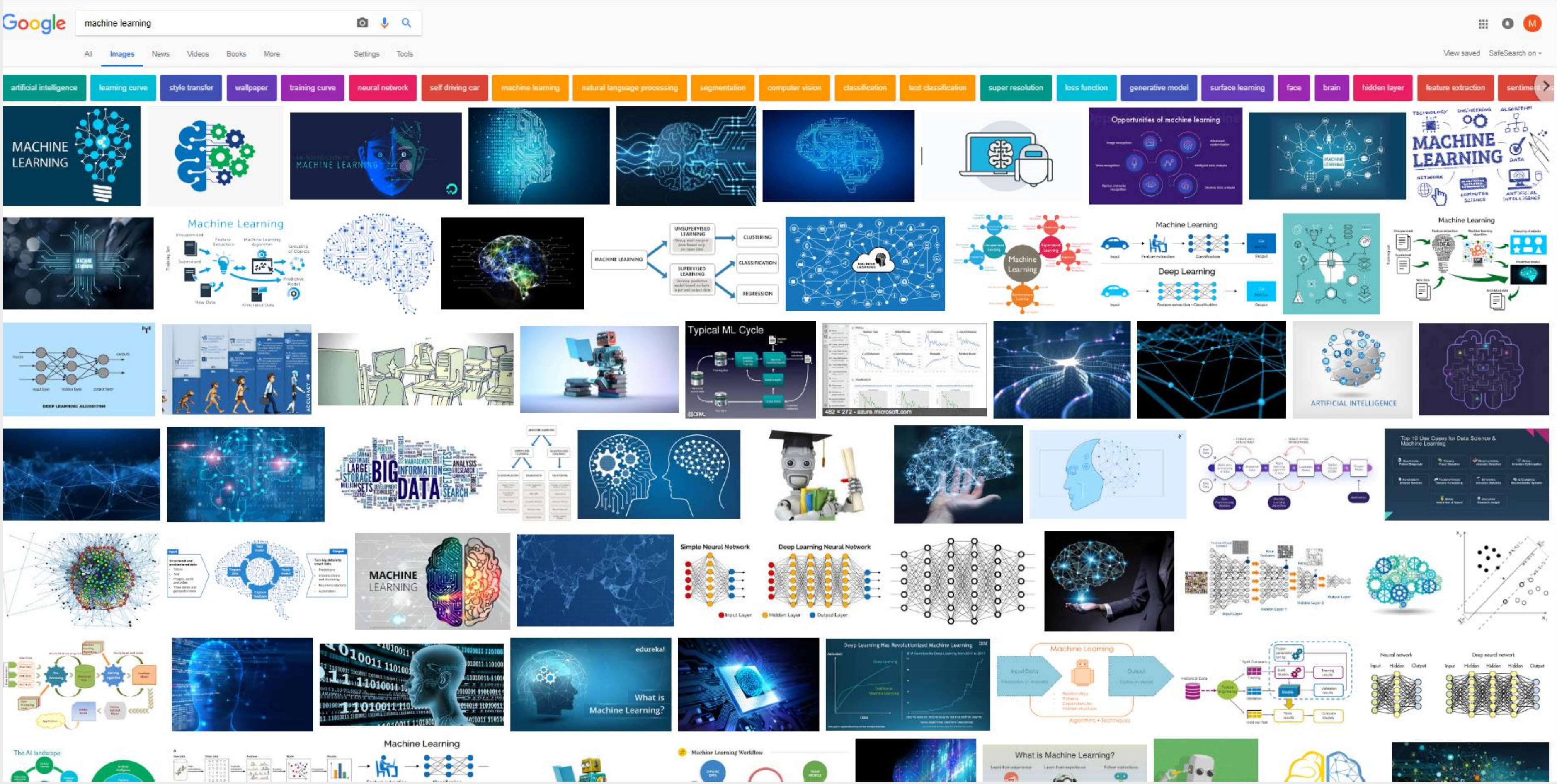
In 1959, Arthur Samuel, a pioneer in the field of machine learning (ML) defined it as the “field of study that gives computers the ability to learn without being explicitly programmed”.

https://theconversation.com/what-is-machine-learning-76759

Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data.[1] It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so.[2] Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.[3]

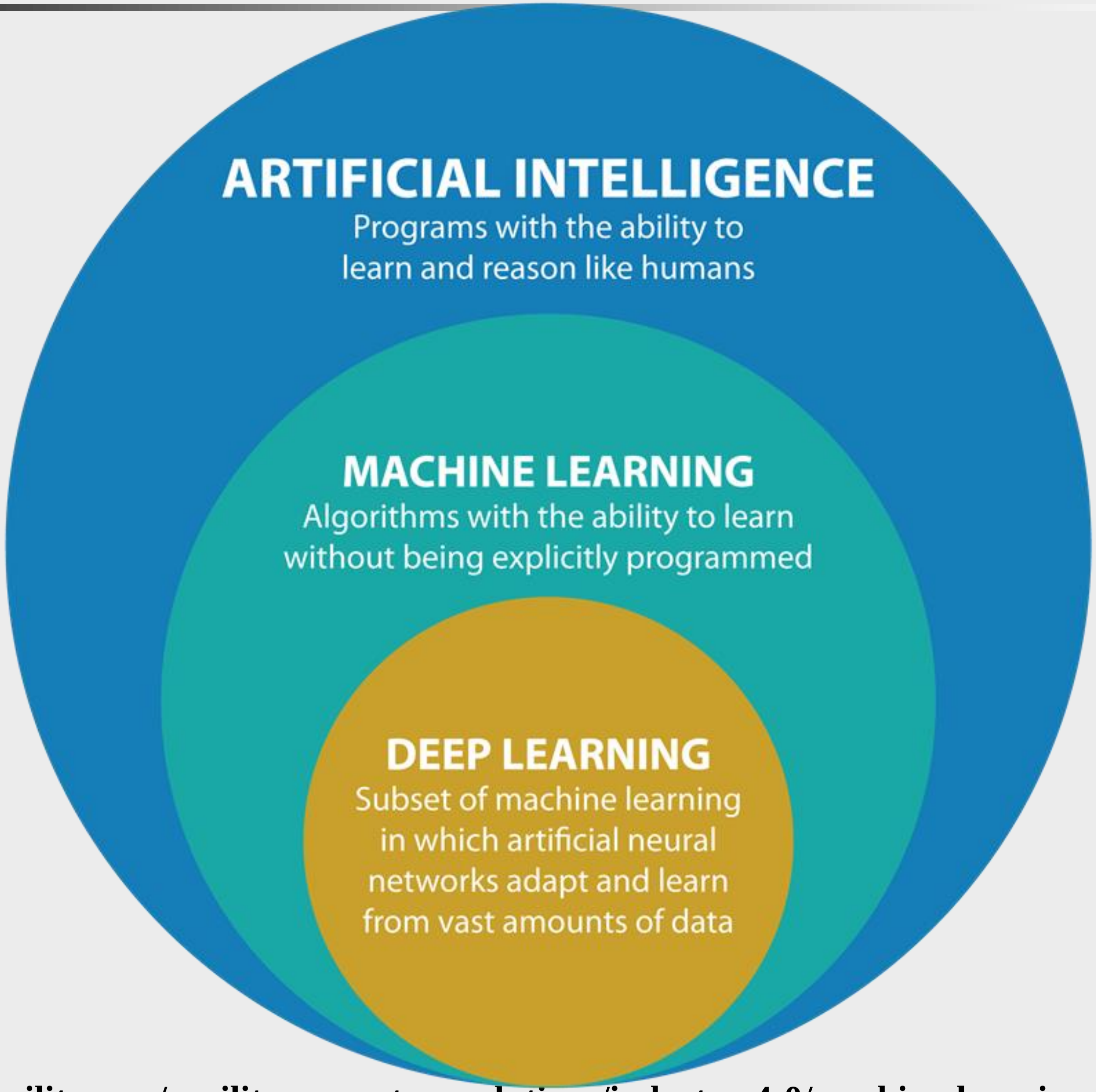
https://en.wikipedia.org/wiki/Machine_learning

Machine Learning



Technique to give computer brain like learning ability through progressively update with data, without being explicitly programmed.

AI ->Machine Learning->Deep Learning



https://www.argility.com/argility-ecosystem-solutions/industry-4-0/machine-learning-deep-learning/

AI ->Machine Learning->Deep Learning

Artificial Intelligence (AI)

AI is the broadest term, applying to any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning).

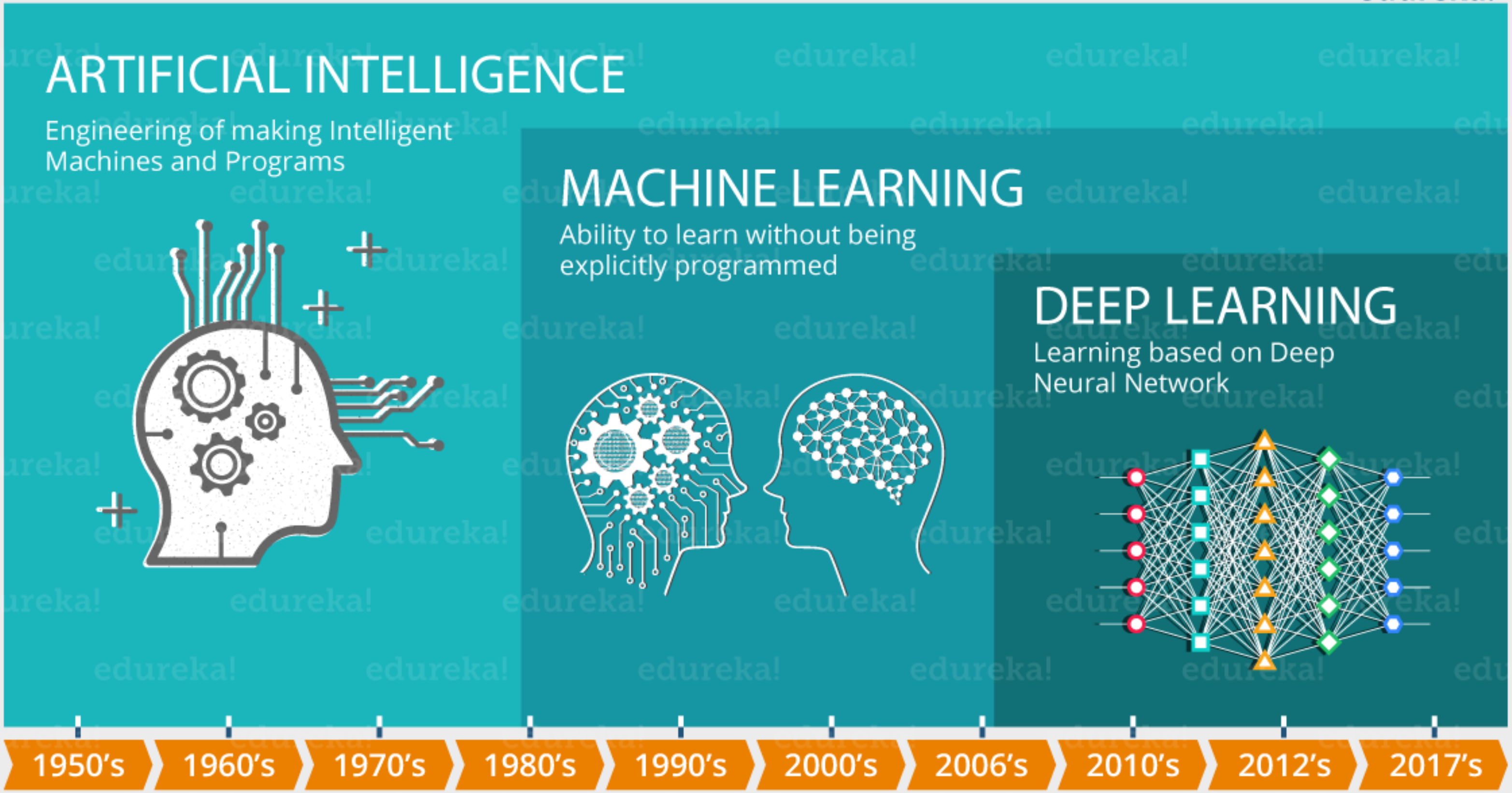
Machine Learning

The subset of AI that includes that enable machines to improve at tasks with experience. The category includes deep learning.

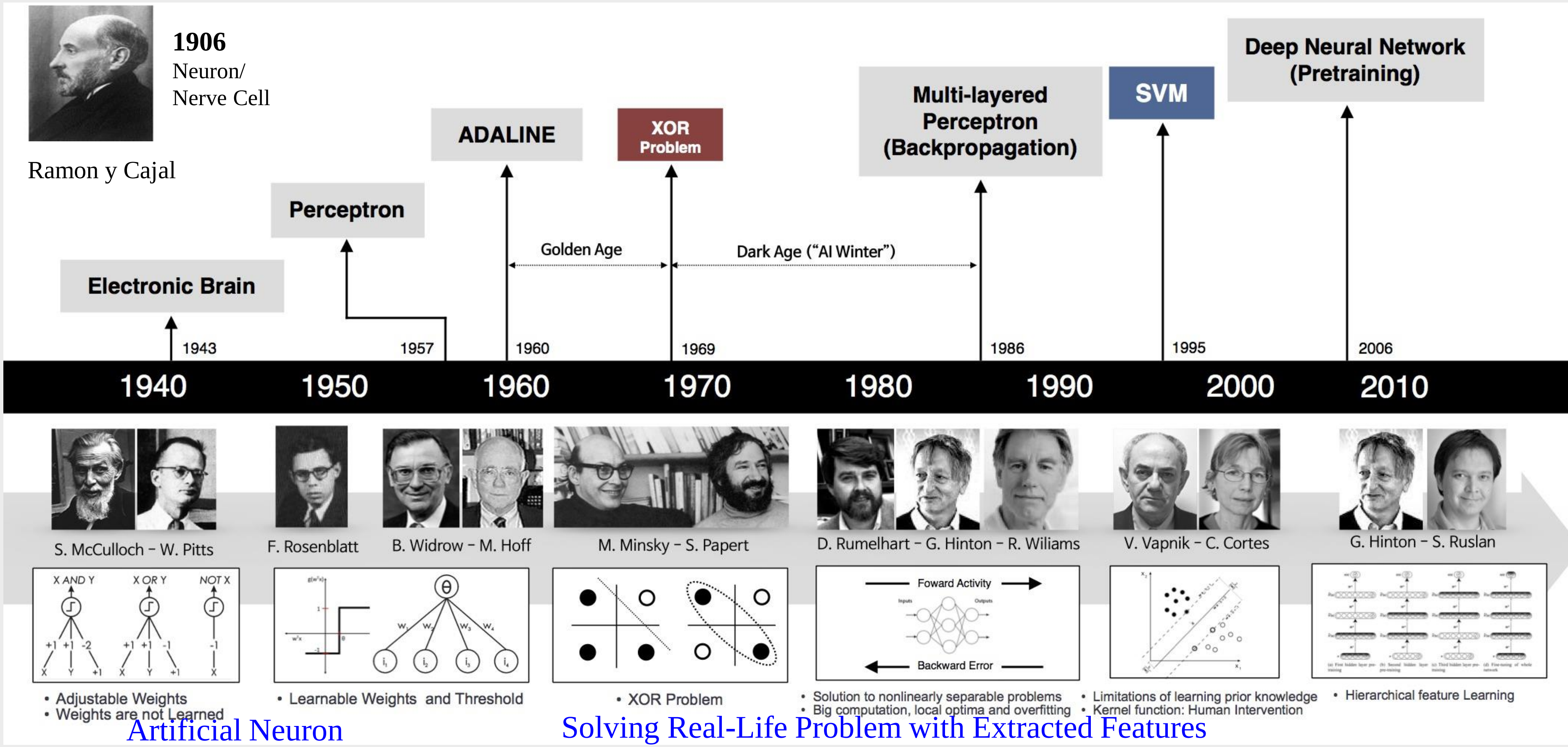
Deep Learning

The subset of machine learning composed of algorithms to train itself to perform tasks, like speech and image recognition, by exposing multilayered neural networks to vast amounts of data.

AI ->Machine Learning->Deep Learning



History of ML and Deep Learning



Artificial Neuron

Solving Real-Life Problem with Extracted Features

Development of Traditional Machine Learning

Deep Leaning

https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html

Learning Related to Pattern Recognition

Learning in Pattern Recognition

- *What Does It Mean to Learn?*
- *What does it mean to have learned?*
- How do we recognize that an algorithm or a method, regardless of how simple or complex, has succeeded in learning?

The object being learned is a classifier *g*(*x*), some sort of black-box / function / algorithm / method / concept / system which takes a given feature *x* and returns the associated class *C*:

$$g(\underline{x}) \in \mathcal{C} = \{C_1, C_2, \ldots, C_K\}$$

We suppose that we are given a dataset $\mathcal{D}$ of N feature-class[1] pairs

$$\mathcal{D} = \{(\underline{x}_i, c_i)\} \qquad \text{where} \quad \underline{x}_i \in \mathbb{R}^n,\ c_i \in \mathcal{C},\ 1 \leq i \leq N$$

Learning in Pattern Recognition

The most primitive type of learning is just memorization, in which case we would consider *g*() to have learned from the given dataset based on the number of feature-class pairs it successfully reproduces:

$$\text{Correct Count} = \sum_i \delta(c_i, g(\underline{x}_i)) \qquad \delta(a,b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}$$

In principle, memorization is actually a credible approach to developing a classifier, however in general there are two significant limitations:

1. Memorizing is hard : memory concern
2. Memorizing is not enough: learning is not just to remember

#Generalize is more important to learning, to be able to reach correct conclusions about instances which you have *not* seen.

$$\hat{\underline{\theta}} = \arg_{\underline{\theta}} \max \sum_i \delta(c_i, g(\underline{x}_i, \underline{\theta}))$$

Learning in Pattern Recognition

True Class			Classification			
c_1	c_2	c_3	$g(\underline{x}_1) = \text{Dog}$	$g(\underline{x}_2) = \text{Dog}$	$g(\underline{x}_3) = \text{Cat}$	
Dog	Dog	Dog	✓	✓	✗	
Dog	Dog	Cat	✓	✓	✓	
Dog	Cat	Dog	✓	✗	✗	
Dog	Cat	Cat	✓	✗	✓	
Cat	Dog	Dog	✗	✓	✗	
Cat	Dog	Cat	✗	✓	✓	
Cat	Cat	Dog	✗	✗	✗	
Cat	Cat	Cat	✗	✗	✓	
			50%	50%	50%	← Average Performance

Fig. 3.1. NO FREE LUNCH: There is no objectively best classifier, averaged over all possible outcomes all classifiers perform the same as random guessing. A given classifier $g()$ is asked to classify three previously-unseen features $\underline{x}_1, \underline{x}_2, \underline{x}_3$ having true associated classes c_1, c_2, c_3, where the features are classified into two possible classes of *Cat* and *Dog*. Averaged over all possible truths (left), the classifier $g()$ does no better than random guessing. Indeed, averaged over all truths, *every* classifier will perform the same.

Averaged over all possibilities for the unseen data, no classifier generalizes better than any other!

Robustness in Learning

$$g(x, \underline{\theta}) = \theta_p x^p + \theta_{p-1} x^{p-1} + \ldots + \theta_1 x^1 + \theta_0 x^0.$$

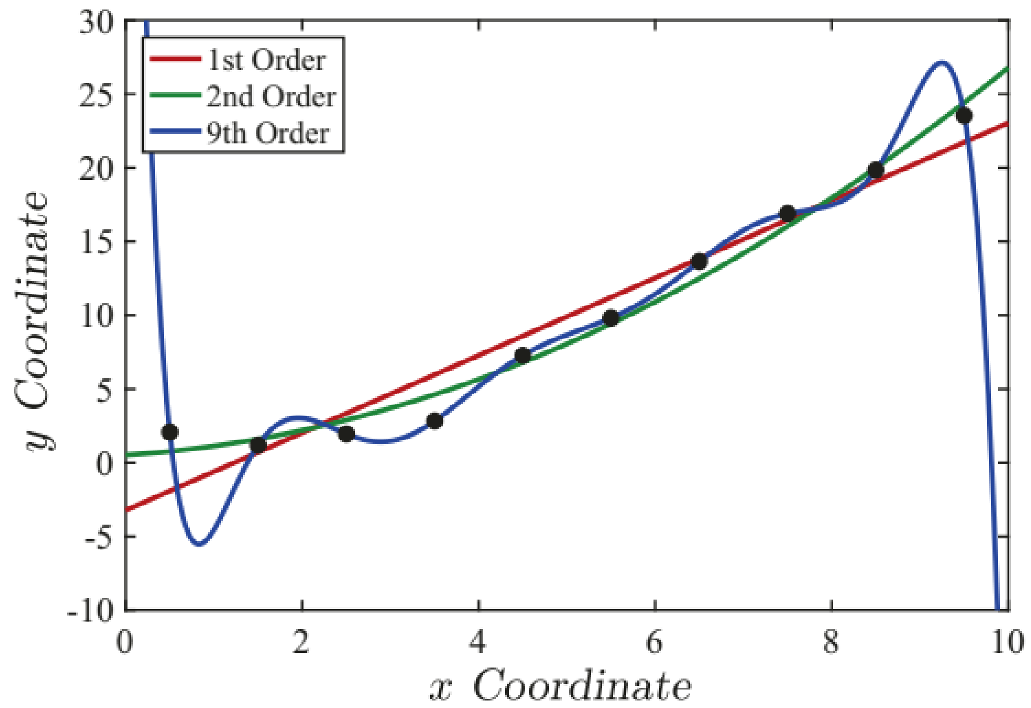
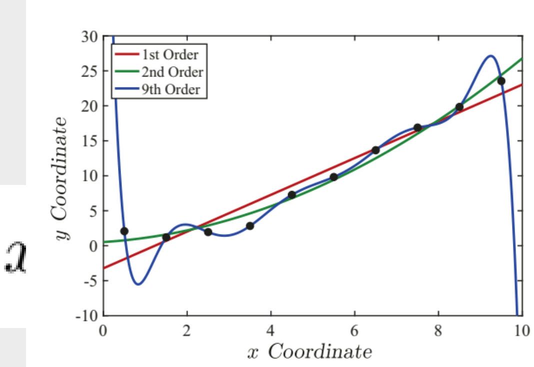


Fig. 3.2. What does it mean to *fit* a model to data? Here we have ten data points (black dots), which come from some unknown model with added random noise. We fit polynomials to the points, for which the first-order (linear regression), second-order (parabolic regression), and ninth-order fits are shown. A *p*th-order polynomial can always be found that passes through $p + 1$ given points, so here the ninth-order polynomial fits the points exactly, however it seems like a very unlikely generalization of the data.

Robustness in Learning

$$g(x, \underline{\theta}) = \theta_p x^p + \theta_{p-1} x^{p-1} + \ldots + \theta_1 x$$



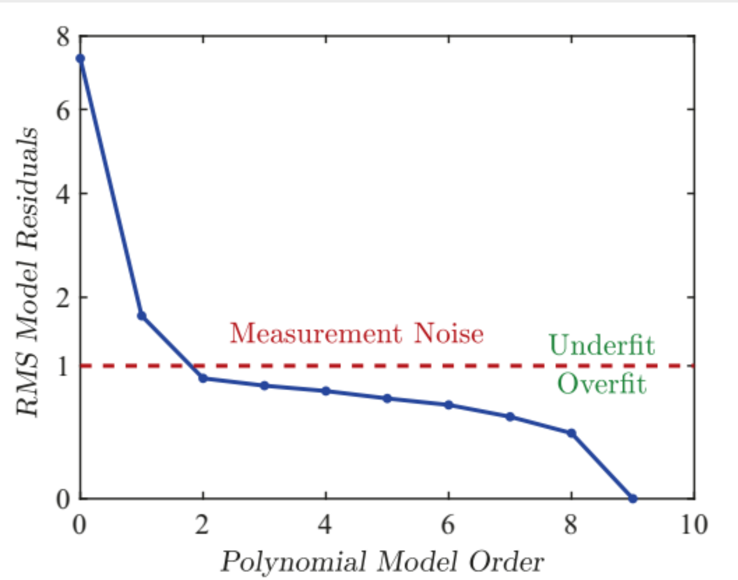


Fig. 3.3. Following up on Figure 3.2, we can plot the root-mean-square (RMS) difference between the polynomial fit and the data points, as a function of the polynomial order. In this case the measurement noise level was assumed to have been provided, it was not learned.

$$\mathrm{RMS}(\underline{\theta}) = \left(\frac{1}{N} \sum_{i=1}^{N} \left(y_i - g(x_i, \underline{\theta})\right)^2 \right)^{1/2},$$

In this example the added noise has a standard deviation of σ = 1, meaning that for the correct model, RMS(θexact) = 1. As a result,

• Any RMS difference below σ must be overfitting, meaning that g(x; θ) is partly fitting to noise, by taking some of the noise into account when learning θ;

• Any RMS difference above σ suggests that the learned model has not adequately generalized, or has not been given adequate flexibility in θ (enough degrees of freedom q) to capture the variations that need to be captured.

Robustness in Learning

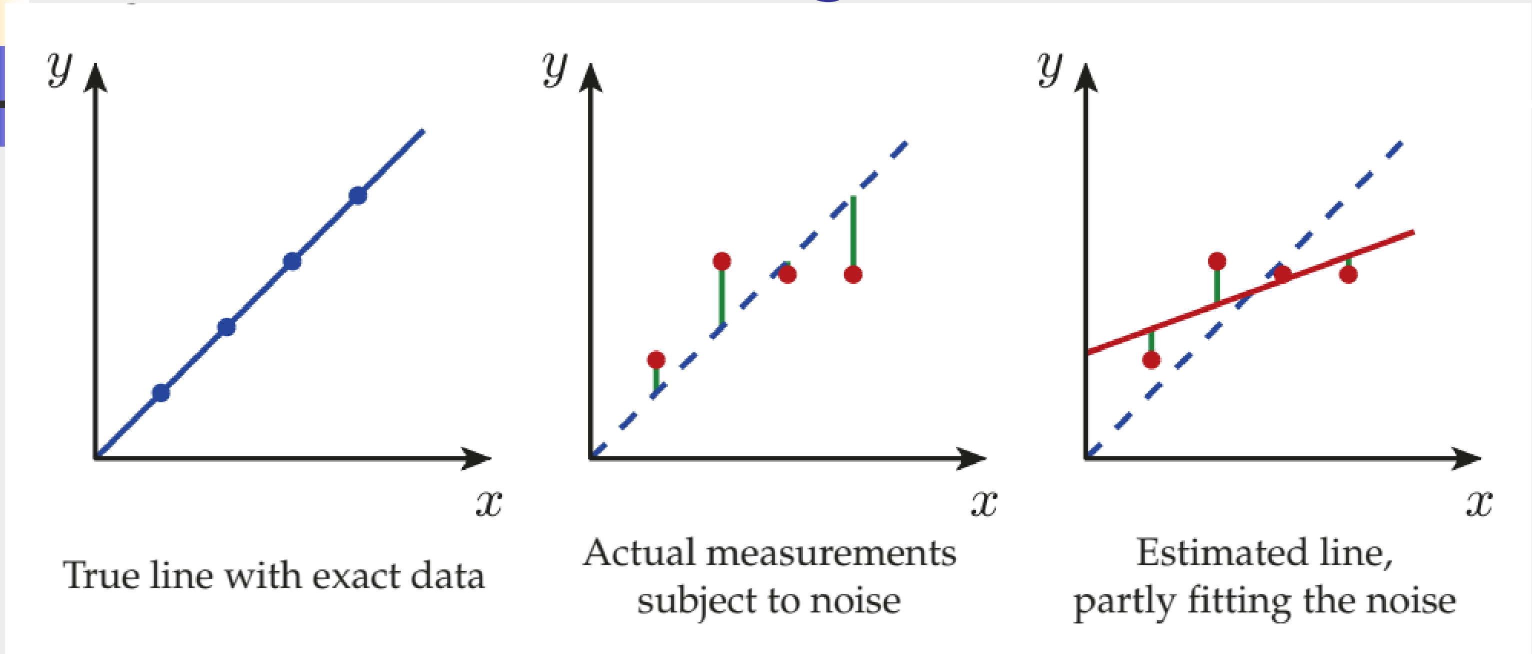
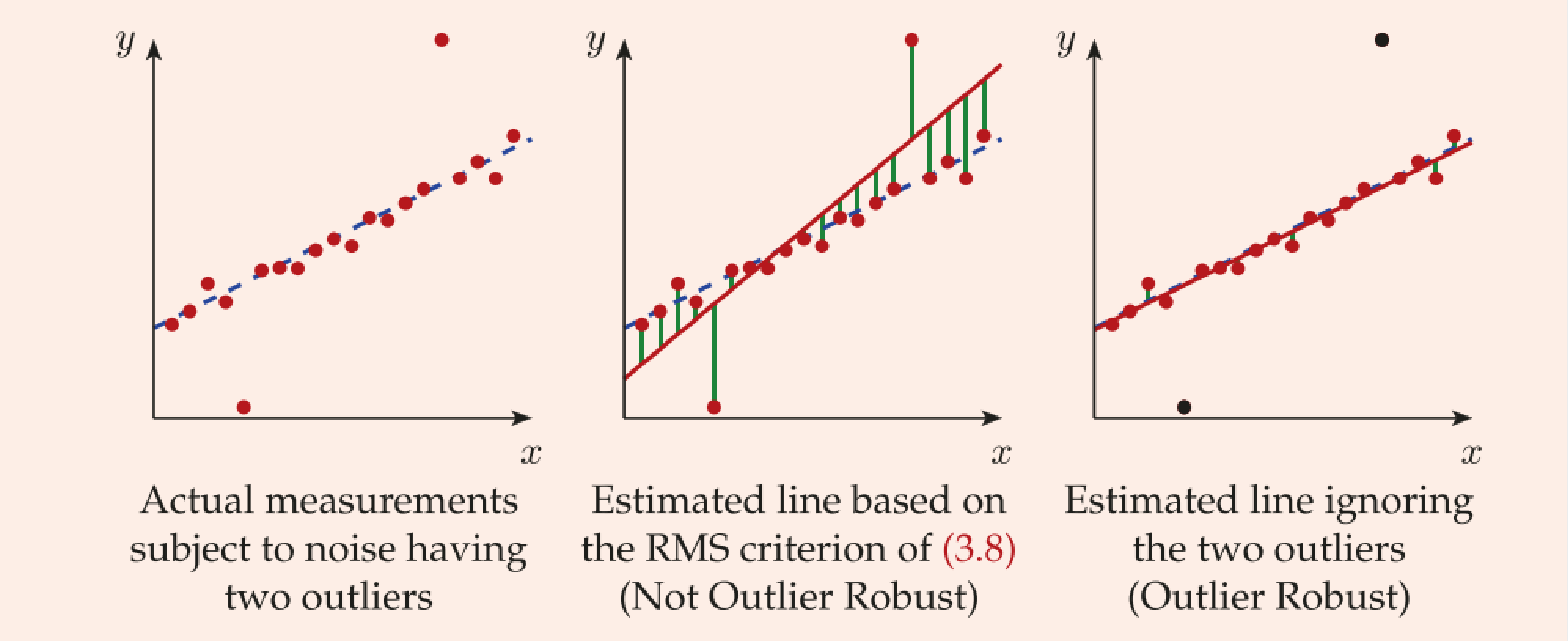


Fig. 3.4. Overfitting: Any learning, whether of linear regression (here) or a pattern recognition classifier (Figure 3.5), is said to be overfitting if it begins to tune its parameters to the behaviour of the noise, rather than of the underlying phenomenon we wish to learn. The estimated line (red) is quite plausible, given the four data points (red dots), however it is clear how the line has accommodated (fit) the noise, to make the residuals (green) smaller

Robustness in Learning



So how do we make learning *robust* to outliers? Really this is a vast topic, which we can only begin to touch here. A variety of approaches is possible:

1. Detect and remove the outliers (as was done in the right-most panel, above),
2. Choose parameters in $\underline{\theta}$ insensitive to outliers, or
3. Choose an optimization metric insensitive to outliers.

Robustness in Learning

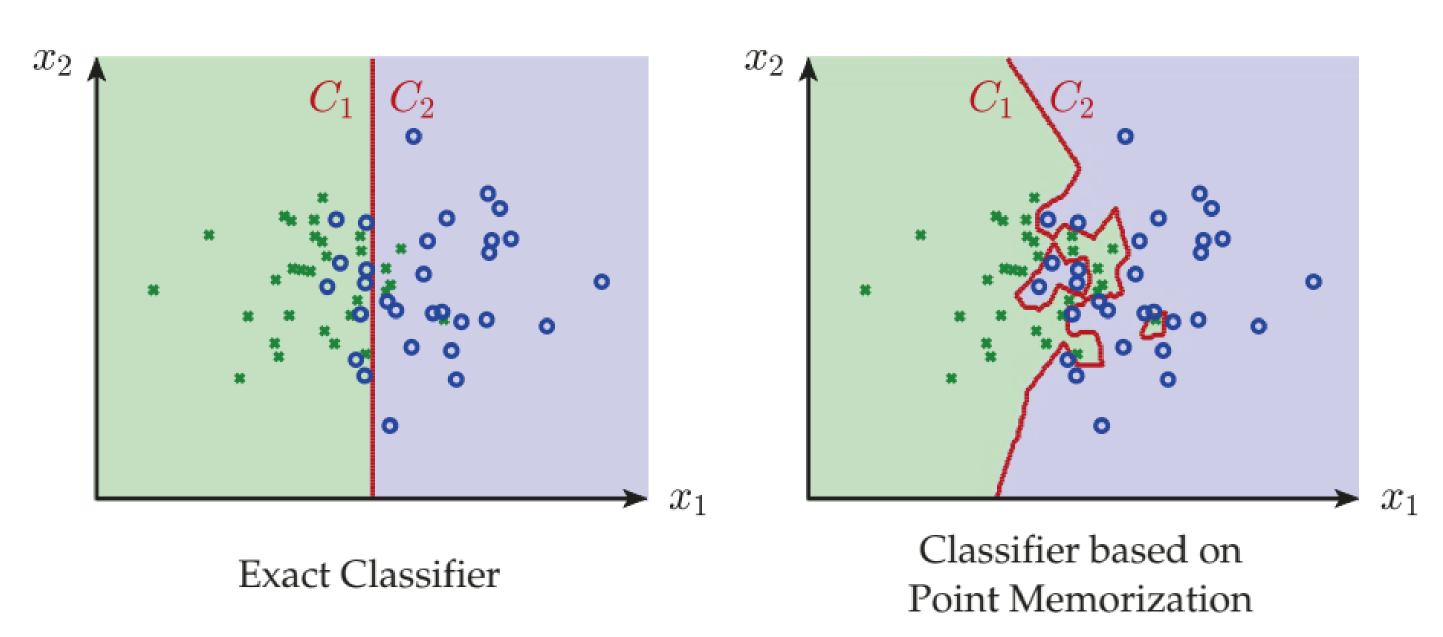
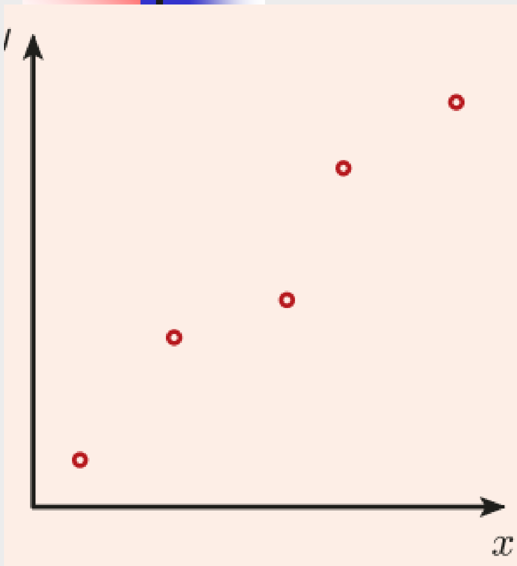
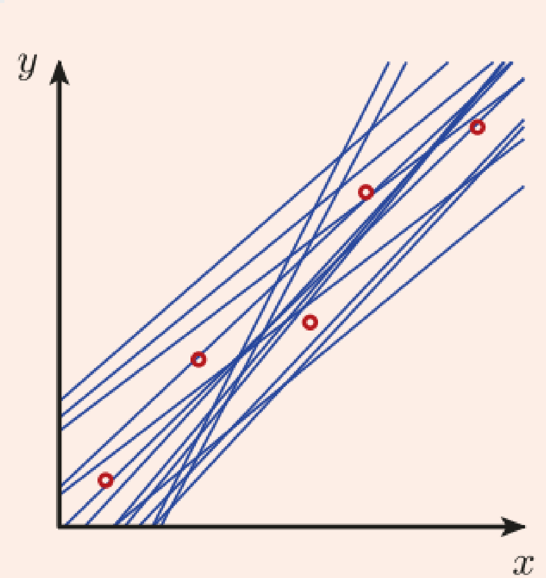


Fig. 3.5. OVERFITTING: As in Figure 3.4, but here for a pattern recognition classifier. We will have to wait for Chapter 6 for the details of the classifier to be discussed, however the principle is the same as in regression: Any learning is overfitting if it tunes its parameters to the behaviour of the noise. That tuning is obvious here, in that the memorized classifier (right) is tuning its decision (coloured background) on the basis of individual training points, significantly distracted from the correct or ideal classification (left).

Regression and Classification



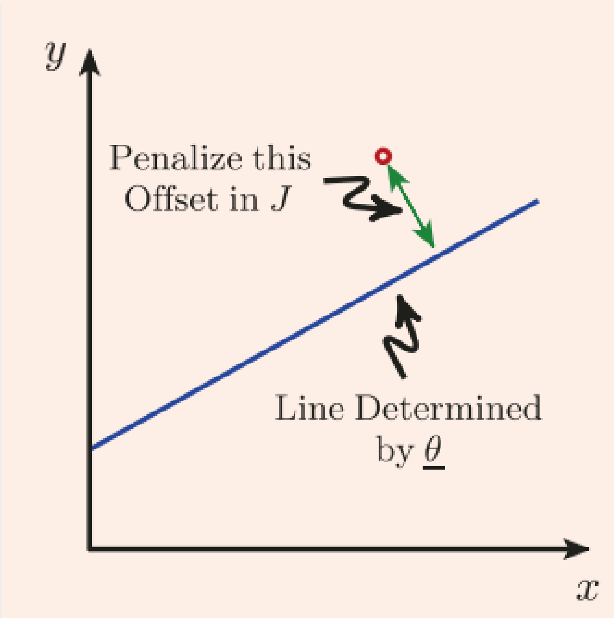


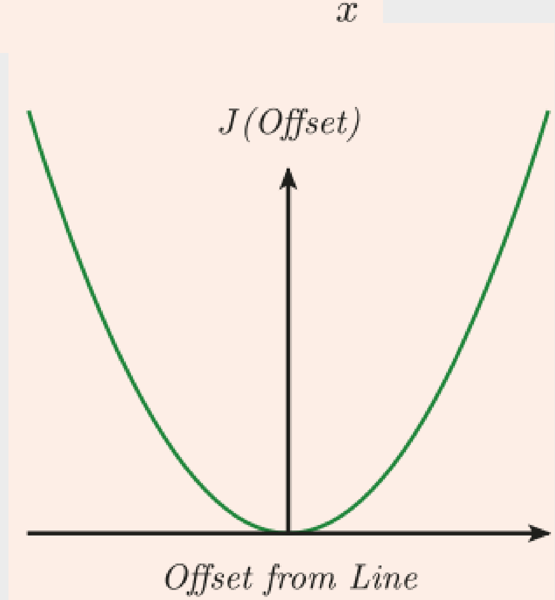
For each possible line, described by

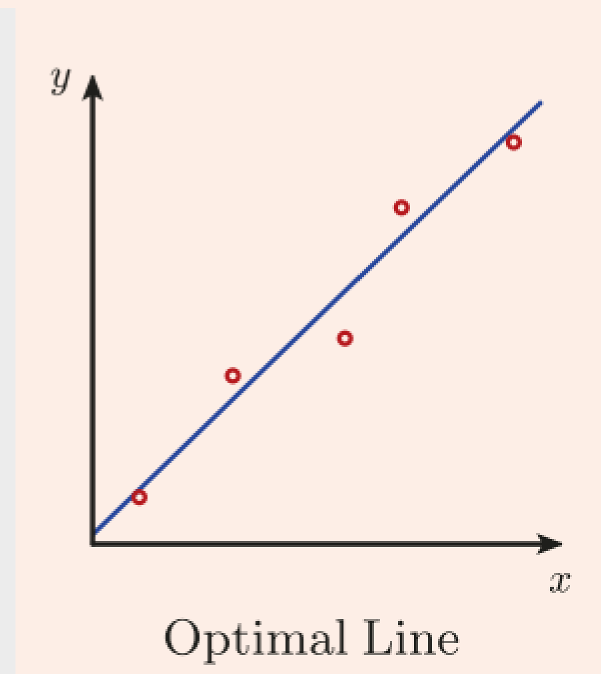
$$\underline{\theta} = \begin{bmatrix} \text{Angle of Line} \\ \text{Y-Intercept of Line} \end{bmatrix}$$

we can assess the penalty associated with the
as

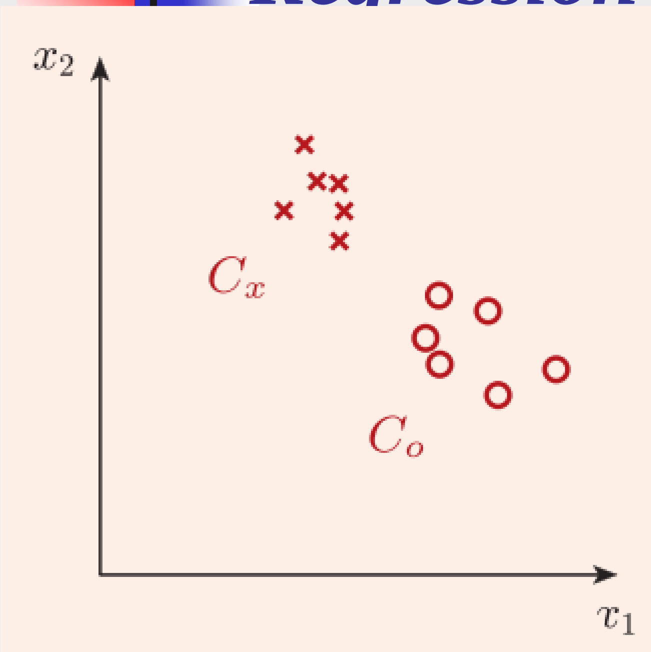
$$\text{Penalty}(\underline{\theta}) = \sum_i J\big(\text{Offset from line } \underline{\theta} \text{ to } (x_i,$$

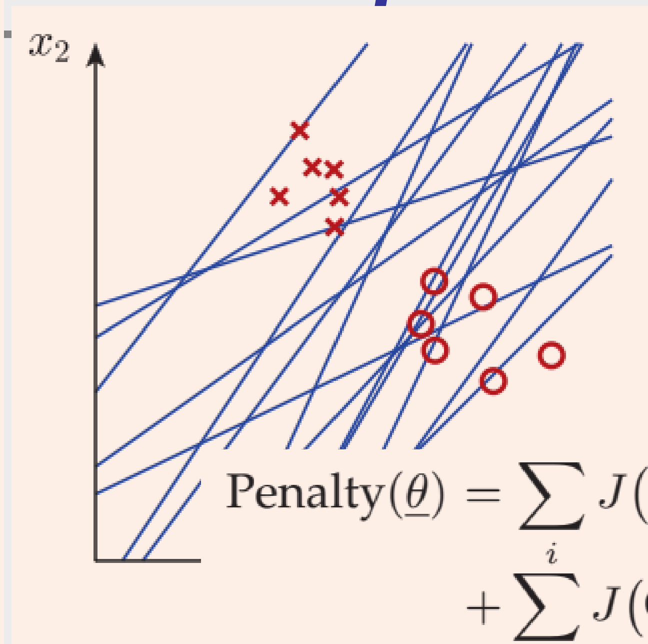




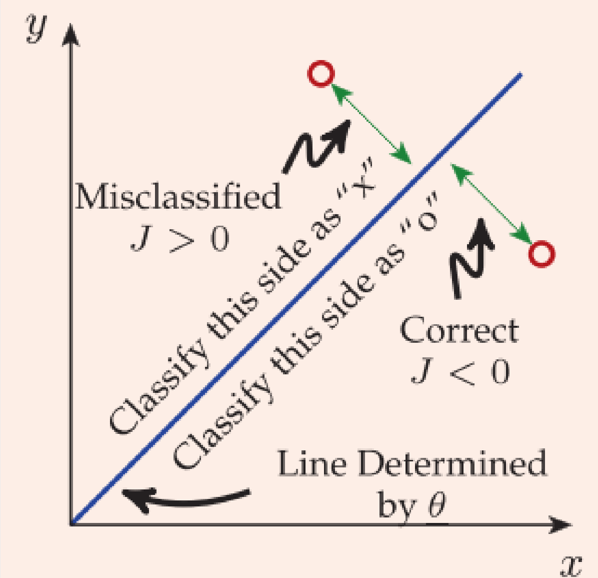


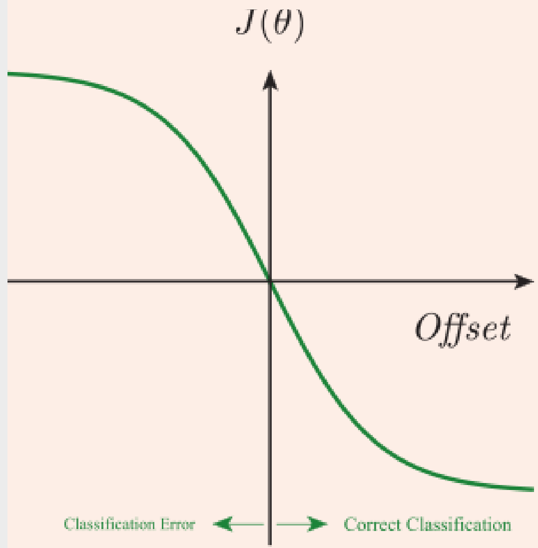
Regression and Classification

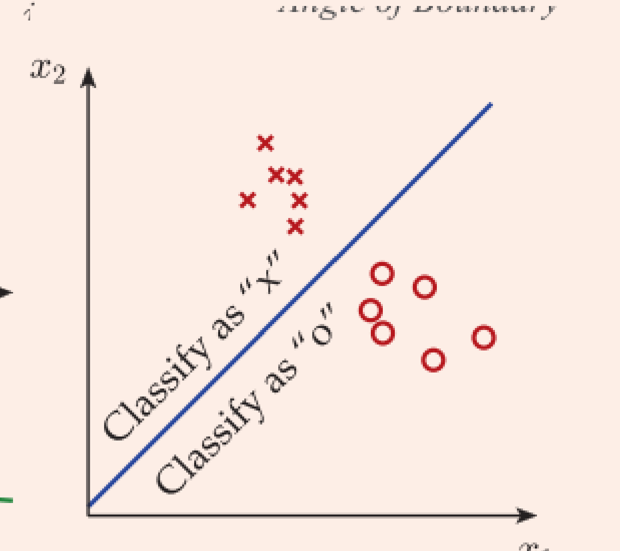




$$\text{Penalty}(\underline{\theta}) = \sum_i J\big(\text{Offset from line } \underline{\theta} \text{ to } \underline{x}_i \in C_x\big) + \sum_j J\big(\text{Offset from line } \underline{\theta} \text{ to } \underline{x}_j \in C_o\big)$$







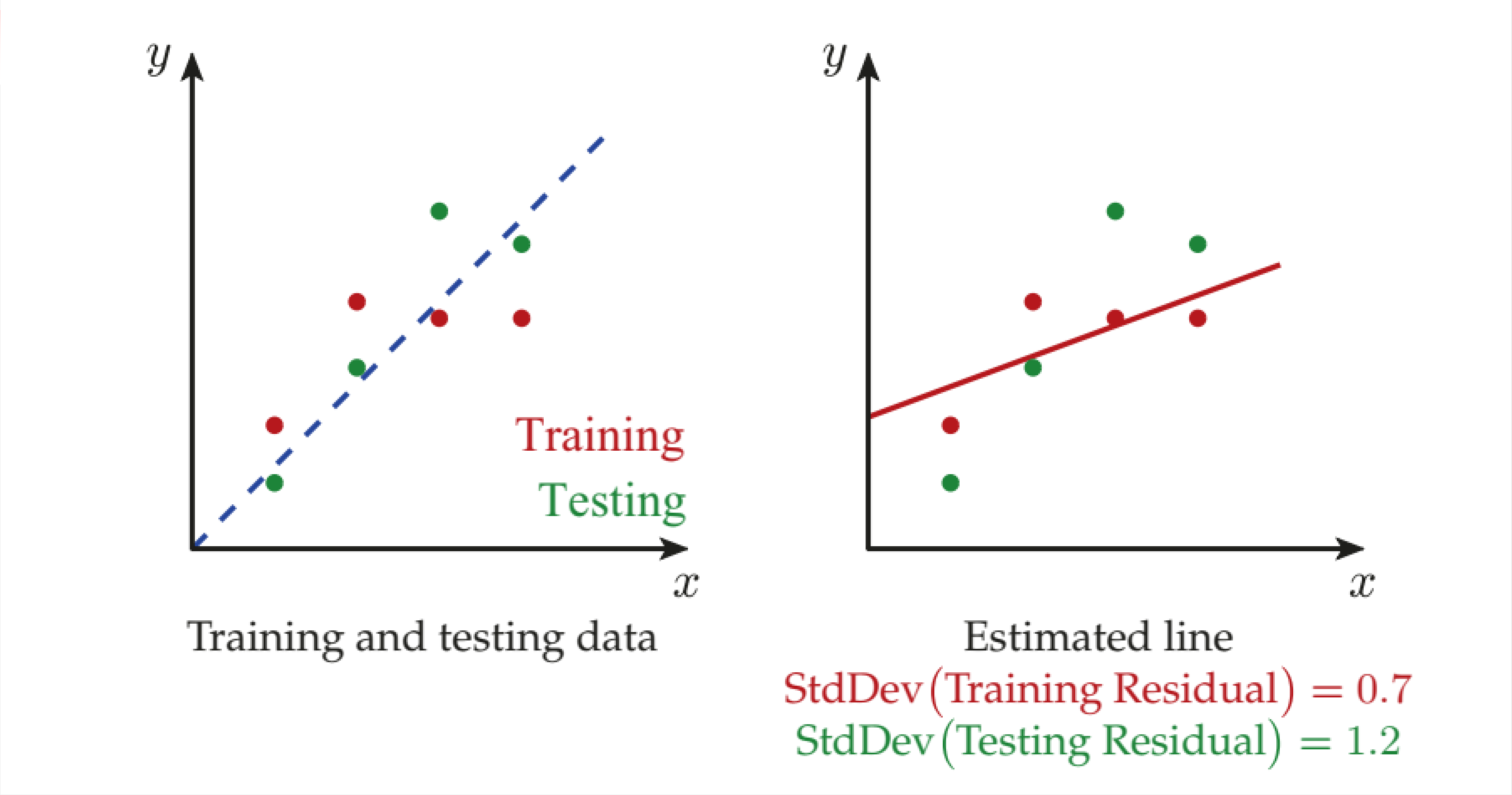
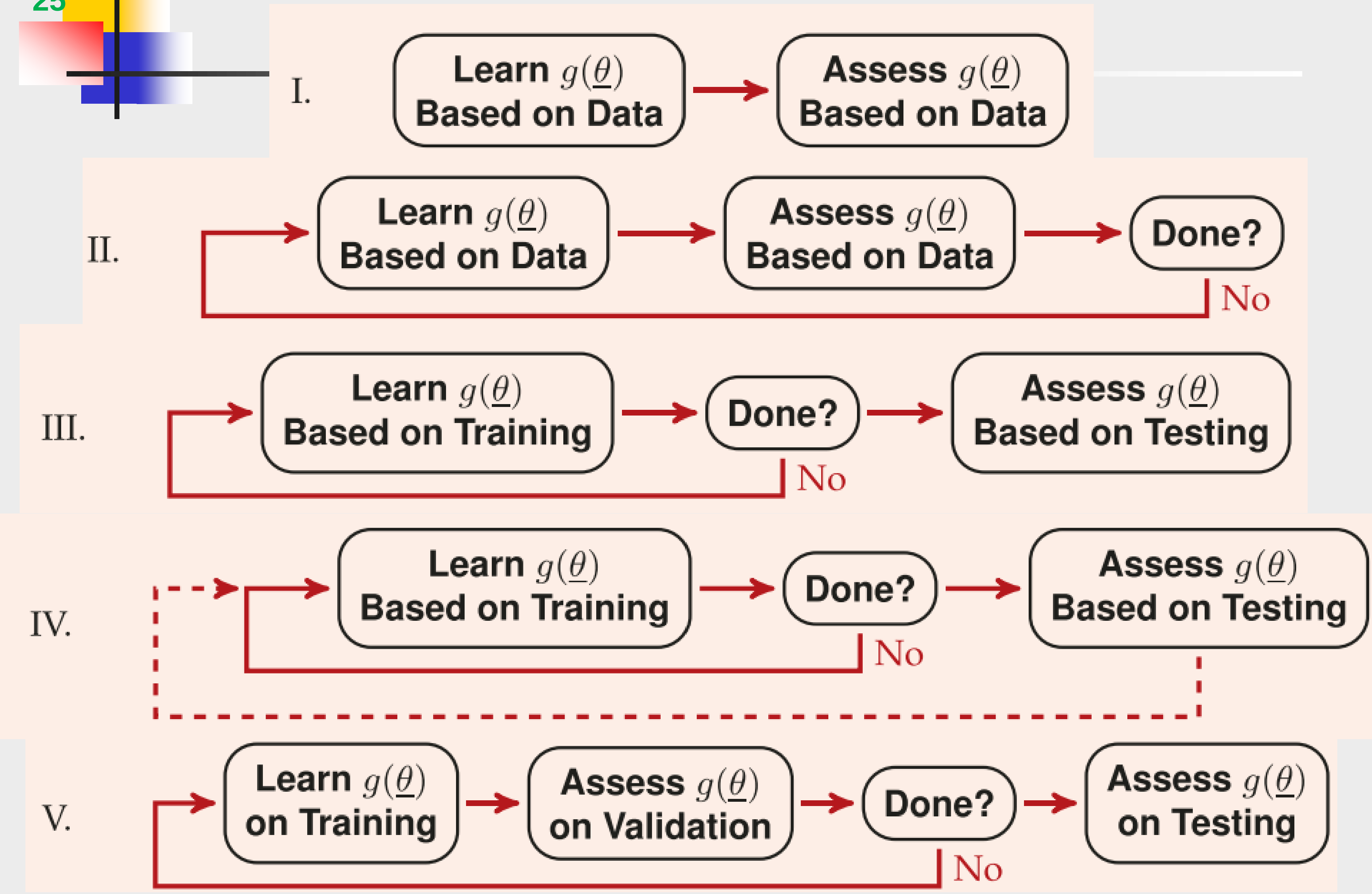


Fig. 3.7. We have separate training and testing data (left), such that the learned model (red line) is deduced from the training data, but assessed against the testing data. Observe the degree to which the estimated line fits to the noise, based on the difference between the fit to training data (overfit, under-reporting model inconsistency) and the fit to testing data (which is an accurate, objective assessment).

Use of Data in Learning



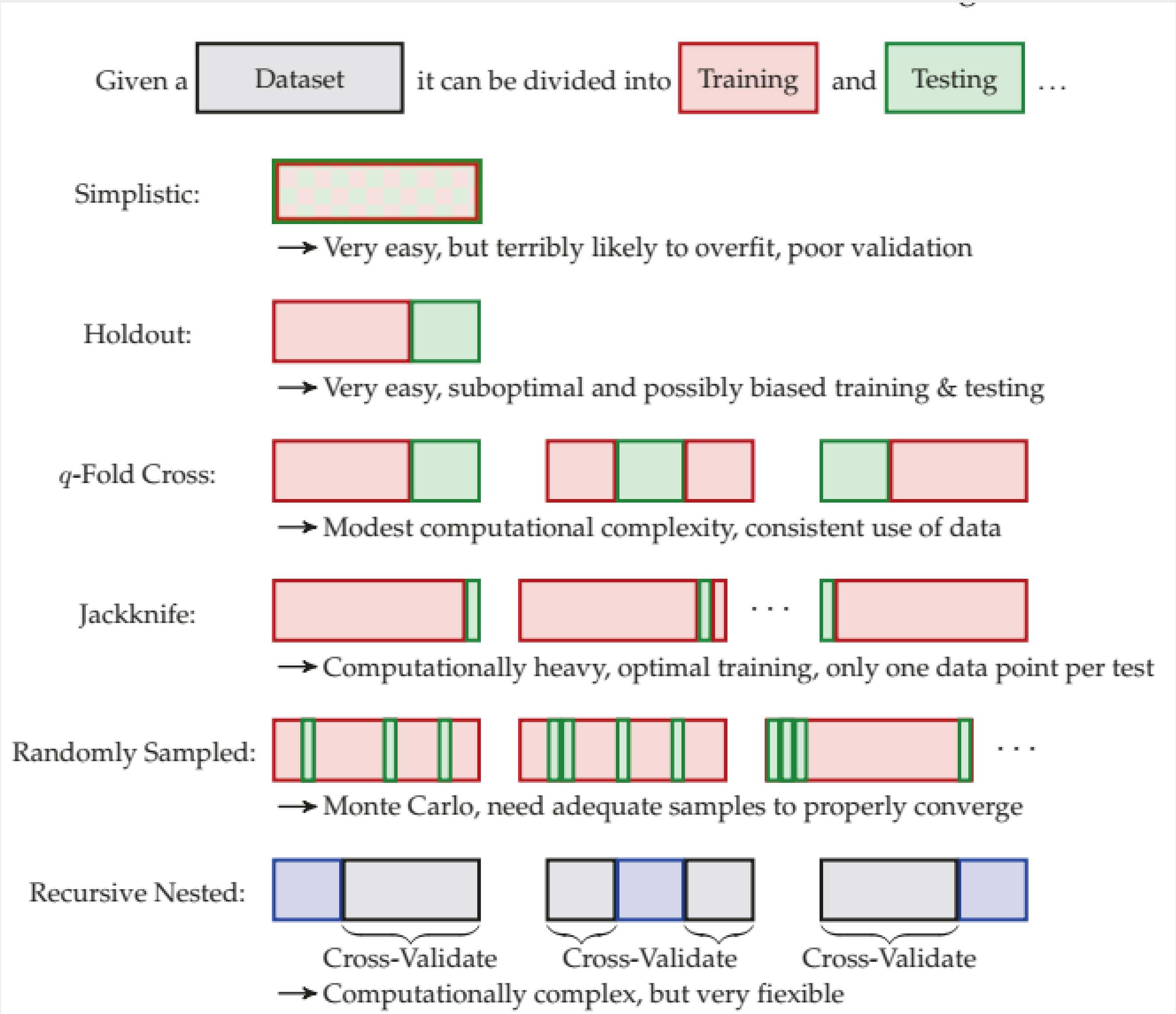
Classifier Evaluation / Performance Measure

1. **Test Set Accuracy**
2. **Test Set Error Rate**
3. **Confusion Matrix**

Given a classifier g, the corresponding confusion matrix from (3.19),

$$
\begin{array}{cc|cccc|}
 & & \multicolumn{4}{c}{\text{Is classified as} \ldots} \\
 & & C_1 & C_2 & \cdots & C_K \\
\hline
 & C_1 & S_g(C_1|C_1) & S_g(C_2|C_1) & \cdots & S_g(C_K|C_1) \\
\text{True Class} & \vdots & \vdots & \vdots & & \vdots \\
 & C_K & S_g(C_1|C_K) & S_g(C_2|C_K) & \cdots & S_g(C_K|C_K) \\
\hline
\end{array}
$$

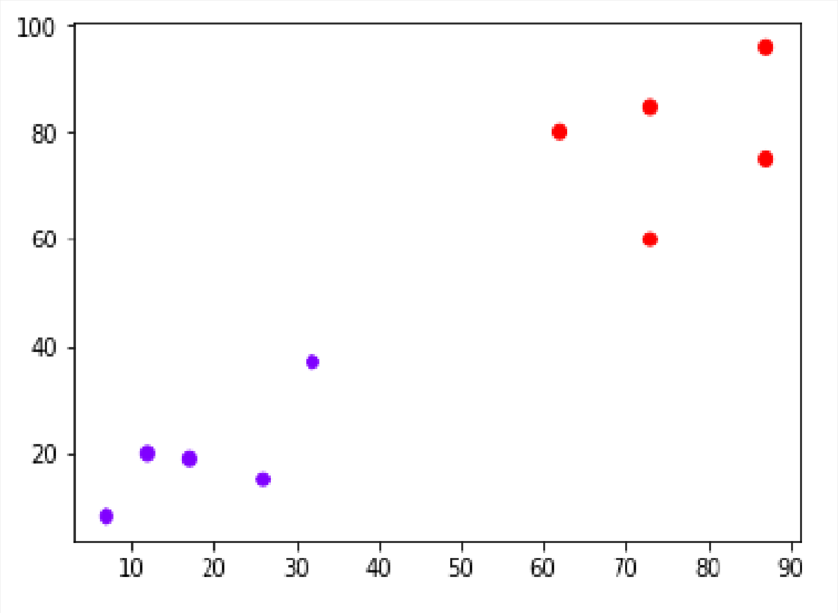
Classifier Validation

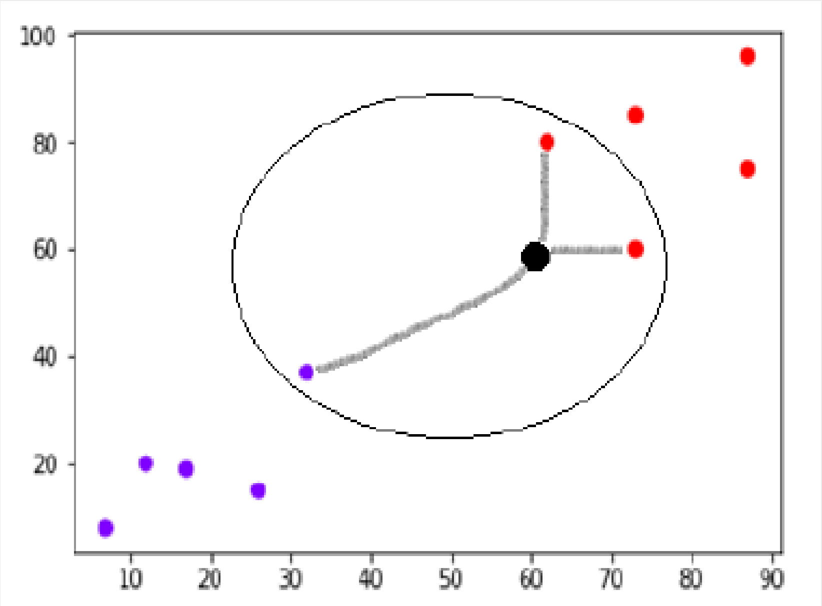


K-nearest neighbors (KNN) Classifier

•**Lazy learning algorithm** − KNN is a lazy learning algorithm because it does not have a specialized training phase and uses all the data for training while classification.

•**Non-parametric learning algorithm** − KNN is also a non-parametric learning algorithm because it doesn’t assume anything about the underlying data.

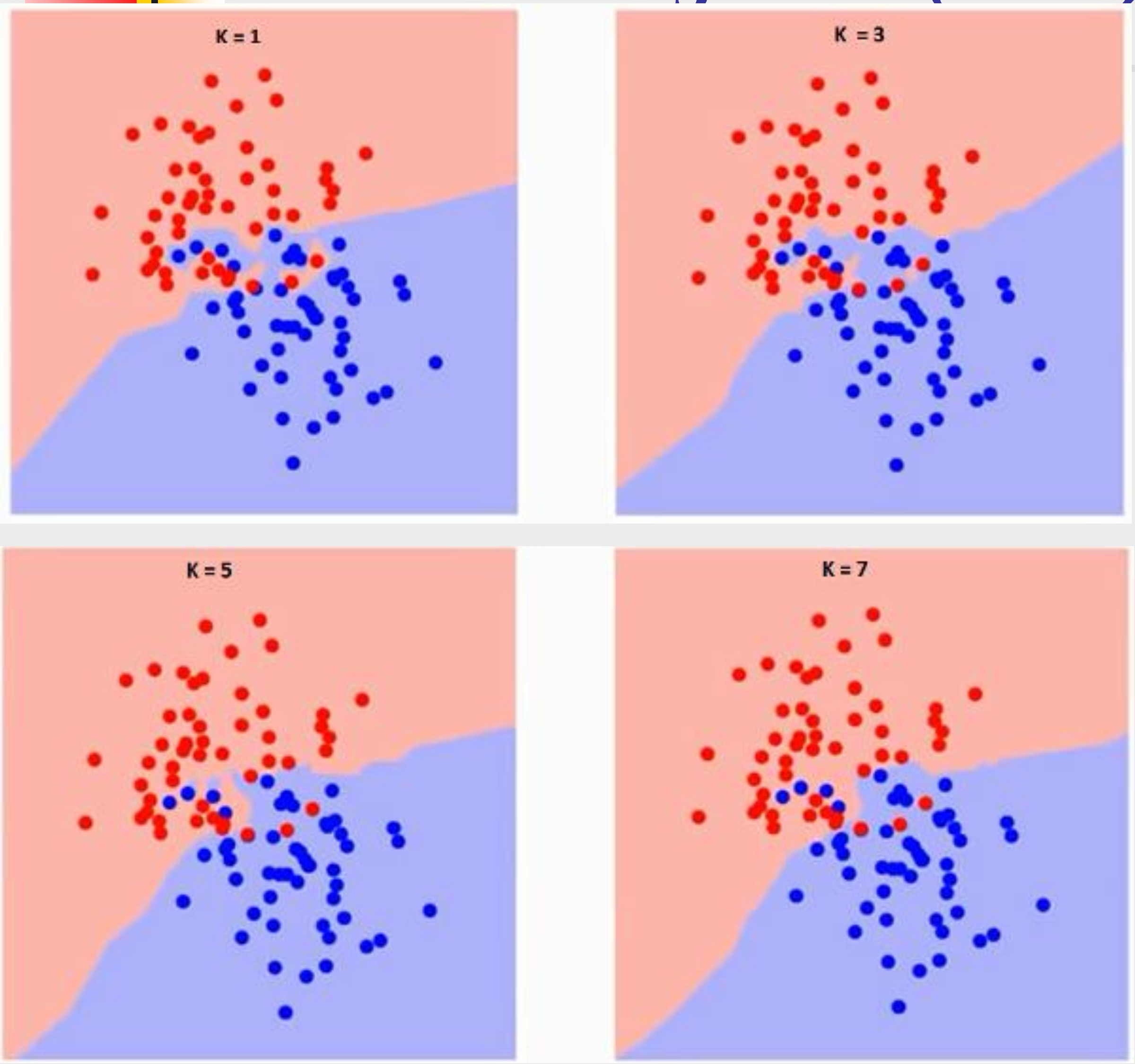




https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm

K-nearest neighbors (KNN) Classifier
- The Simplest Classifier

K-nearest neighbors (KNN) Classifier



The boundary becomes smoother with increasing value of K.

With K increasing to infinity it finally becomes all blue or all red depending on the total majority.

https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/#h-what-is-knn-k-nearest-neighbor-algorithm

K-nearest neighbors (KNN) Classifier

Pros of KNN

- Very simple algorithm to understand and interpret.
- Very useful for nonlinear data because there is no assumption about data in this algorithm.
- Versatile algorithm as we can use it for classification as well as regression.
- High accuracy but there are much better supervised learning models than KNN.

Cons of KNN

- Computationally a bit expensive algorithm because it stores all the training data.
- High memory storage required as compared to other supervised learning algorithms.
- Prediction is slow in case of big N.
- Very sensitive to the scale of data as well as irrelevant features.

Applications of KNN

Banking System: to predict weather an individual is fit for loan approval? Does that individual have the characteristics similar to the defaulters one?

Calculating Credit Ratings: can be used to find an individual's credit rating by comparing with the persons having similar traits.

Other areas in which KNN algorithm can be used are Speech Recognition, Handwriting Detection, Image Recognition and Video Recognition.

Thanks for your attention

Question and Answer