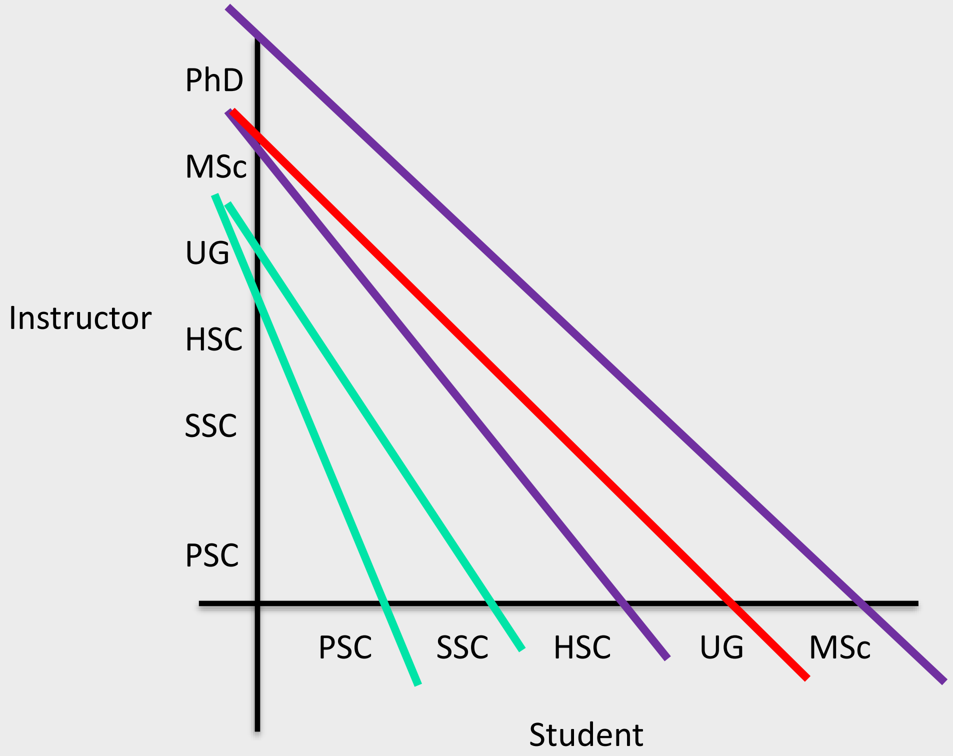
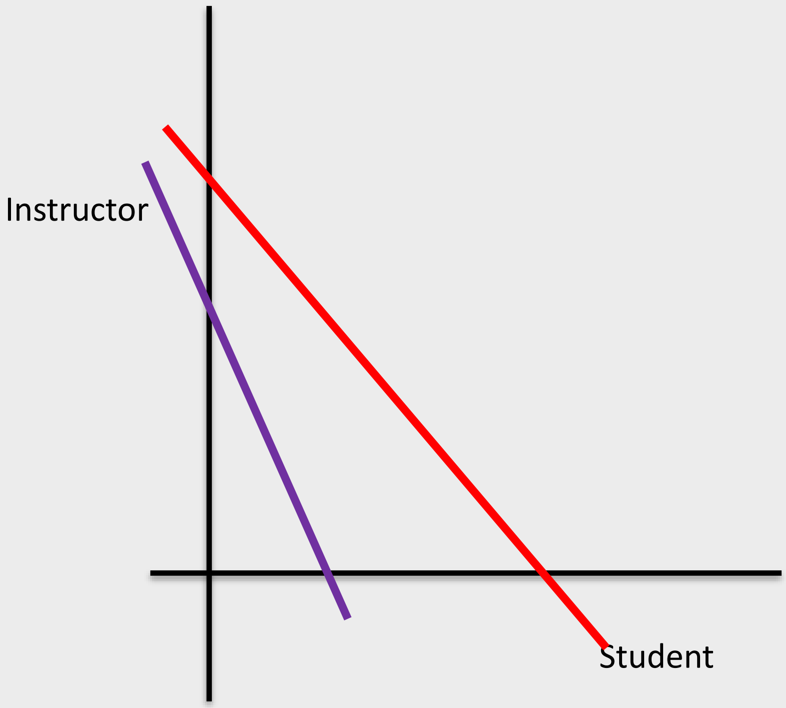
MCSE 666:
Pattern and Speech Recognition
Aspect & Prospect

Dr. Md. Aminul Haque Akhand
Dept. of CSE, SUB

PG Course: Instructor vs. Student Qualification and Contribution



Qualification



Contribution

Pattern and Speech Recognition (PSR)

Pattern

Speech

Recognition

PSR Individual Terms: Pattern

pattern

noun

1. a repeated decorative design.
   "a neat blue herringbone pattern"
   Similar: design, decoration, motif, marking, ornament, ornamentation
2. a model or design used as a guide in needlework and other crafts.
   "make a pattern for the zigzag edge"
   Similar: sample, specimen, swatch

verb

1. decorate with a repeated design.
   "he was sitting on a soft carpet patterned in rich colours"
   Similar: decorated, ornamented, figured, tessellated, mosaic, goffered
2. give a regular or intelligible form to.
   "the brain not only receives information, but interprets and patterns it"
   Similar: shape, influence, form, model, fashion, mould, style, affect

Translate to Bangla

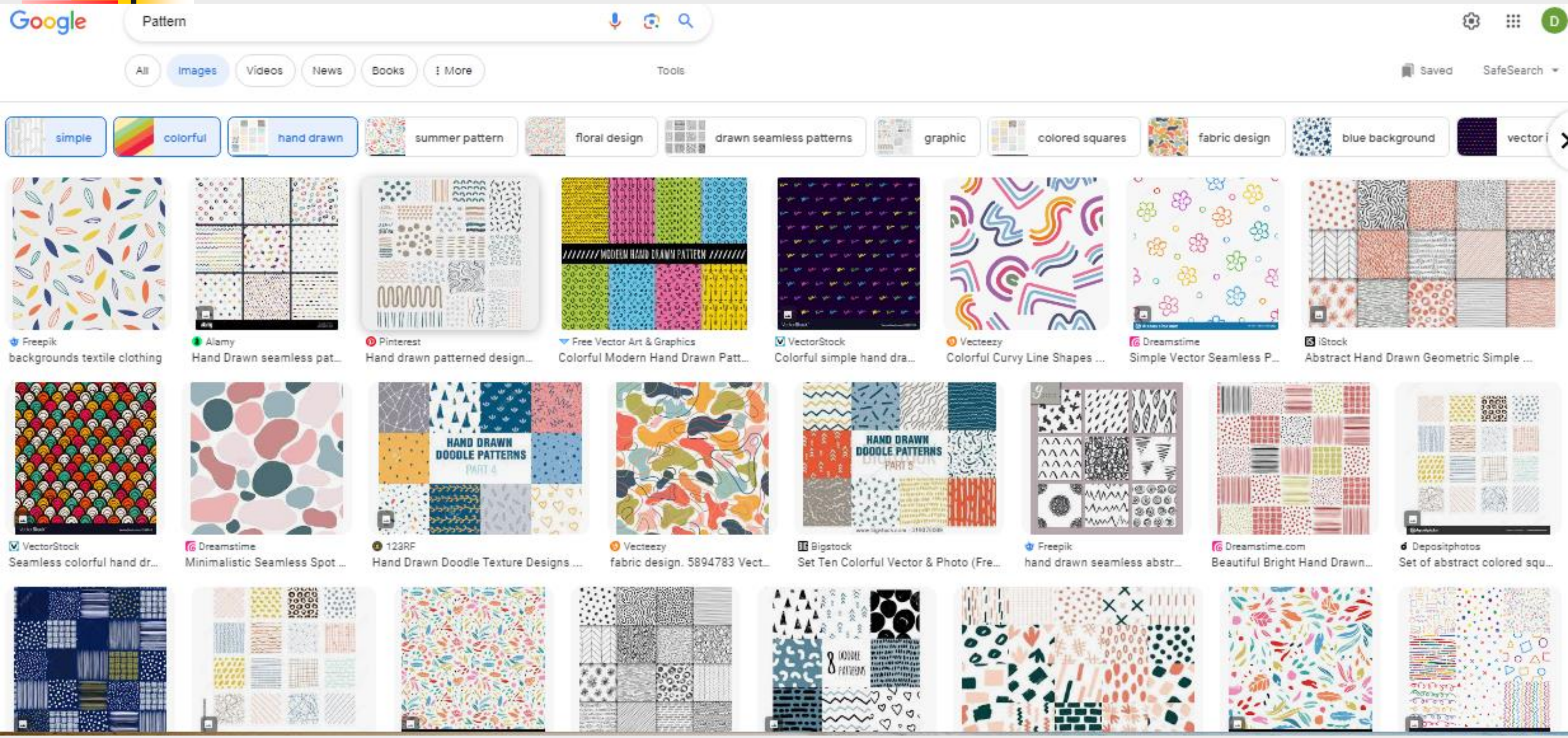
noun

1. প্যাটার্ন
2. আদর্শ

verb

1. আদর্শরূপে গঠন করা
2. আদর্শানুযায়ী গঠন করা

PSR Individual Terms: Pattern



A pattern is a regularity in the world, in human-made design, or in abstract ideas. The elements of a pattern repeat in a predictable manner. A geometric pattern is a kind of pattern formed of geometric shapes and typically repeated like a wallpaper design.

PSR Individual Terms: Speech

Translate to Bangla

noun

1. বক্তৃতা
2. কথাবার্তা

1. কথা
2. বাকশক্তি

speech

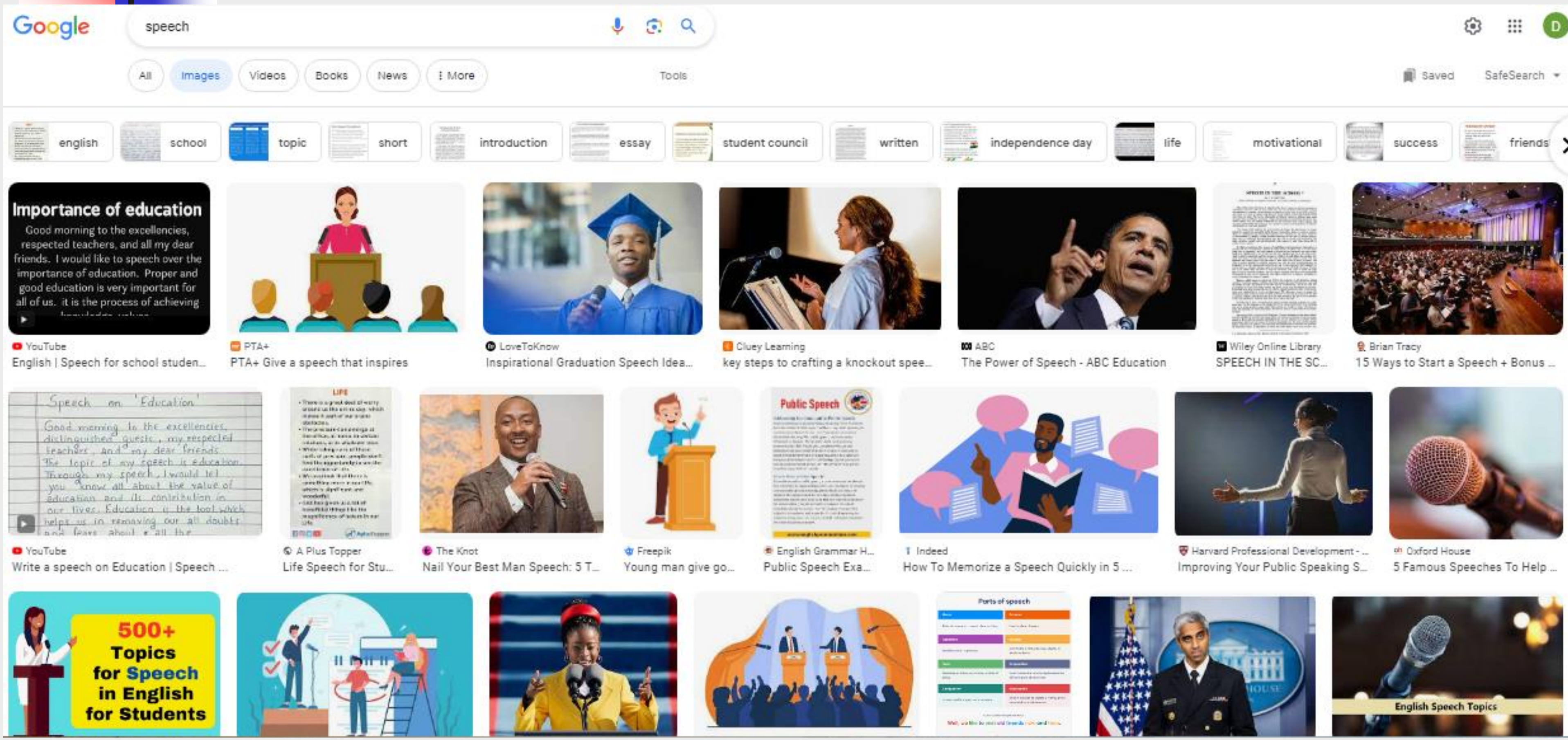
noun

noun: **speech**; plural noun: **speeches**

1. the expression of or the ability to express thoughts and feelings by articulate sounds.
   "he was born deaf and without the power of speech"
   Similar: speaking talking verbal communication verbal expression articulation
   - a person's style of speaking.
     "she wouldn't accept his correction of her speech"
     Similar: diction elocution manner of speaking articulation enunciation
2. a formal address or discourse delivered to an audience.
   "he gave a speech about the company"
   Similar: talk address lecture discourse oration disquisition
   - a sequence of lines written for one character in a play.
     "Antony's speech over Caesar's body"

7

PSR Individual Terms: Speech



Speech, language refer to the means of communication used by people. Speech is the expression of ideas and thoughts by means of articulate vocal sounds, or the faculty of thus expressing ideas and thoughts.

PSR Individual Terms: Recognition

recognition

noun

noun: **recognition**; plural noun: **recognitions**

1. identification of someone or something or person from previous encounters or knowledge.
   "she saw him pass by without a sign of recognition"

   Similar: identification recollection recall remembrance

2. acknowledgement of the existence, validity, or legality of something.
   "the unions must receive proper recognition"

   Similar: acknowledgement acceptance admission conceding concession

   - appreciation or acclaim for an achievement, service, or ability.
     "his work was slow to gain recognition"

     Similar: appreciation gratitude thanks congratulations a pat on the back

   - formal acknowledgement by a country that another political entity fulfils the conditions of statehood and is eligible to be dealt with as a member of the international community.
     noun: **diplomatic recognition**; plural noun: **diplomatic recognitions**
     "they are granting full recognition to the republic"

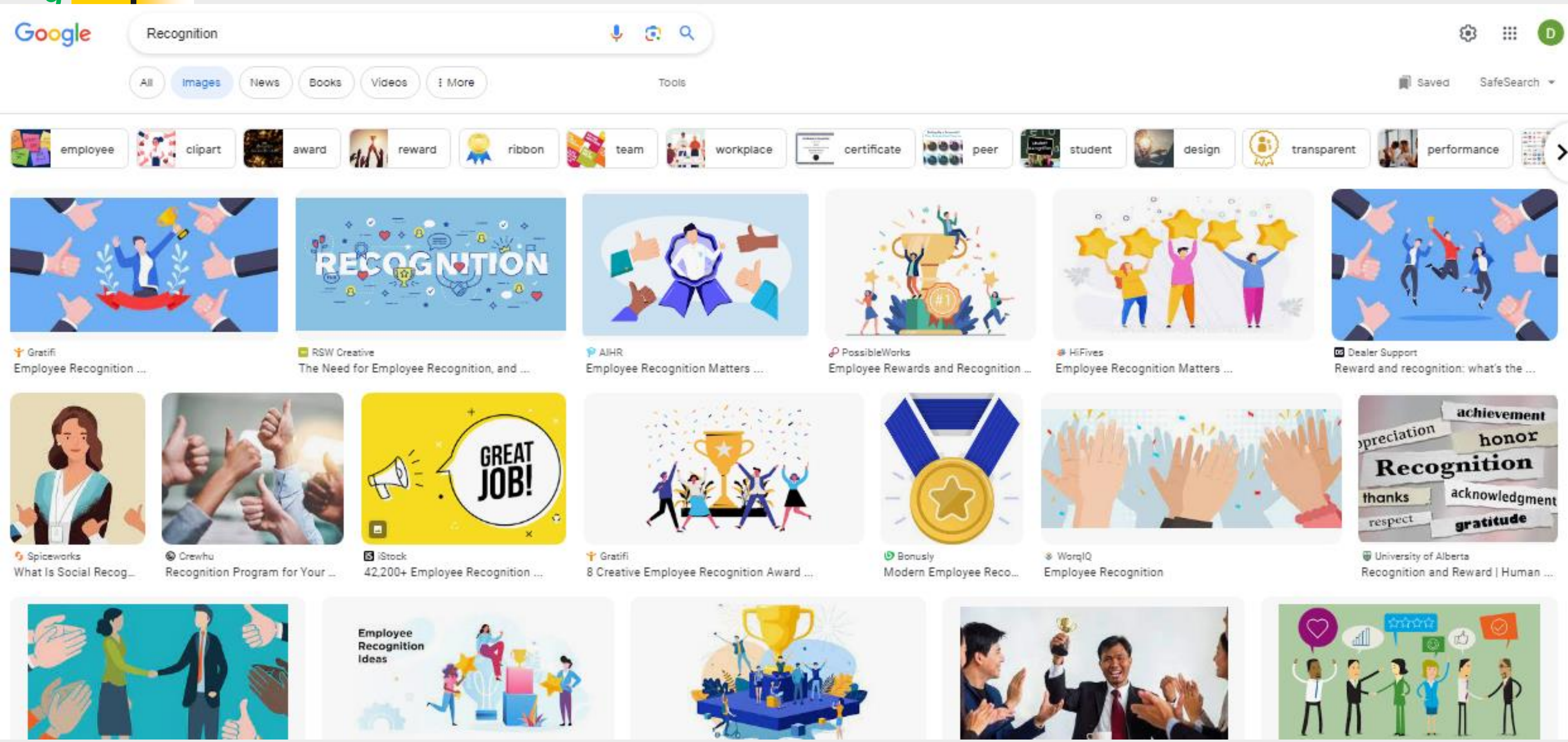
Translate to Bangla

noun

1. স্বীকার
2. ঠাহর

1. স্বীকৃতিদান হওয়া
2. স্বীকৃত হওয়া

PSR Individual Terms: Recognition



- Recognition in sociology is the public acknowledgment of a person's status or merits (achievements, virtues, service, etc.).
- Another example of recognition is when some person is accorded some special status, such as title or classification.

Pattern and Speech Recognition (PSR)

Pattern Speech Recognition

Speech Pattern Recognition

Pattern Recognition and
Speech Recognition

Intelligent Pattern Recognition and
Intelligent Speech Recognition

Examples of Pattern Recognition

Table 2 Examples of Pattern Recognition Applications

Problem Domain	Application	Input Pattern	Pattern Class
Bioinformatics	Sequence Analysis	DNA/Protein sequence	Known type of genes/patterns
Data mining	Searching for meaningful patterns	Points in multi dimension space	Compact and well separated clusters
Document classification	Internet search	Text document	Semantic categories
Document image analysis	Reading machine for the blind	Document image	Alphanumeric characters / words
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective / non defective nature of product
Multimedia database retrieval	Internet search	Video clip	Video genres e.g. action, dialogue etc
Biometric recognition	Personal identification	Face, iris & finger print	Authorized user for access control
Remote sensing	Forecasting crop yield	Multispectral image	Land use categories, growth pattern of crops
Speech recognition	Telephone directory enquiry with operator	Speech waveform	Spoken words

Mayank Parasher et al., Anatomy on Pattern Recognition, June 2011, Indian Journal of Computer Science and Engineering 2(3)

Examples of Pattern Recognition

What Are the Real-World Applications of Pattern Recognition?

https://serokell.io/blog/applications-of-pattern-recognition

10 Real Life Examples Of Pattern Recognition

https://numberdyslexia.com/pattern-recognition-real-life-examples/

Pattern recognition and use in real life problem solving

https://suresolv.com/problem-solving-techniques/pattern-recognition-and-use-real-life-problem-solving

What is pattern recognition and why it matters? Definitive guide

https://theappsolutions.com/blog/development/pattern-recognition-guide/

MCSE 666: Pattern and Speech Recognition

SUB MCSE 666 Syllabus or Course Outline

Introduction to formal languages and patterns; string languages for pattern description; higher dimensional pattern grammars; syntax analysis as a recognition procedure; stochastic languages; error correcting parsing for string languages; error-correcting tree automata; cluster analysis for syntactic patterns; grammatical inference for syntactic pattern recognition; syntactic approach to texture analysis;

Speech signal: production, perception and characterization; signal processing and analysis; pattern comparison techniques: distortion measures, spectral-distortion measures, time alignment and normalization; recognition system design and implementation: source-coding, template training, and performance analysis; connected word models; continuous speech recognition: sub word units, statistical modeling, and context-dependent units; task oriented models.

Textbooks

Open Discussion on Course Conduction

MCSE 666:Pattern and Speech Recognition

Introduction to Pattern Recognition

Dr. Md. Aminul Haque Akhand
Dept. of CSE, SUB

What Is Pattern Recognition?

Pattern recognition (PR) is a process by which some input is measured, analyzed, and then classified as belonging to one of a set of classes

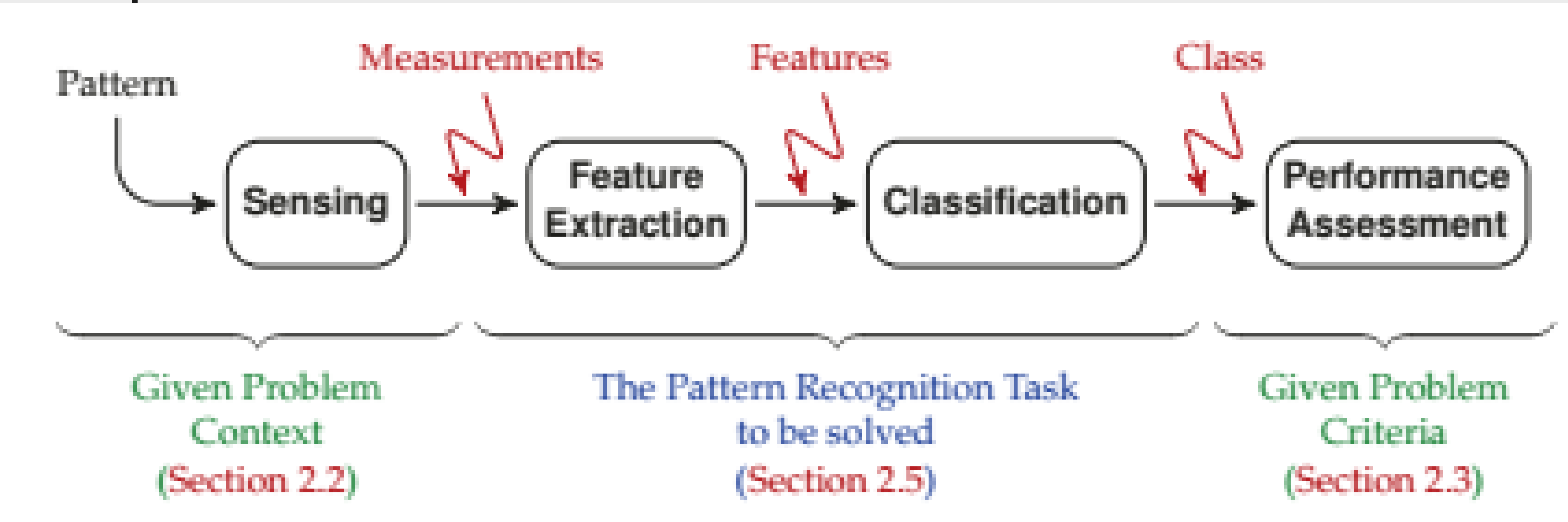
Process of PR and classification is a continual, never ending aspect of every-day human existence:

Pattern recognition task	Possible classes
What is in front of you as you walk?	Door vs. Window Sidewalk vs. Road
What music are you listening to?	Familiar or Unfamiliar Genre (Rock, Classical, . . .) Name of Composer or Group
Is the traffic intersection safe to cross?	Green vs. Red light Pedestrian Walk vs. Stop Car Present vs. Not Present
Reading a page in a textbook	Letters of the Alphabet Text vs. Graphics Languages
You smell something in your	Cookies finished baking? (Yes/No) Is something burning? (Yes/No)

What Is Pattern Recognition? Cont.

- PR as a human experience refers to a perceptual process: some form of sensory input is sensed, analyzed, and recognized (classified), either subconsciously (by instinct) or consciously (based on previous experience).
- Patterns may be presented in any sensory modality: vision, hearing, touch, taste, or smell.
- PR as a technical discipline: a process in which an input object is measured, analyzed, and classified by a machine as being more or less similar to some class in a set of classes.

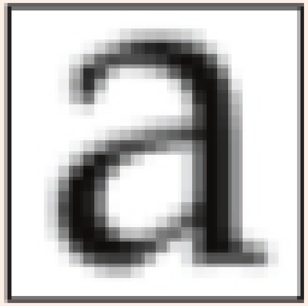
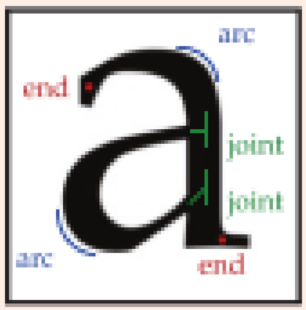
Pattern Recognition Framework



The goal of PR is to provide a machine with a kind of perceptual capability to automatically extract useful information from measured data.

PR Illustration with Example

Pattern Recognition of Text

Pattern	Attributes to Measure	Measurements	Strengths and Weaknesses
"a" →		Vector of pixel values	Fast, easy, explicit Sensitive to changes in font style, size, and rotation
"a" →		Vector of shape properties	Robust to changes in size and rotation Complicated features to extract
"a" →	Complex Nonlinear Algorithm	Vector of values, but with no intuition	Possibly very flexible May be very hard to learn Difficult to analyze

PR Illustration with Example

Pattern Recognition of the Mind

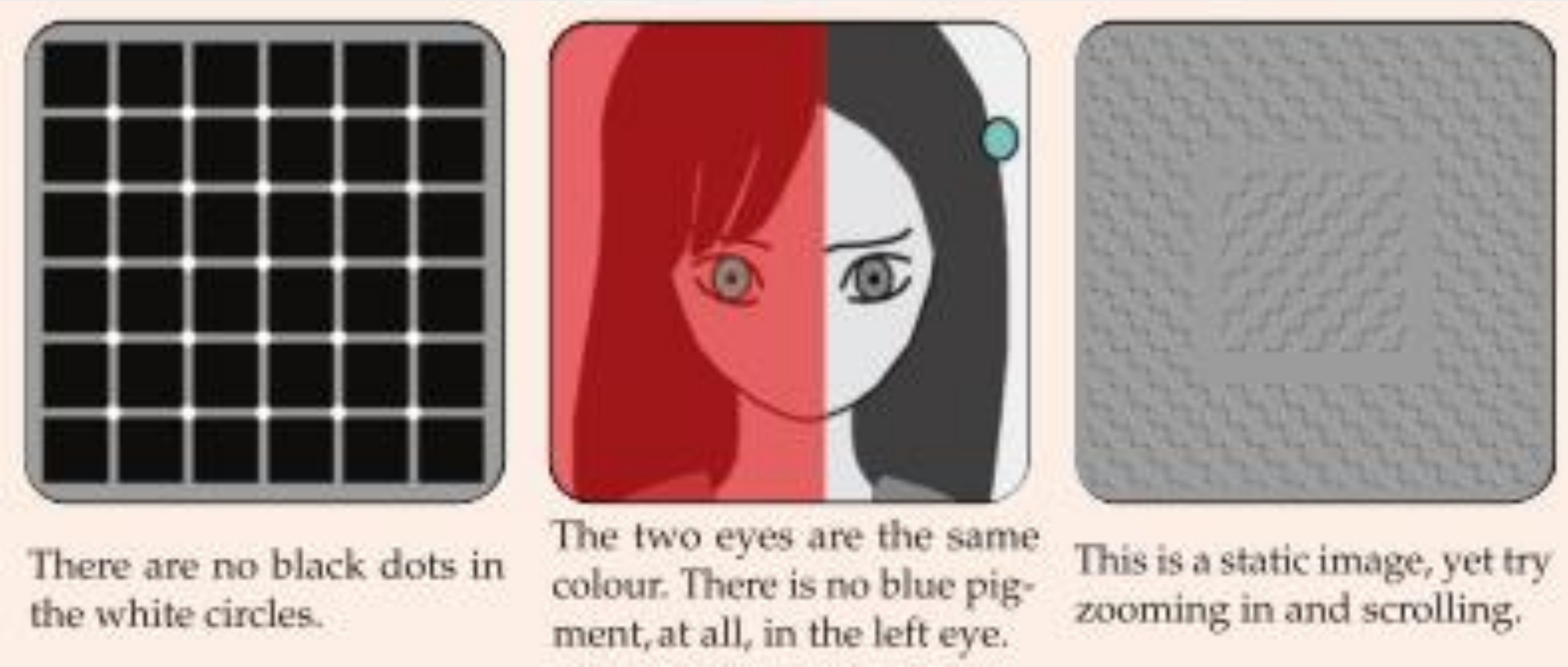
The typesetting may be weird, but for the human brain this is very easy to read.

It deons't mettar in waht oerdr the lrttees in a wrod are, as lnog as the frist and lsat ltteres rmeain in the rgiht pacle.

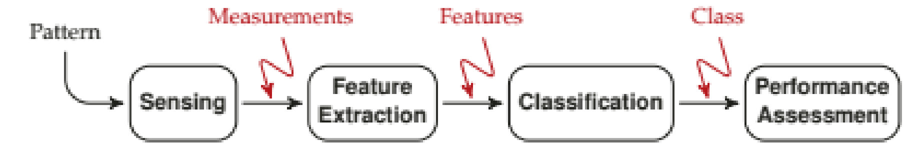
- The retina in the eye is densely packed with light-sensitive cells, so that it may be tempting to think
- Our brain effectively sees and perceives the world as a great many pixels.
- There are many simple mind tricks or optical illusions

PR Illustration with Example

Pattern Recognition of the Mind



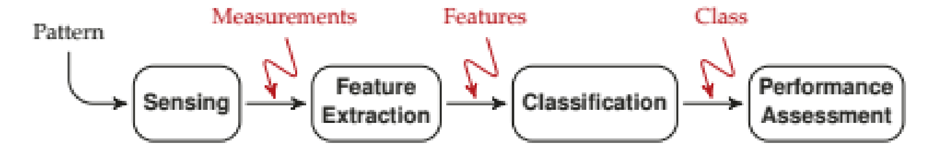
- The eye is not just sending pixellated images to the brain.
- There is a great deal of feature extraction taking place, much already in the retina
- It is very helpful in running through a forest, but perhaps not so useful in staring at deliberately manipulated images on a page.



Features from Patterns

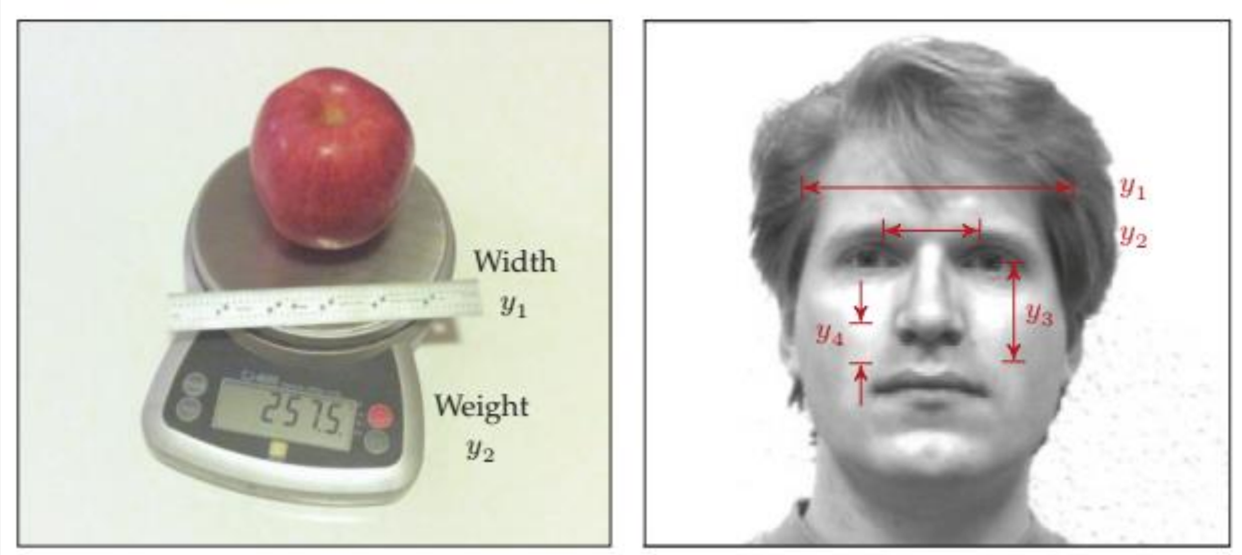
- The word “pattern” may bring to mind texture, fabric, or shape.
- In the context of *pattern recognition*, the notion of pattern is far more broad, and can apply to any *thing* that can be distinguished from another *thing*.
- Identity view point: infer the unknown *identity* of an object
- Type of wildflower, type of songbird, or the name of the person facing a camera — each of these has a certain identity to determine from measurements.

A pattern is assumed to have certain *properties* or *attributes* which distinguishes it from other patterns.



Features from Patterns (Cont.)

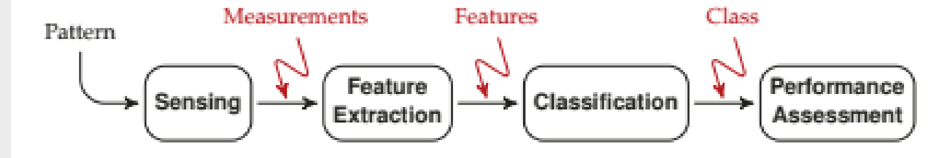
One or more measurements are taken of a pattern



Size / weight of a piece

Various dimensions from face

Does all measurements from patterns important to distinguish?

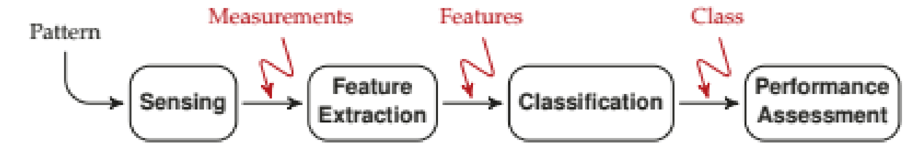


Features from Patterns (Cont.)

- The selection of appropriate measurements is an essential
- Measurements may cost money and/or time, and
- Poor measurements lead to poor performance of the resulting classifier.

The process of transforming measurements into features facilitate classification, normally in one or both of the following ways:
1. By reducing the dimensionality of the problem: $n < m$
2. By creating features in which patterns are more clearly distinguished.

$$\text{Measurements} \quad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \quad \xrightarrow{\underline{x} = f(\underline{y})} \quad \text{Features} \quad \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

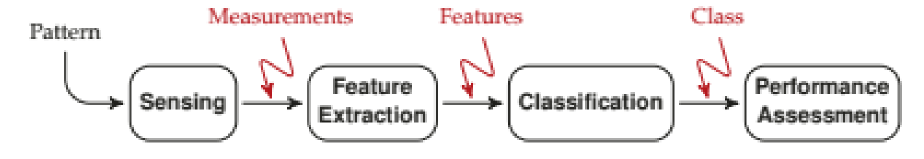


Features from Patterns (Cont.)

$$\text{Measurements } \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \xrightarrow{\underline{x} = f(\underline{y})} \text{Features } \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- The feature extraction function *f()* can focus the information from x, or it can remove irrelevant information from x, but *f()* never adds information.
- Data Processing Theorem: *x can never have more information than was present in y.*

- An effective feature extraction function *f*() can make the PR problem *easier,*
- However, in principle, the best possible classifier based on the measurements *y* should perform at least as well as the best possible classifier based on the features *x*.



Features from Patterns (Cont.)

Features may be intuitive or they may be quite abstract. Consider, the measurements of an electric motor:

$$\text{Measurements} \quad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \text{Motor voltage (Volts)} \\ \text{Motor current (Amperes)} \\ \text{Motor speed (RPM)} \\ \text{Motor winding temperature (°C)} \\ \text{Surrounding air temperature (°C)} \end{bmatrix}$$

$$x = f(\underline{y}) = y_1 \cdot y_2$$

Power as Feature: Understandable

$$x = f(\underline{y}) = \sqrt{y_1 - y_2} - \frac{y_3}{y_4}$$

- Uninterpretable feature do not make any physical sense,
- But perhaps be effective as a feature for classification.

Classes and Classification

- The whole purpose of PR or classification consists of assigning an object to a class.
- A class is a particular pattern, or possibly a group of patterns which are similar or equivalent in some sense.

In a given problem, the set of classes *C* is defined as

$$C = \{C_1; C_2; -; -; -; -; C_K\};$$

Have to choose one class from *K* different classes.

Members of a class share some common properties or attributes

In PR problems the class set *C* is predefined and has been specified as part of the problem to be solved.

Classes and Classification

So, for example, in a face recognition problem each person is their own class, so the set of classes would be defined as

$$\mathcal{C} = \{\text{“Paul Fieguth”}, \text{“Bob”}, \text{“Jane”}, \text{“Ali”}, \ldots\}$$

so that I would be a member of the “Paul Fieguth” class:

$\in$ “Paul Fieguth”

At a university we might define a different set of classes

$$\mathcal{C} = \{\text{“Professors”}, \text{“Staff”}, \text{“Undergraduate Students”}, \text{“Graduate Students”}\}$$

such that

$$\text{“Paul Fieguth”} \in \text{“Professor”}$$

One *could* use pattern recognition to estimate age, in which case one might have classes like

$$\mathcal{C} = \{\text{“0–10 years”}, \text{“10–20 years”}, \text{“20–30 years”}, \ldots\}$$

such that now I appear in a class as

$$\text{“Paul Fieguth”} \in \text{“50–60 years”}$$

Classes and Classification

Way of Describing Classes

Via Prototype Idealized representation or notion of the “essence” of the class. Pros: each class is unambiguously defined, Cons: no scope for variability.

Via Parameterized Shape: A generalization of the prototype; the class has a known shape (e.g., rectangular or elliptical), the shape is described in some number of parameters (e.g., ellipse centre, rotation, and axis lengths). Pros. more flexible than that of a single prototype, Cons: still requires the type of shape to be assumed or known.

Via Statistical Distribution: Some description of the likelihood or probability of a class member having a particular set of measurements or features. Pros.: very comprehensive, Cons: There will be circumstances when the statistics are not known and may be difficult to infer.

Via Samples: A set of given samples (many apples, or tigers, or bicycles) directly characterizes the class. Pros: Highly convenient, since nothing further needs to be done to describe the class Cons: Storage and computational challenges, since all of the data need to be saved.

Classes and Classification

Variability in a Class

Patterns do not need to be identical to belong to the same class: not all pictures of a person the same, or of tigers, or of apples.

There are at least two sources of variability present in the measurements associated with a single class:

1. The inherent variability within a class: Every class will consist of members which differ in some way. The degree and nature of the inherent variability will depend greatly on the class definition.

“Fruit” class contains all manners of variability in colour, size, and shape; “Apple” class is much more specific, but apples do come in different colours and patterning; the “Granny Smith Apple” class is even more specific, but still will have apples of different sizes or with more or fewer blemishes.

2. Noise or random variations in measurement: Every measurement involves some sort of physical process which will be subject to error, such as thermal noise in electronics, or quantization noise in converting an analogue signal to a digital representation.

Open Discussion on Course Conduction

MCSE 666:Pattern and Speech Recognition

Introduction to Pattern Recognition

Dr. Md. Aminul Haque Akhand
Dept. of CSE, SUB

What Is Pattern Recognition?

Pattern recognition (PR) is a process by which some input is measured, analyzed, and then classified as belonging to one of a set of classes

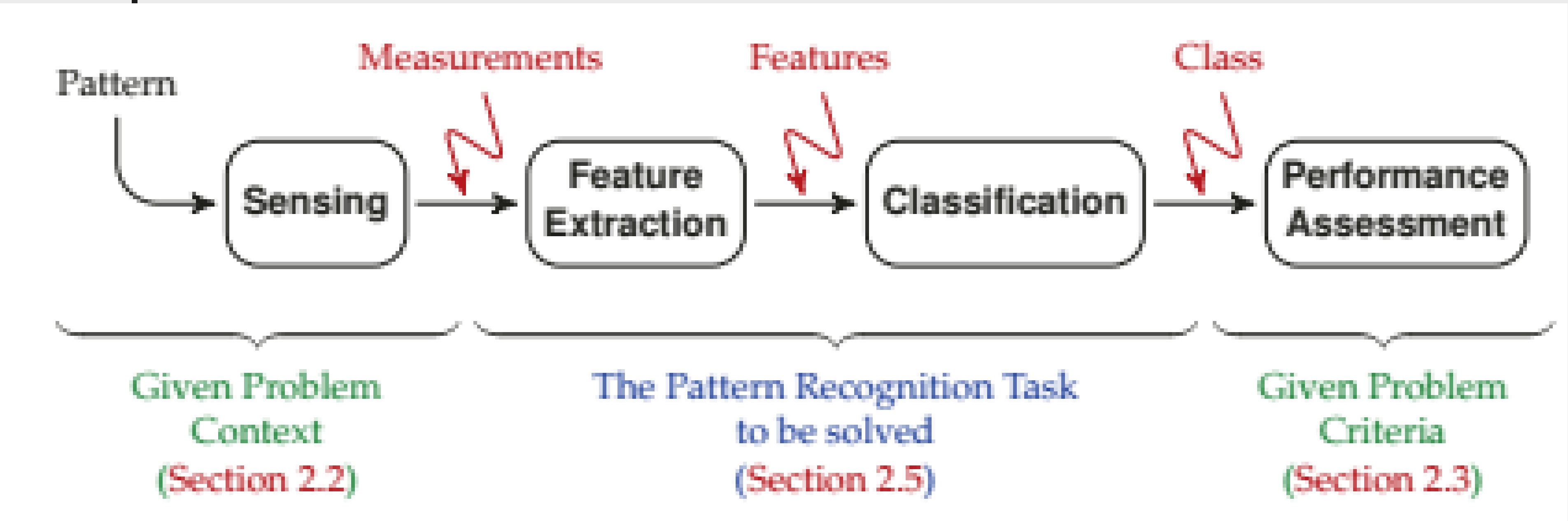
Process of PR and classification is a continual, never ending aspect of every-day human existence:

Pattern recognition task	Possible classes
What is in front of you as you walk?	Door vs. Window Sidewalk vs. Road
What music are you listening to?	Familiar or Unfamiliar Genre (Rock, Classical, ...) Name of Composer or Group
Is the traffic intersection safe to cross?	Green vs. Red light Pedestrian Walk vs. Stop Car Present vs. Not Present
Reading a page in a textbook	Letters of the Alphabet Text vs. Graphics Languages
You smell something in your	Cookies finished baking? (Yes/No) Is something burning? (Yes/No)

What Is Pattern Recognition? Cont.

- PR as a human experience refers to a perceptual process: some form of sensory input is sensed, analyzed, and recognized (classified), either subconsciously (by instinct) or consciously (based on previous experience).
- Patterns may be presented in any sensory modality: vision, hearing, touch, taste, or smell.
- PR as a technical discipline: a process in which an input object is measured, analyzed, and classified by a machine as being more or less similar to some class in a set of classes.

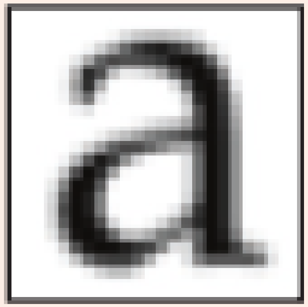
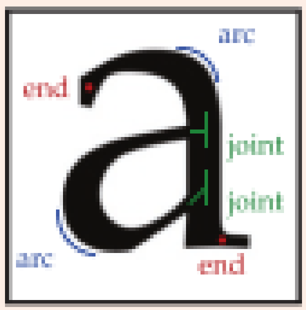
Pattern Recognition Framework



The goal of PR is to provide a machine with a kind of perceptual capability to automatically extract useful information from measured data.

PR Illustration with Example

Pattern Recognition of Text

Pattern	Attributes to Measure	Measurements	Strengths and Weaknesses
"a" →		Vector of pixel values	Fast, easy, explicit Sensitive to changes in font style, size, and rotation
"a" →		Vector of shape properties	Robust to changes in size and rotation Complicated features to extract
"a" →	Complex Nonlinear Algorithm	Vector of values, but with no intuition	Possibly very flexible May be very hard to learn Difficult to analyze

PR Illustration with Example

Pattern Recognition of the Mind

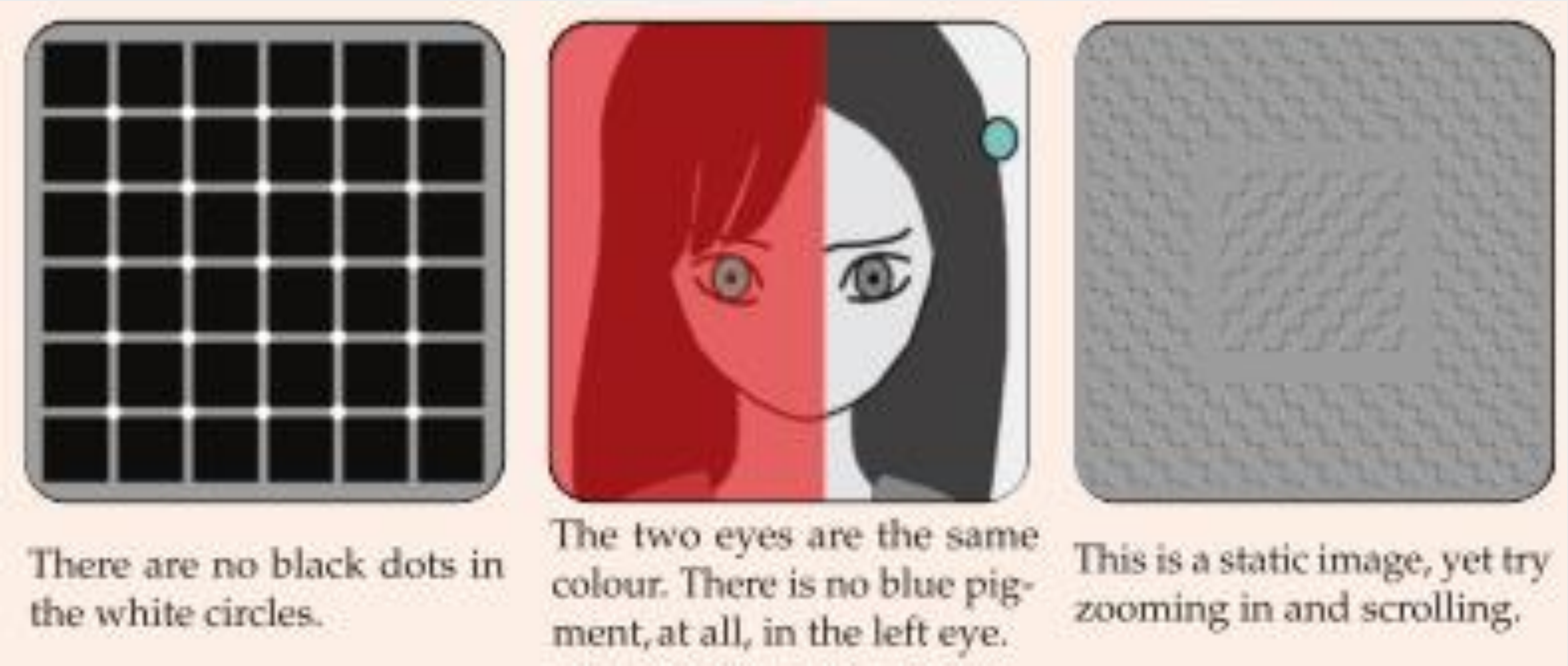
The typesetting may be weird, but for the human brain this is very easy to read.

It deons't mettar in waht oerdr the lrttees in a wrod are, as lnog as the frist and lsat ltteres rmeain in the rgiht pacle.

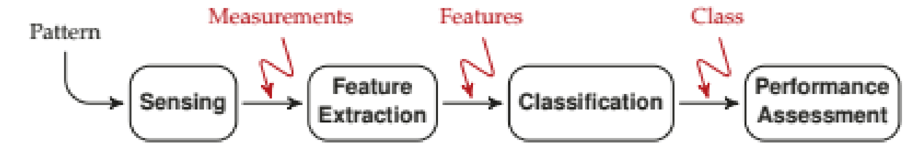
- The retina in the eye is densely packed with light-sensitive cells, so that it may be tempting to think
- Our brain effectively sees and perceives the world as a great many pixels.
- There are many simple mind tricks or optical illusions

PR Illustration with Example

Pattern Recognition of the Mind



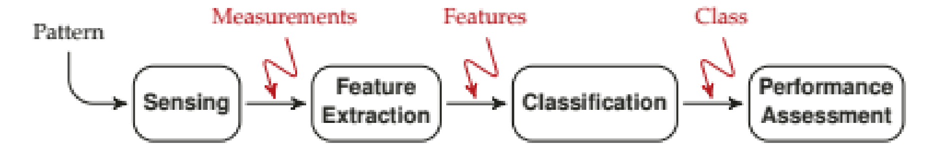
- The eye is not just sending pixellated images to the brain.
- There is a great deal of feature extraction taking place, much already in the retina
- It is very helpful in running through a forest, but perhaps not so useful in staring at deliberately manipulated images on a page.



Features from Patterns

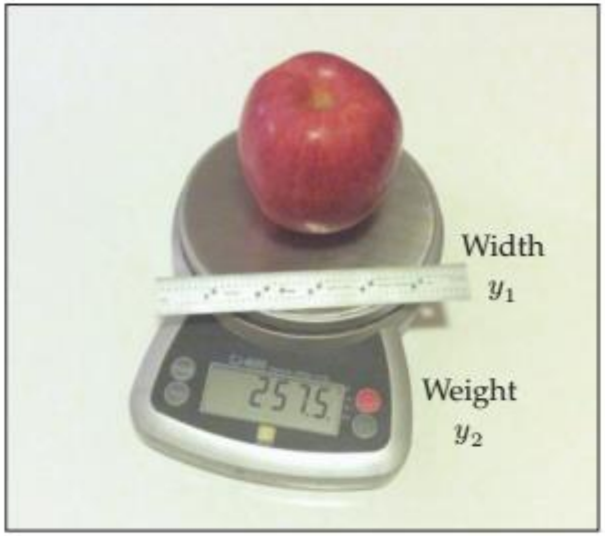
- The word “pattern” may bring to mind texture, fabric, or shape.
- In the context of *pattern recognition*, the notion of pattern is far more broad, and can apply to any *thing* that can be distinguished from another *thing*.
- Identity view point: infer the unknown *identity* of an object
- Type of wildflower, type of songbird, or the name of the person facing a camera — each of these has a certain identity to determine from measurements.

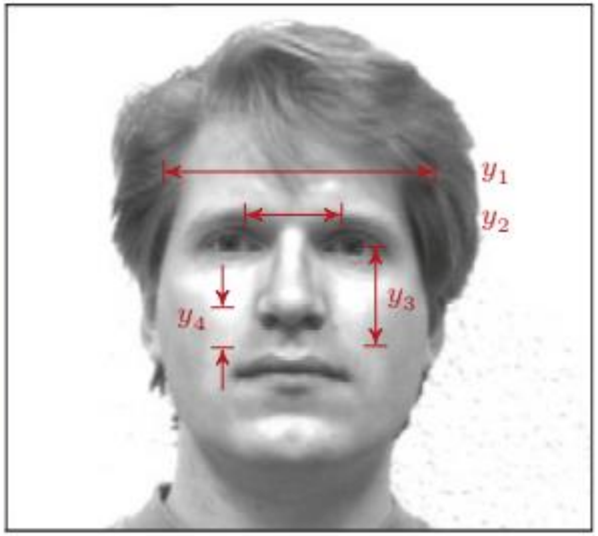
A pattern is assumed to have certain *properties* or *attributes* which distinguishes it from other patterns.



Features from Patterns (Cont.)

One or more measurements are taken of a pattern

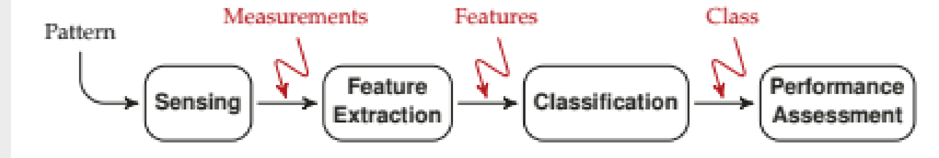




Size / weight of a piece

Various dimensions from face

Does all measurements from patterns important to distinguish?

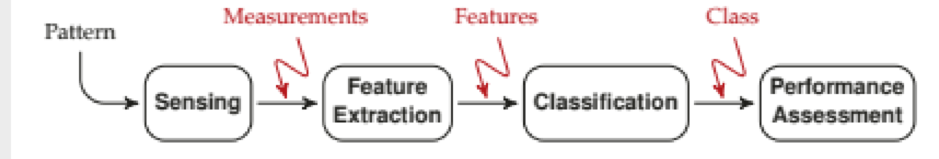


Features from Patterns (Cont.)

- The selection of appropriate measurements is an essential
- Measurements may cost money and/or time, and
- Poor measurements lead to poor performance of the resulting classifier.

The process of transforming measurements into features facilitate classification, normally in one or both of the following ways:
1. By reducing the dimensionality of the problem: $n < m$
2. By creating features in which patterns are more clearly distinguished.

$$\text{Measurements} \quad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \quad \xrightarrow{\underline{x} = f(\underline{y})} \quad \text{Features} \quad \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

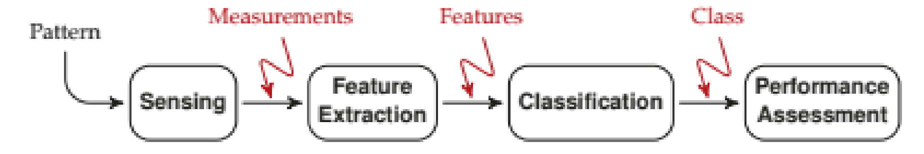


Features from Patterns (Cont.)

$$\text{Measurements}\ \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \xrightarrow{\underline{x} = f(\underline{y})} \text{Features}\ \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- The feature extraction function *f()* can focus the information from x, or it can remove irrelevant information from x, but *f()* never adds information.
- Data Processing Theorem: *x can never have more information than was present in y.*

- An effective feature extraction function *f*() can make the PR problem *easier,*
- However, in principle, the best possible classifier based on the measurements *y* should perform at least as well as the best possible classifier based on the features *x*.



Features from Patterns (Cont.)

Features may be intuitive or they may be quite abstract. Consider, the measurements of an electric motor:

$$\text{Measurements} \quad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \text{Motor voltage (Volts)} \\ \text{Motor current (Amperes)} \\ \text{Motor speed (RPM)} \\ \text{Motor winding temperature (°C)} \\ \text{Surrounding air temperature (°C)} \end{bmatrix}$$

$$x = f(\underline{y}) = y_1 \cdot y_2$$

Power as Feature: Understandable

$$x = f(\underline{y}) = \sqrt{y_1 - y_2} - \frac{y_3}{y_4}$$

- Uninterpretable feature do not make any physical sense,
- But perhaps be effective as a feature for classification.

Classes and Classification

- The whole purpose of PR or classification consists of assigning an object to a class.
- A class is a particular pattern, or possibly a group of patterns which are similar or equivalent in some sense.

In a given problem, the set of classes C is defined as

$$C = \{C_1; C_2; -; -; -; -; C_K\};$$

Have to choose one class from K different classes.

Members of a class share some common properties or attributes

In PR problems the class set C is predefined and has been specified as part of the problem to be solved.

Classes and Classification

So, for example, in a face recognition problem each person is their own class, so the set of classes would be defined as

$$\mathcal{C} = \{\text{"Paul Fieguth"}, \text{"Bob"}, \text{"Jane"}, \text{"Ali"}, \ldots\}$$

so that I would be a member of the "Paul Fieguth" class:

$\in$ "Paul Fieguth"

At a university we might define a different set of classes

$$\mathcal{C} = \{\text{"Professors"}, \text{"Staff"}, \text{"Undergraduate Students"}, \text{"Graduate Students"}\}$$

such that

$$\text{"Paul Fieguth"} \in \text{"Professor"}$$

One *could* use pattern recognition to estimate age, in which case one might have classes like

$$\mathcal{C} = \{\text{"0–10 years"}, \text{"10–20 years"}, \text{"20–30 years"}, \ldots\}$$

such that now I appear in a class as

$$\text{"Paul Fieguth"} \in \text{"50–60 years"}$$

Classes and Classification

Way of Describing Classes

Via Prototype Idealized representation or notion of the “essence” of the class. Pros: each class is unambiguously defined, Cons: no scope for variability.

Via Parameterized Shape: A generalization of the prototype; the class has a known shape (e.g., rectangular or elliptical), the shape is described in some number of parameters (e.g., ellipse centre, rotation, and axis lengths). Pros. more flexible than that of a single prototype, Cons: still requires the type of shape to be assumed or known.

Via Statistical Distribution: Some description of the likelihood or probability of a class member having a particular set of measurements or features. Pros.: very comprehensive, Cons: There will be circumstances when the statistics are not known and may be difficult to infer.

Via Samples: A set of given samples (many apples, or tigers, or bicycles) directly characterizes the class. Pros: Highly convenient, since nothing further needs to be done to describe the class Cons: Storage and computational challenges, since all of the data need to be saved.

Classes and Classification

Variability in a Class

Patterns do not need to be identical to belong to the same class: not all pictures of a person the same, or of tigers, or of apples.

There are at least two sources of variability present in the measurements associated with a single class:

1. The inherent variability within a class: Every class will consist of members which differ in some way. The degree and nature of the inherent variability will depend greatly on the class definition.

“Fruit” class contains all manners of variability in colour, size, and shape; “Apple” class is much more specific, but apples do come in different colours and patterning; the “Granny Smith Apple” class is even more specific, but still will have apples of different sizes or with more or fewer blemishes.

2. Noise or random variations in measurement: Every measurement involves some sort of physical process which will be subject to error, such as thermal noise in electronics, or quantization noise in converting an analogue signal to a digital representation.

Classifier and Classification

In PR tasks the class set C is predefined and has been specified as part of the problem to be solved.

A *classifier* is some function *g*(), possibly analytical (i.e., an equation) or a computer algorithm, which assigns a class label to a given feature:

$$g(x) \in C = \{C_1; C_2; : : : ; C_K\}$$

There is a strong relationship between feature extraction and classification, such that

- Good features allow for simpler classifiers, whereas
- Complex classifiers can compensate for weaker features, which are unable to fully separate the pattern classes.

Classifier and Classification

Two Fundamental Steps Associated With Classification:
I. Classifier Learning and II. Classifier Testing or Validation

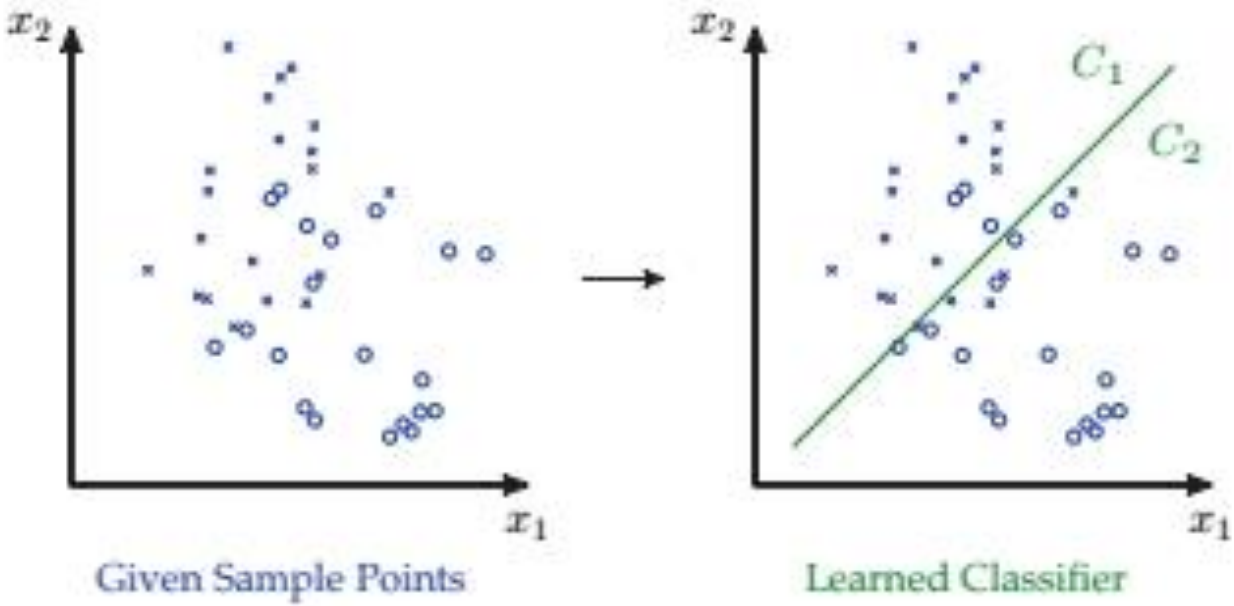


Fig. 2.3. PATTERN RECOGNITION I — CLASSIFIER LEARNING: A classifier, here a straight line (right) dividing a feature space into classifications C_1 and C_2, can be *learned* from a given set of sample points (left).

Classifier and Classification

II. Classifier Testing or Validation

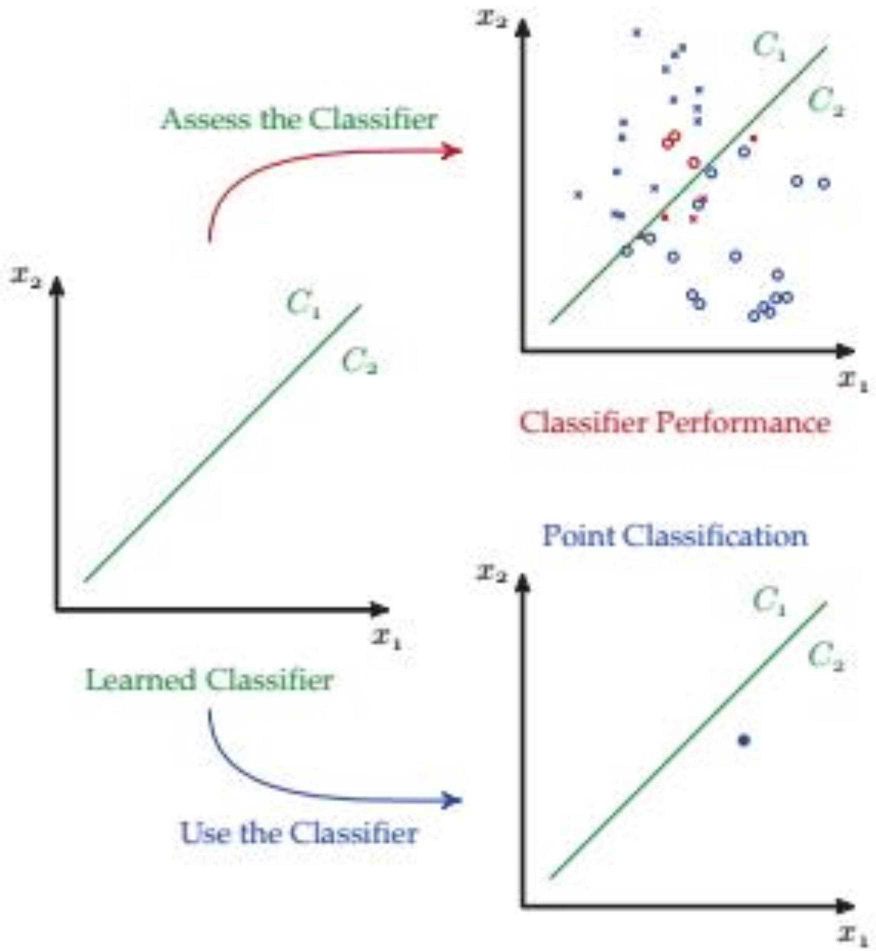


Fig. 2.4. PATTERN RECOGNITION II — CLASSIFIER TESTING: What can we do with the learned classifier, left, from Figure 2.3? We could assess its performance (top), for example by counting how many sample points are classified correctly (blue) and incorrectly (red). Or we could apply the classifier (bottom) to a new, unknown point and then classify it.

Classifier and Classification

Illustrative Examples of Classification

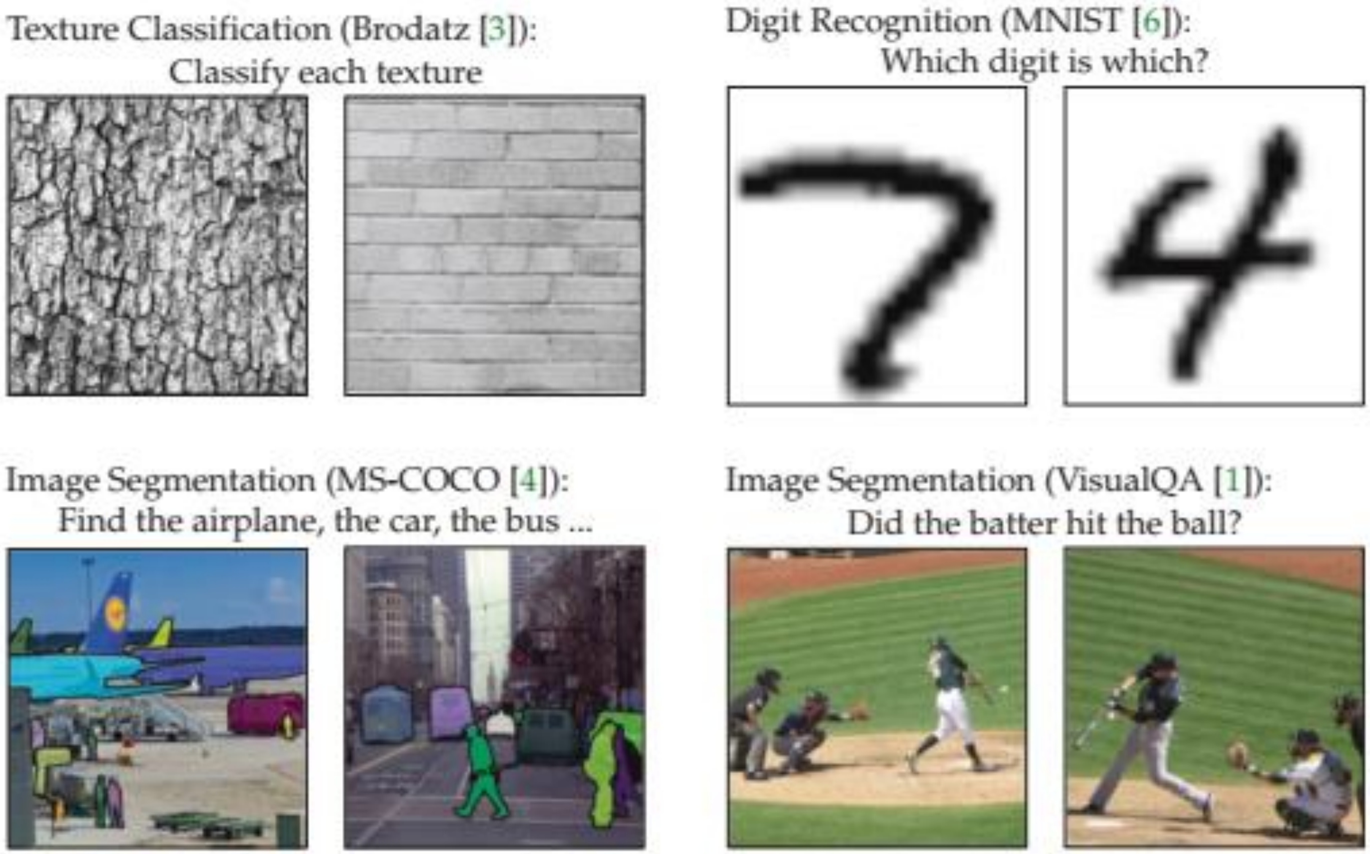


Fig. 2.5. PATTERN RECOGNITION ON IMAGES: Since the human visual system is so dominant in human perception, a great deal of pattern recognition focuses on image-related problems. Here four examples are shown, from comparatively straightforward (top), the classification or recognition of whole images, to rather advanced (bottom), such as recognizing the objects within an image or being able to answer complex high-level questions.

Classifier and Classification

Pattern Recognition Problems

Four Fundament Types PR Problems:

1. Model-Based
2. Supervised
3. Semi-Supervised
4. Unsupervised

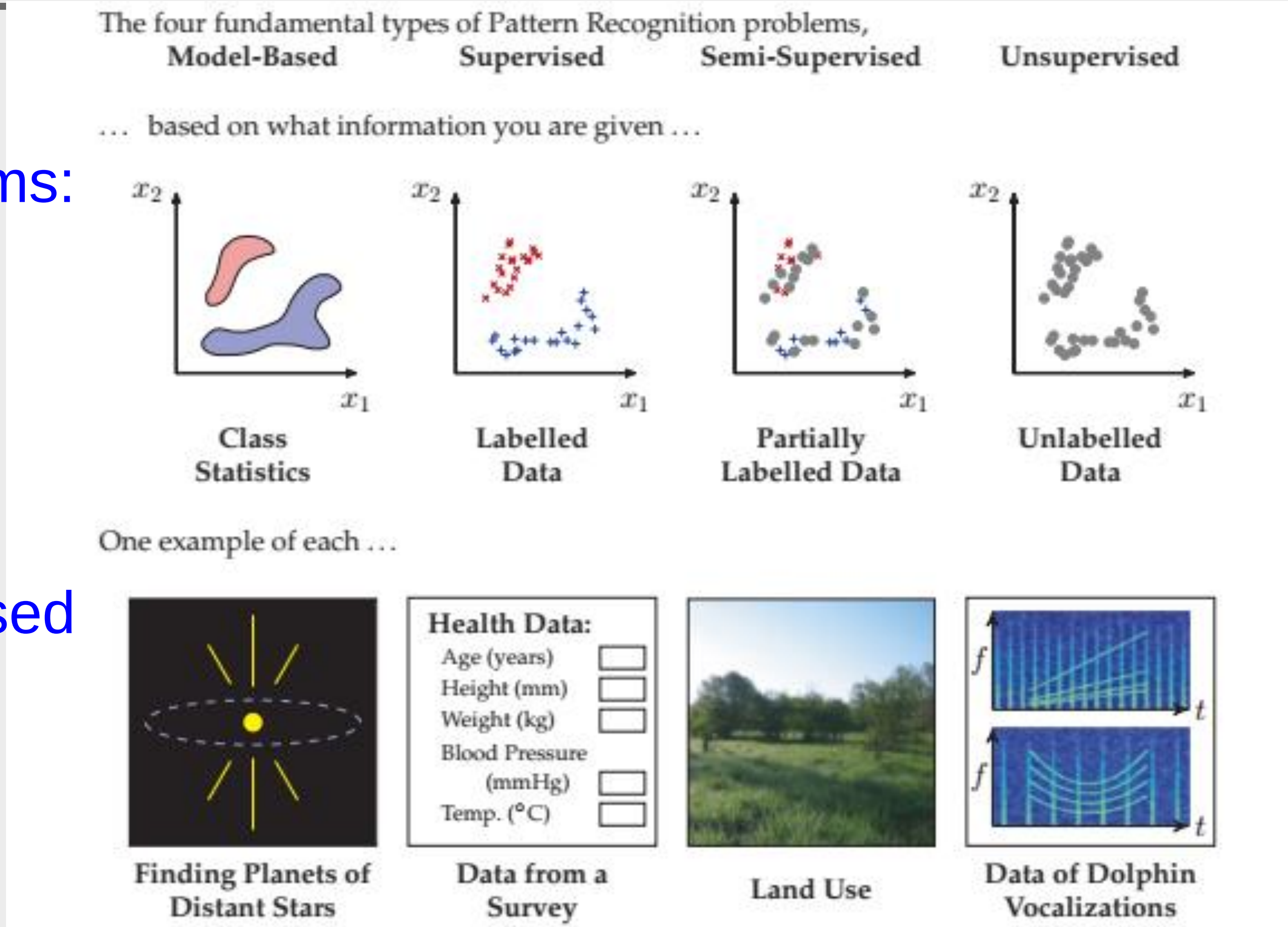


Fig. 2.6. THE FUNDAMENTAL PROBLEMS: There are four fundamental pattern recognition problems, ordered from the most detailed problem description (left) to the most ambiguous (right).

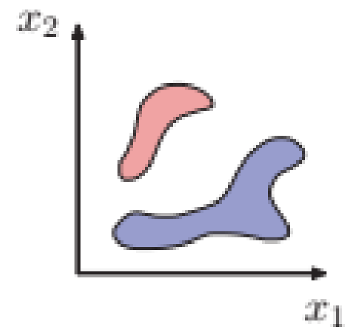
Classifier and Classification

Pattern Recognition Problem: Model-Based

Scenario 1 (Model-Based): Model is known or given

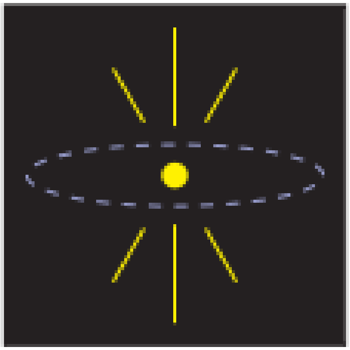
- ➢ Hope to have most information regarding a PR problem the behaviour of the measurements for each of the pattern classes.
- ➢ Normally characterized in a statistical fashion, such as $p(y|C)$ = The distribution of measurement vector y given class C

- ❖ Such detailed information will be available only in those contexts where the physical process is known by which a given pattern class gives rise to measurements.
- ❖ Very convenient to have detailed information, since statistical decision theory allow to explicitly define the optimal classifier, in the sense of minimizing the probability of classification error



Class Statistics

One example of each . . .



Finding Planets of Distant Stars

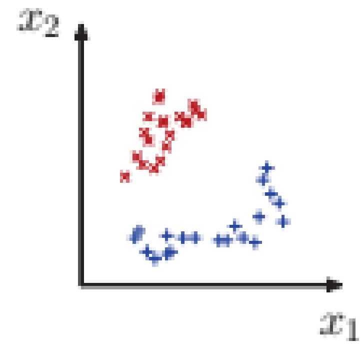
Classifier and Classification

Pattern Recognition Problem: Supervised

Scenario 2 (Supervised): Model is not known, labelled data are available

- Do not have an exact description of the problem, as in Scenario 1
- Given labelled data, meaning data pairs of the form

$\{y_i; C_K\}$ -> The *i*th measurement vector y_i is in class C_K



Labelled Data

- Labelled data do not just *magically* appear; labelled or tagged by a human observer, so refer scenario as *supervised*
- Exceptionally expensive or labour-intensive with larger datasets.
- May derive a classifier directly from the given labelled or learn an empirical probability model as of Scenario 1

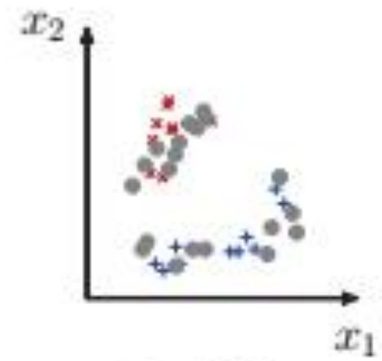
Data from a Survey

Classifier and Classification

Pattern Recognition Problem: Semi-Supervised

Scenario 3 (Semi-Supervised): Model is not known, some are labelled data, some are not

- Labelled data can be expensive, requiring manual labelling. A huge data be unlabeled and Scenario 2 ignores all unlabeled data.
- Problem refers to as semi-supervised when some degree of human input is required
- In semi-supervised case, a small set of samples have been manually labelled for classification (e.g., face images tagged by a human observer), but then also to leverage a very large set of unlabeled data



Partially Labelled Data

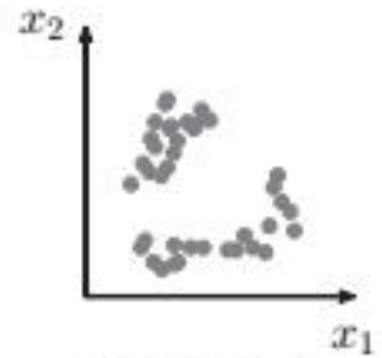

Land Use

Classifier and Classification

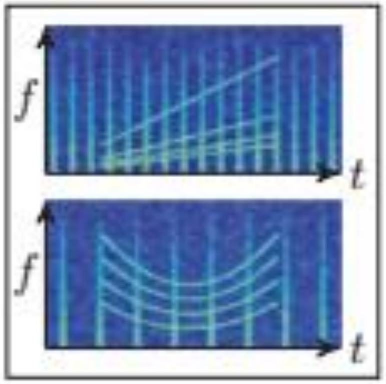
Pattern Recognition Problem: Unsupervised

Scenario 4 (Unsupervised): Model is not known, no labelled data available

- Pattern measurements are available, however the points have no associated class information; this is known as an unsupervised problem.
- The range of problems here is still very broad, depending on whether we are told the number of classes, or their typical size or separation, or perhaps nothing at all.
- Clustering problems are in Unsupervised category



Unlabelled Data



Data of Dolphin Vocalizations

PR Case Study: Biometric Recognition

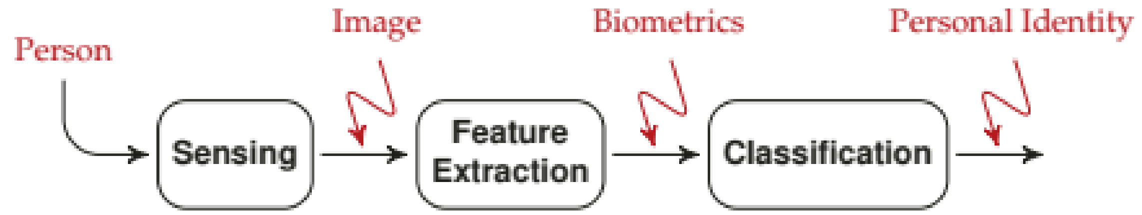


Fig. 2.7. BIOMETRIC RECOGNITION: In order to recognize a specific person on the basis of remotely-sensed biometrics, we need to acquire (sense) an image of some part of the body which has a unique signature (fingerprint, retina etc), extract biometric features from this image, and then develop a classifier that reliably recognizes the individual.

Biometrics

- Face
- Fingerprint
- Iris (the colored region in your eye around the pupil)
- Retina (the pattern of arteries in the back of your eye)
- Veins (vein structure in hand or arm)
- DNA

Basic Components of Biometric Recognition:

1. Image Acquisition
2. Feature Extraction
3. Classification

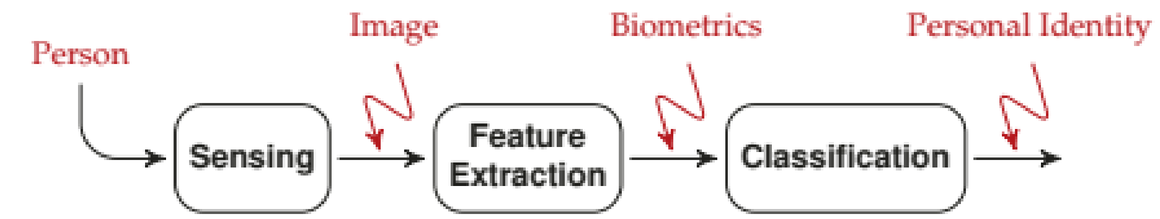
The classifier returns one member of the class set

$$C = \{NoMatch; Person_1; Person_2; : : : ; Person_K\}$$

Classifying x as *NoMatch* where *Person*1 is correct: Frustration

Classifying x as *Person*2 where *Person*1 is correct: Security Breach

PR Case Study: Biometric Recognition



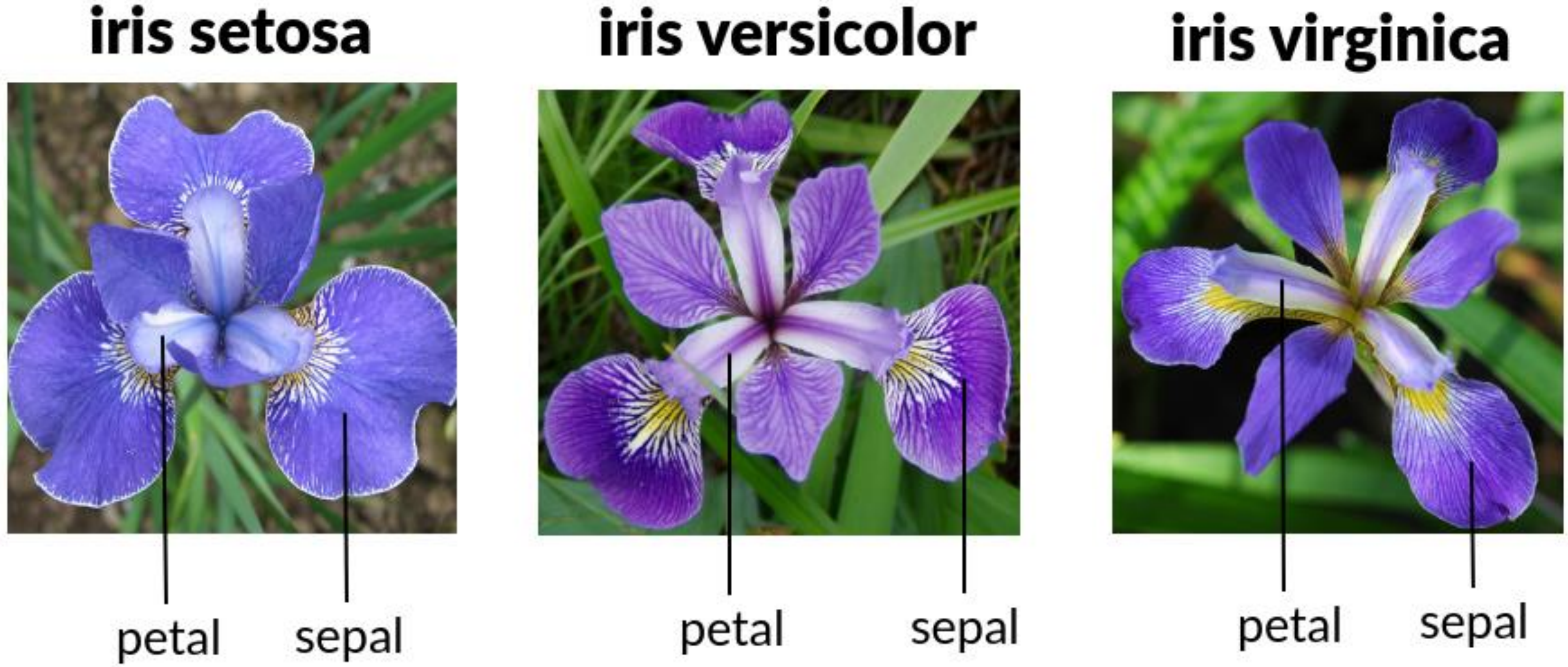
Classifying *x* as *NoMatch* where *Person*1 is correct: Frustration
Classifying *x* as *Person*2 where *Person*1 is correct: Security Breach

Avoiding frustration and security breaches, a successful biometric strategy must also satisfy

- Universality: Every person should be measurable, regardless of age and health
- Uniqueness: The feature vectors extracted for a given person should be robustly unique
- Consistency: For a given person the feature vector should be highly repeatable from one try to the next, and should be slowly (or not at all) varying over time.

Other criteria: non-invasiveness, social acceptability, or how easily the system would be to defeat via nefarious means and so on

PR Hands-on: Iris Flower Recognition



$C = \{C_1; C_2; C_3\}$ = {"Iris Setosa", "Iris Versicolor", "Iris Virginica"}

Each plant four measurements were taken:

$$y = \begin{pmatrix} \text{Sepal Length} \\ \text{Sepal Width} \\ \text{Petal Length} \\ \text{Petal Width} \end{pmatrix}$$

Open Discussion

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

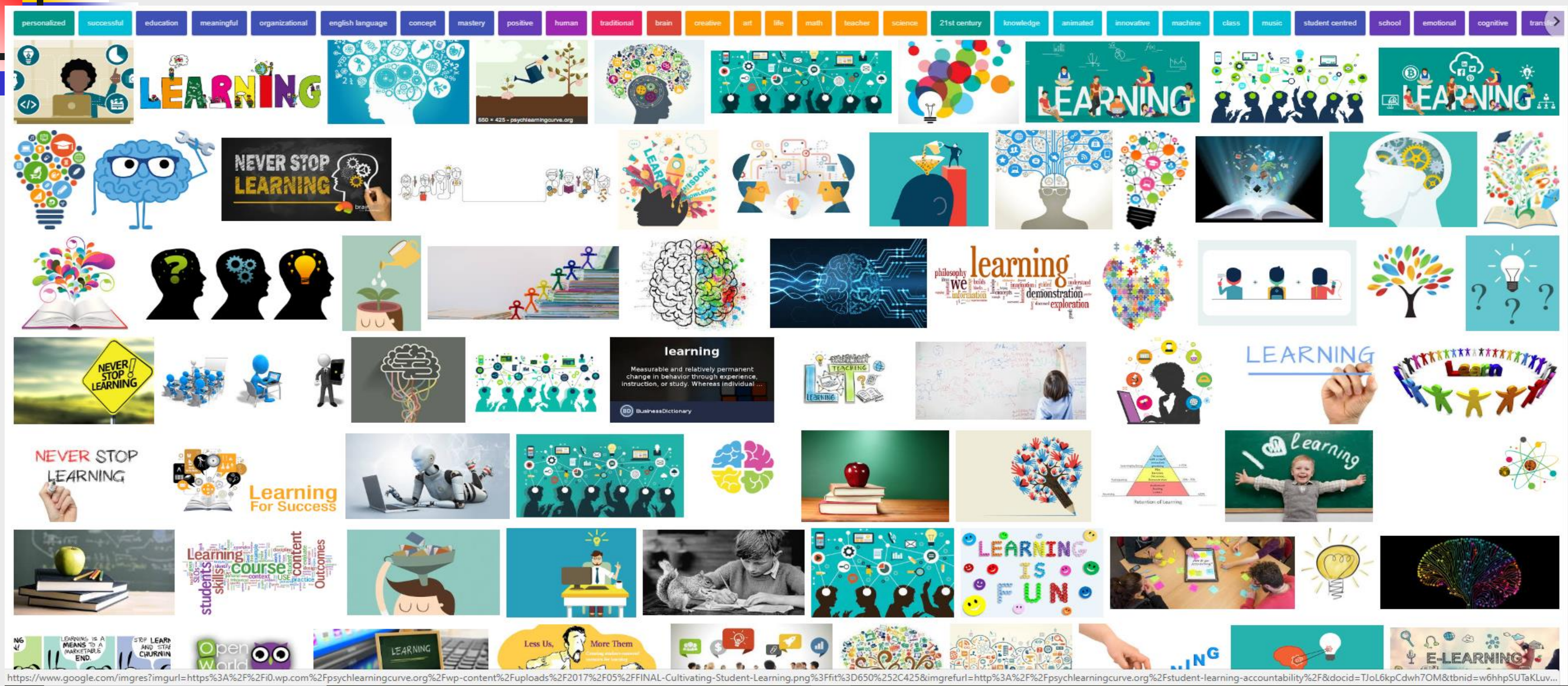
Learning is "a process that leads to change, which occurs as a result of experience and increases the potential for improved performance and future learning"
---(Ambrose et al, 2010, p.3).

"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***

3

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'*

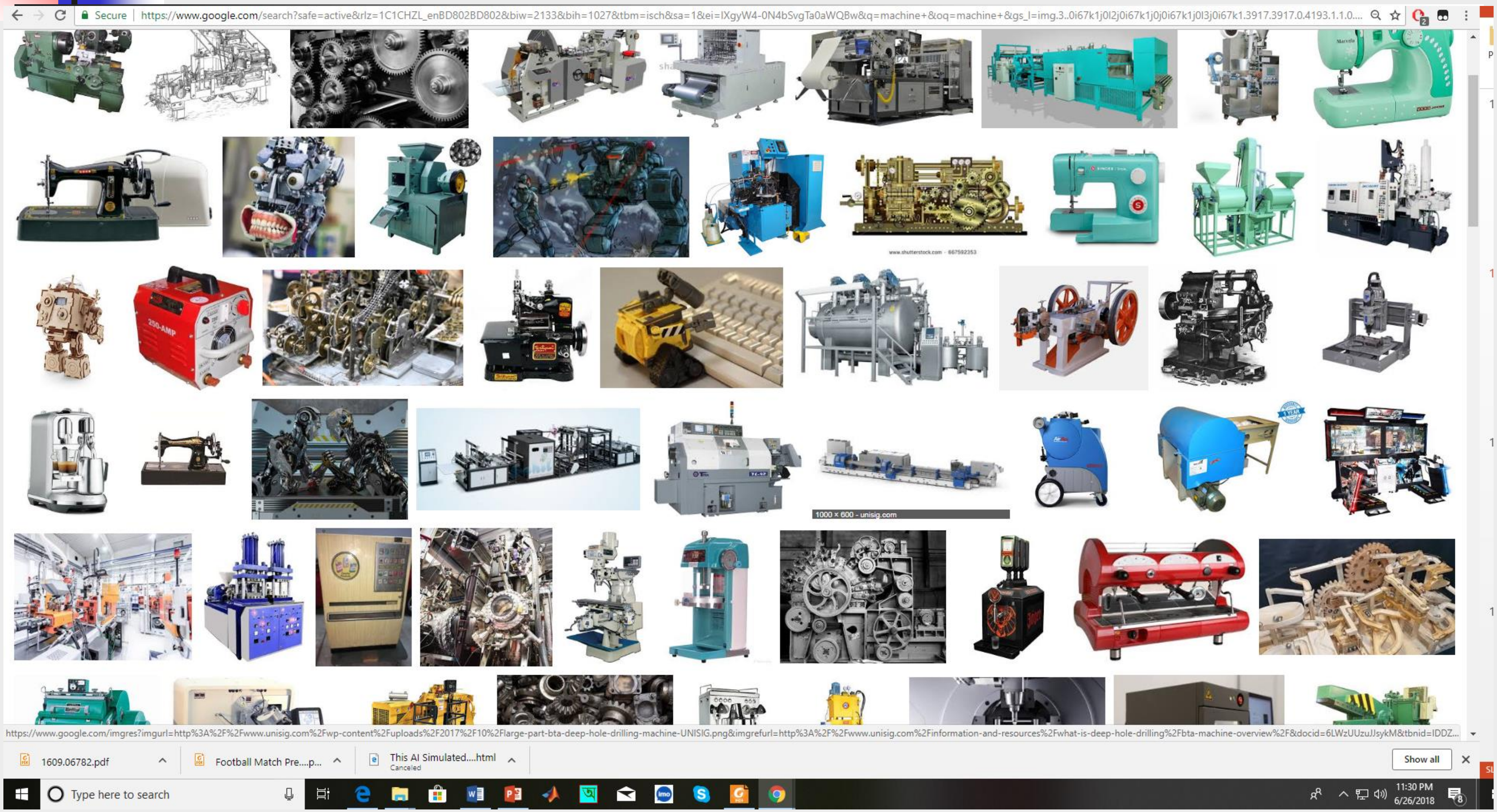
Oxford Dictionary

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

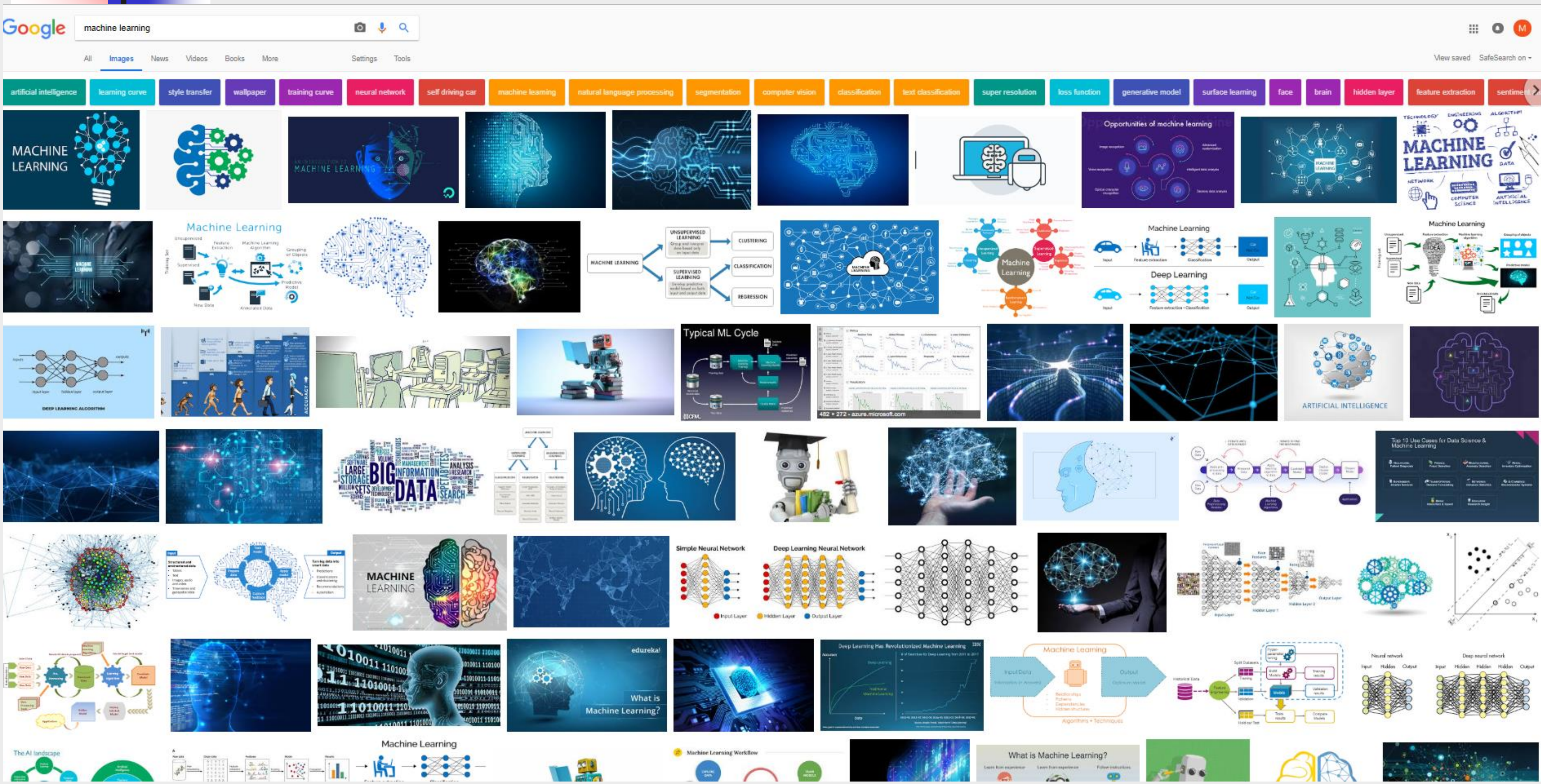
In 1959, Arthur Samuel, a pioneer in the field of machine learning (ML) defined it as the "field of study that gives computers the ability to learn without being explicitly programmed".

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]

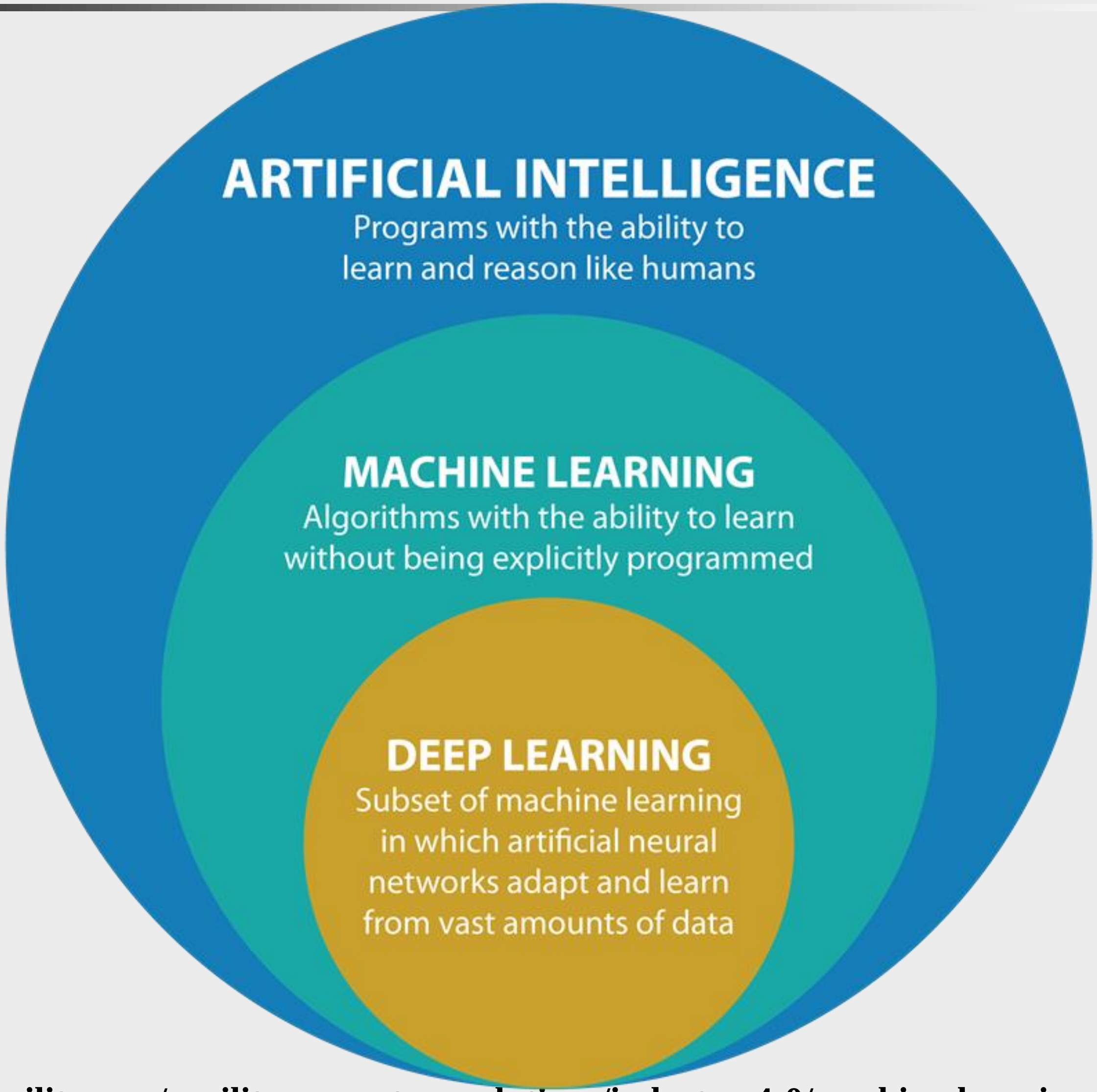
https://en.wikipedia.org/wiki/Machine_learning

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).

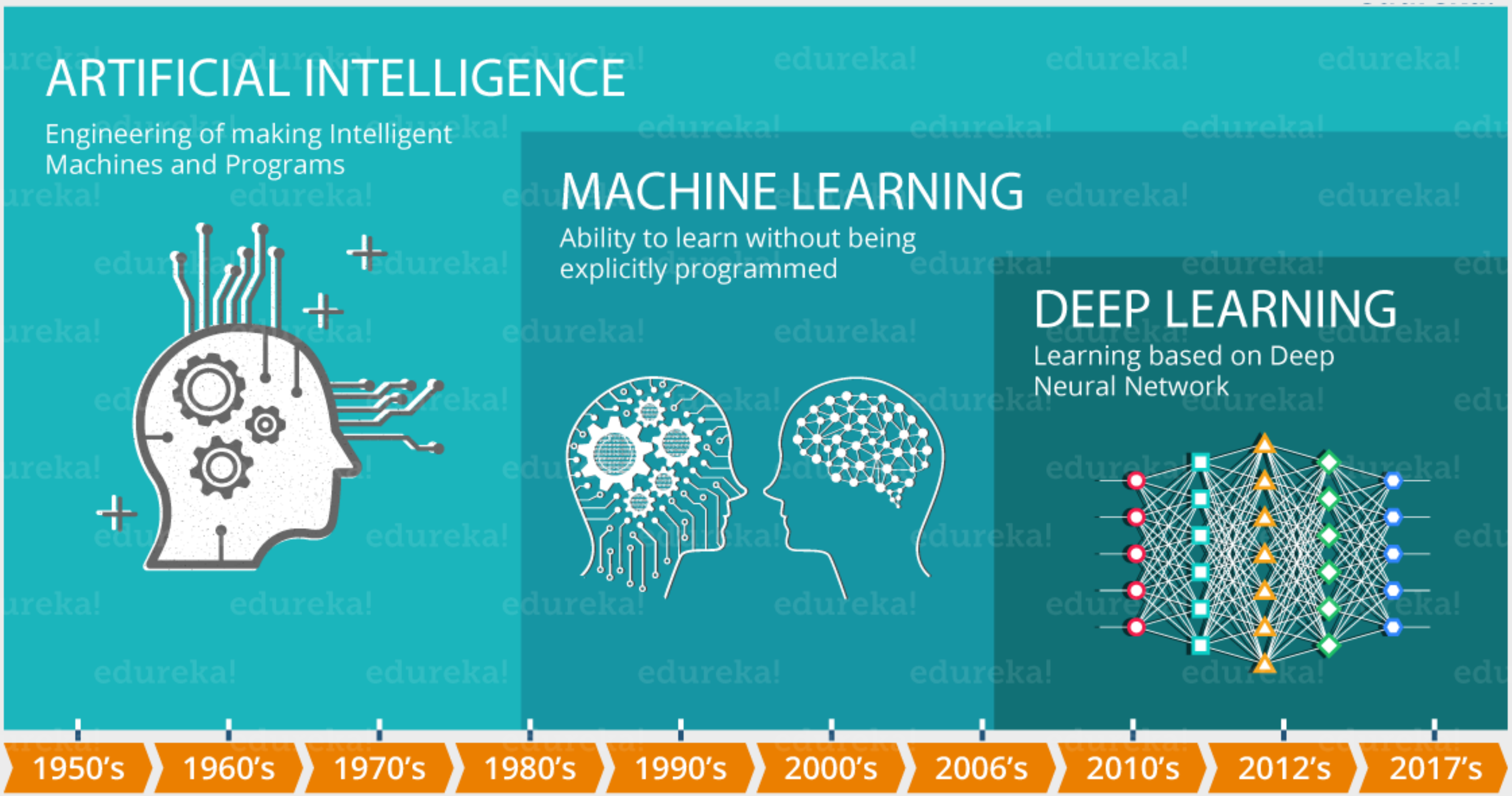
Machine Learning

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

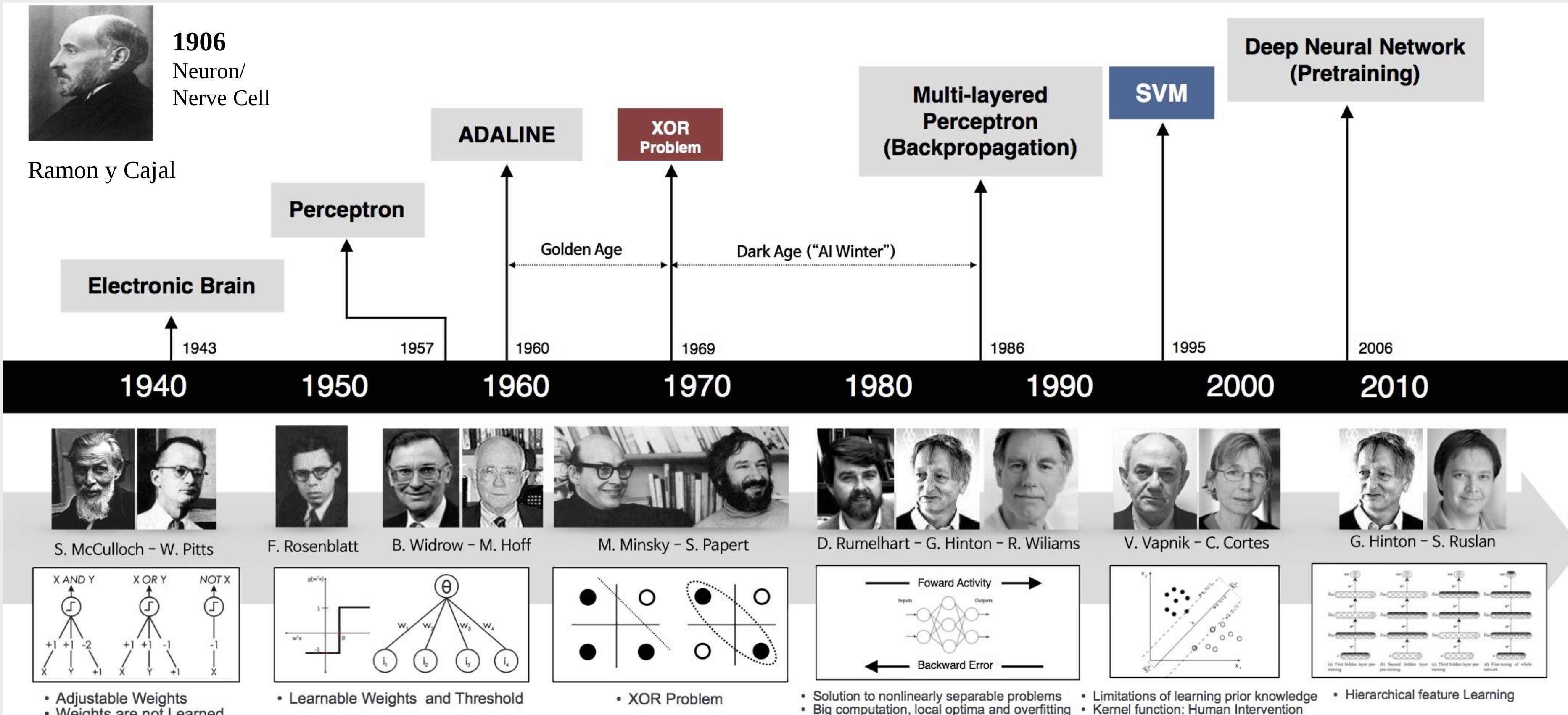
Deep Learning

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*:

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

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.

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

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

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

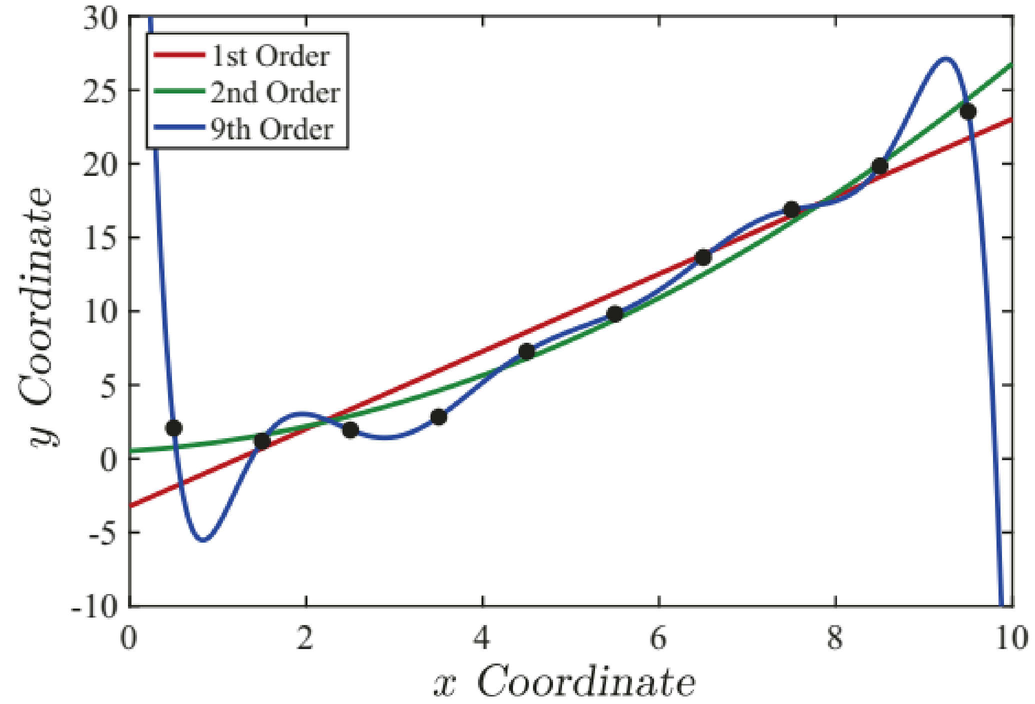
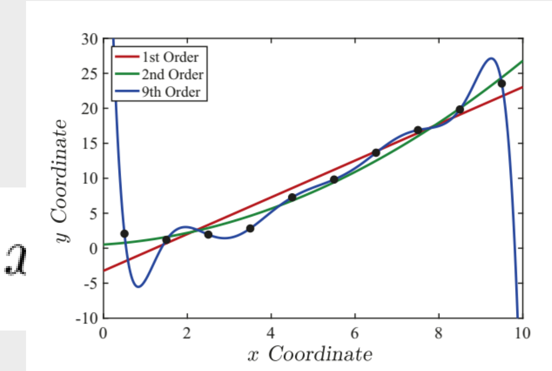


Fig. 3.2. What does it mean to *fit* a model to data? Here we have ten data points (black dots), which come from some unknown model with added random noise. We fit polynomials to the points, for which the first-order (linear regression), second-order (parabolic regression), and ninth-order fits are shown. 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

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



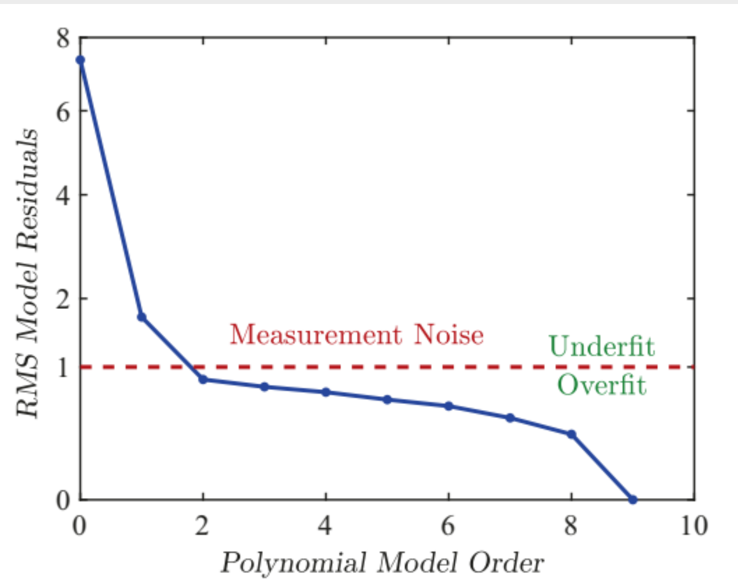


Fig. 3.3. Following up on Figure 3.2, we can plot the root-mean-square (RMS) difference between the polynomial fit and the data points, as a function of the polynomial order. In this case the measurement noise level was assumed to have been provided, it was not learned.

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

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

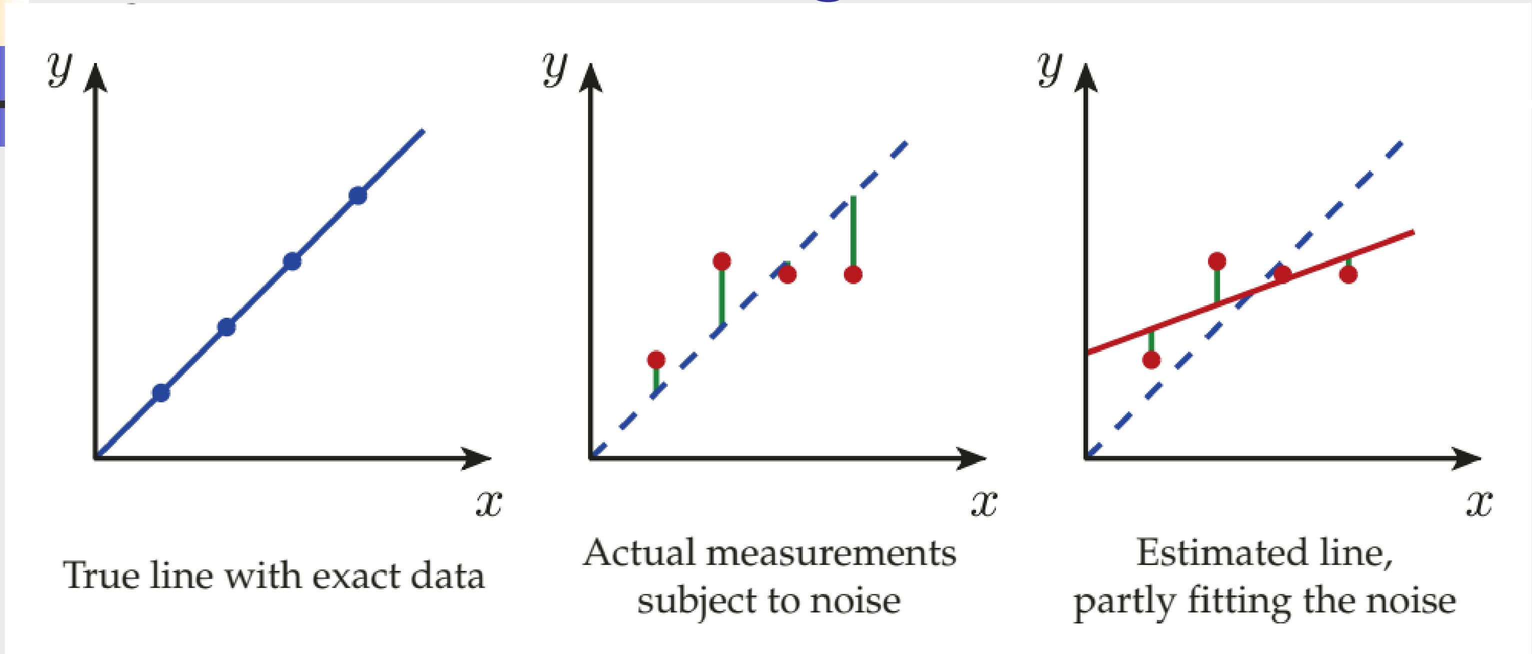
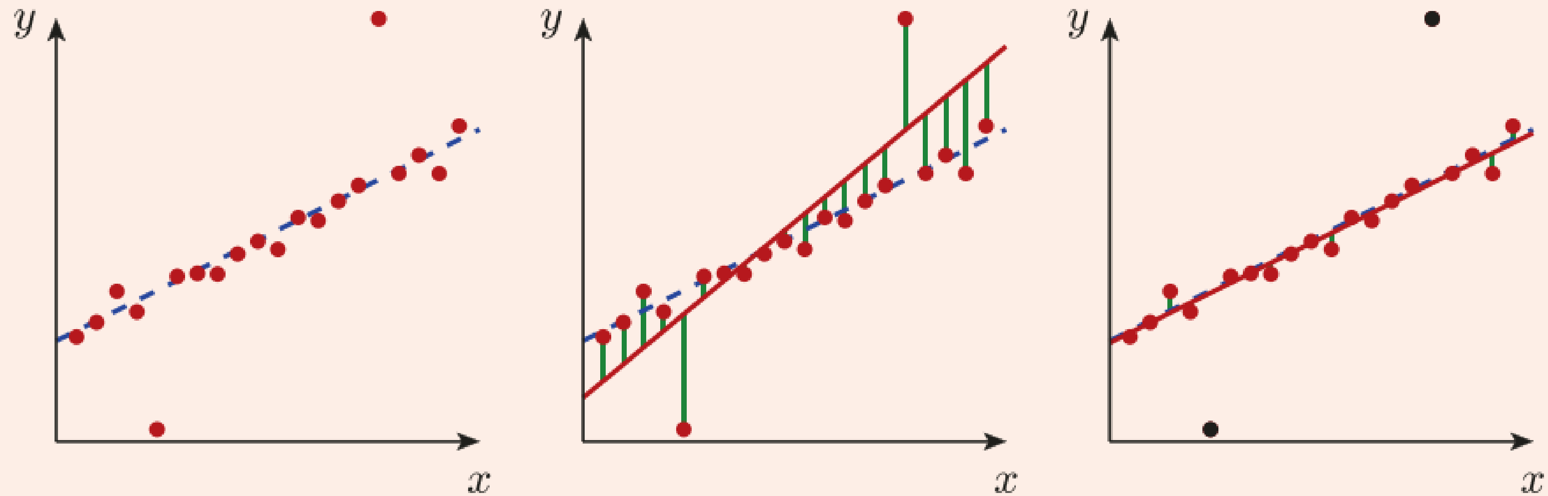


Fig. 3.4. Overfitting: Any learning, whether of linear regression (here) or a pattern recognition classifier (Figure 3.5), is said to be overfitting if it begins to tune its parameters to the behaviour of the noise, rather than of the underlying phenomenon we wish to learn. The estimated line (red) is 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



Actual measurements subject to noise having two outliers

Estimated line based on the RMS criterion of (3.8) (Not Outlier Robust)

Estimated line ignoring the two outliers (Outlier Robust)

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

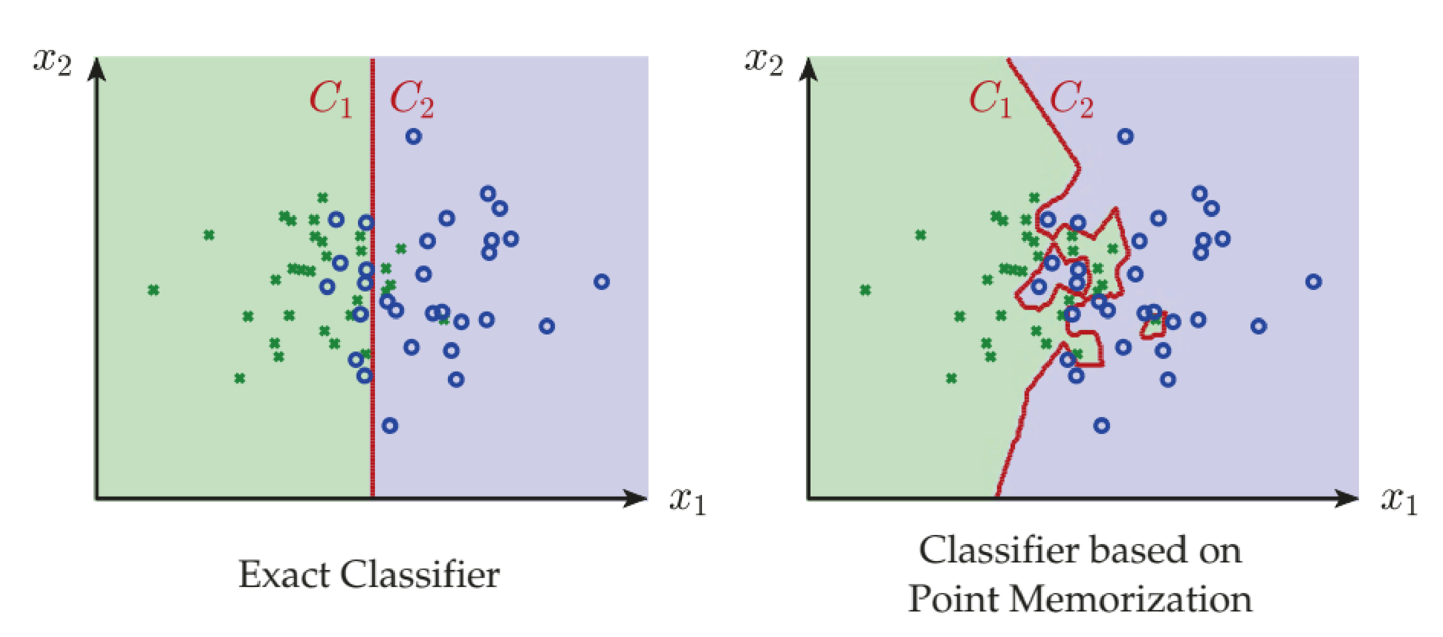
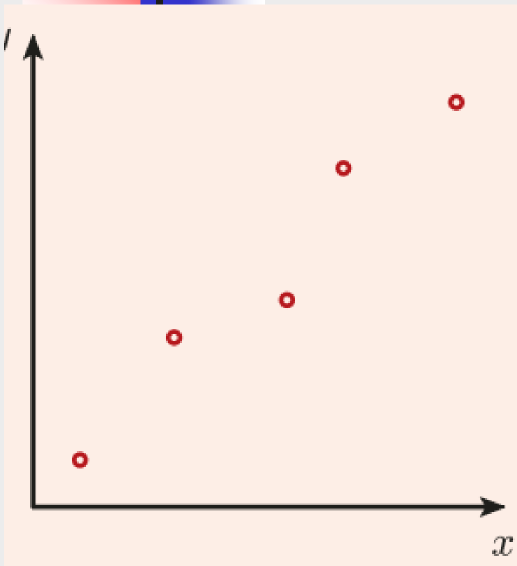
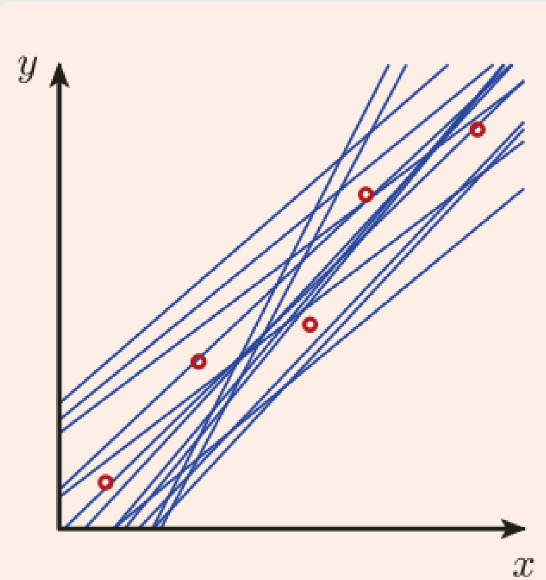


Fig. 3.5. OVERFITTING: As in Figure 3.4, but here for a pattern recognition classifier. We will have to wait for Chapter 6 for the details of the classifier to be discussed, however the principle is the same as in regression: Any learning is overfitting if it tunes its parameters to the behaviour of 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



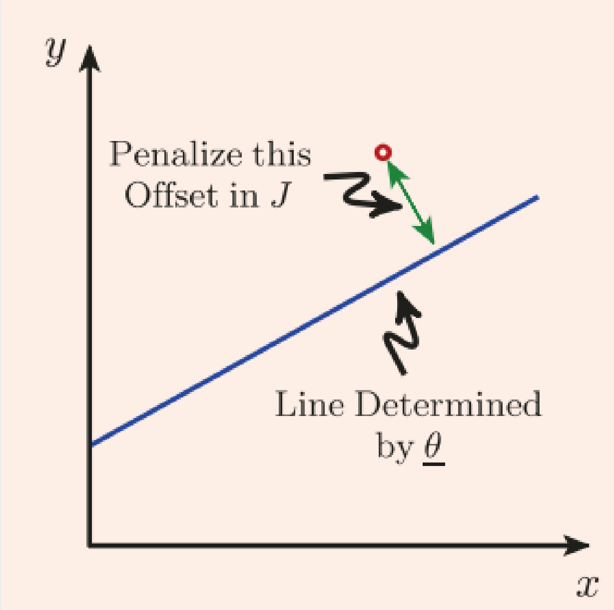


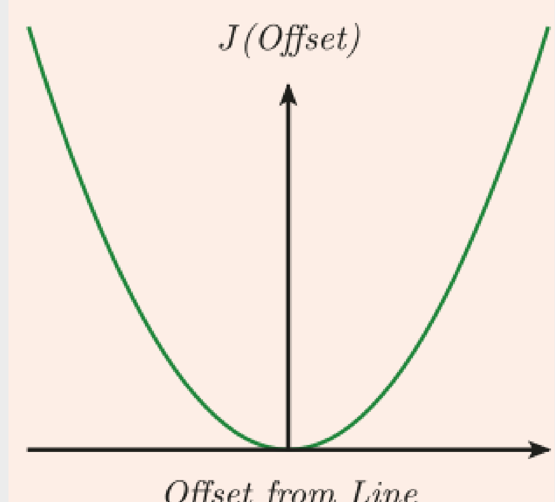
For each possible line, described by

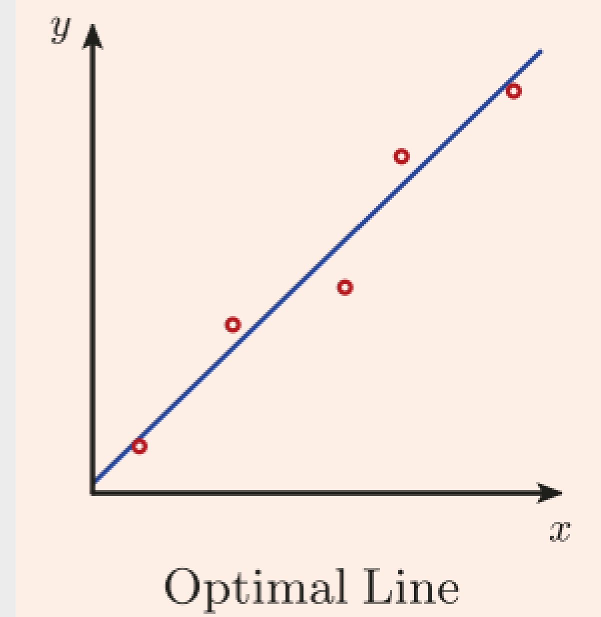
$$\underline{\theta} = \begin{bmatrix} \text{Angle of Line} \\ \text{Y-Intercept of Line} \end{bmatrix}$$

we can assess the penalty associated with the
as

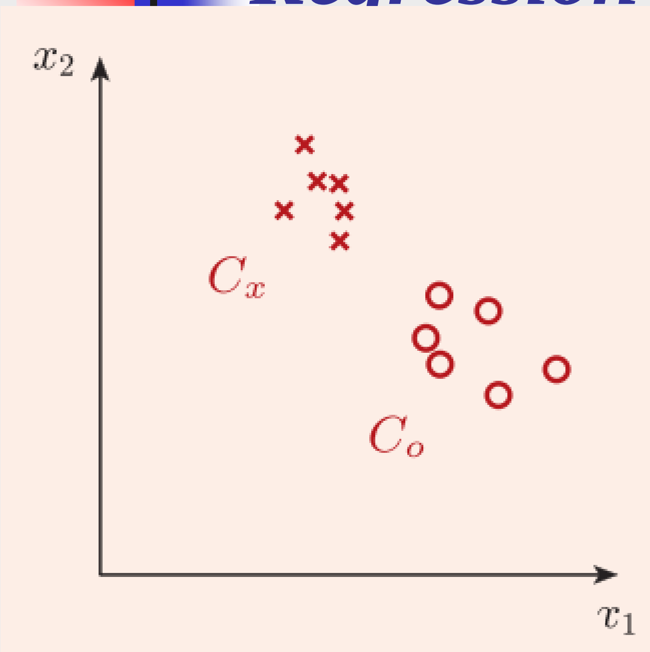
$$\text{Penalty}(\underline{\theta}) = \sum_i J\big(\text{Offset from line } \underline{\theta} \text{ to } (x_i,$$

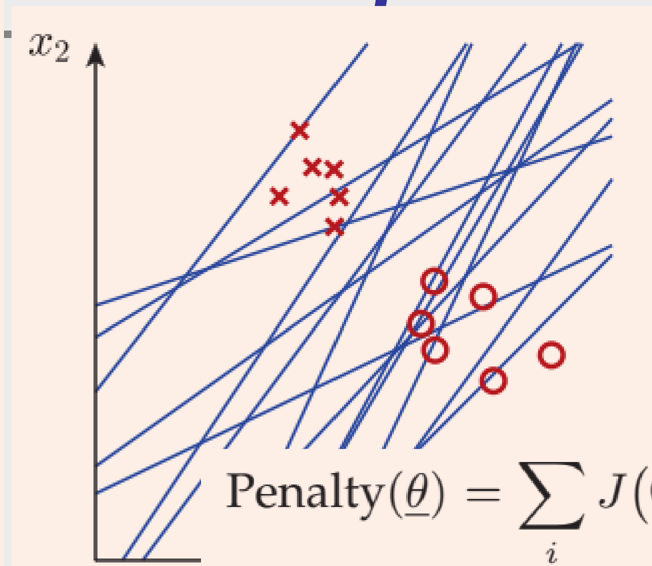




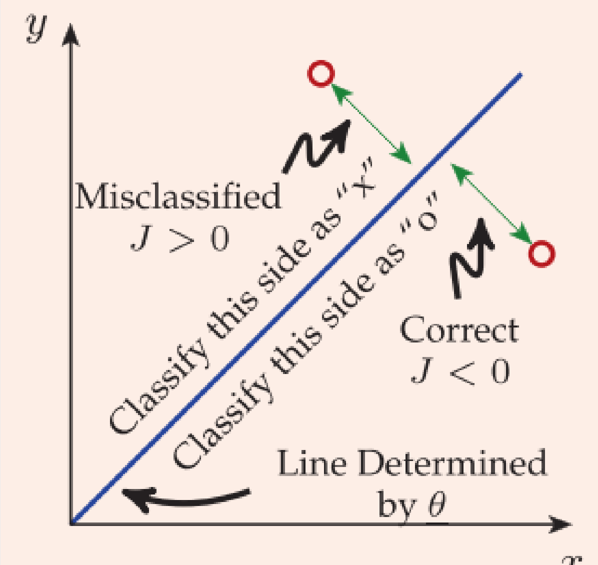


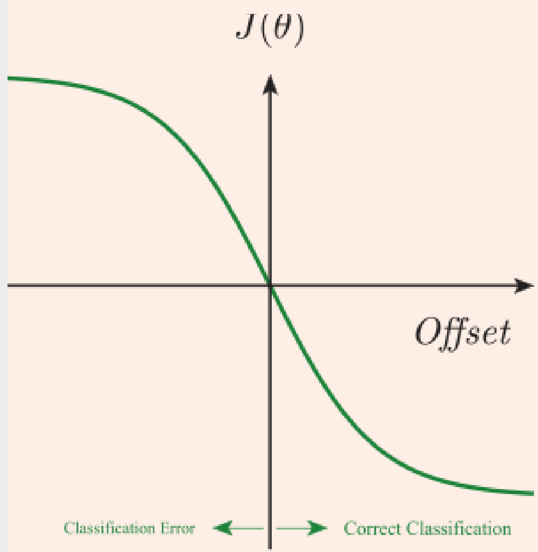
Regression and Classification

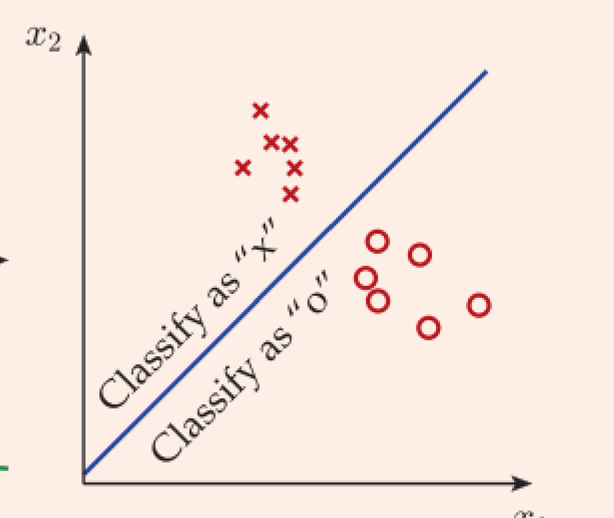




$$\text{Penalty}(\underline{\theta}) = \sum_i J\big(\text{Offset from line } \underline{\theta} \text{ to } \underline{x}_i \in C_x\big) + \sum_j J\big(\text{Offset from line } \underline{\theta} \text{ to } \underline{x}_j \in C_o\big)$$







Regression and Classification

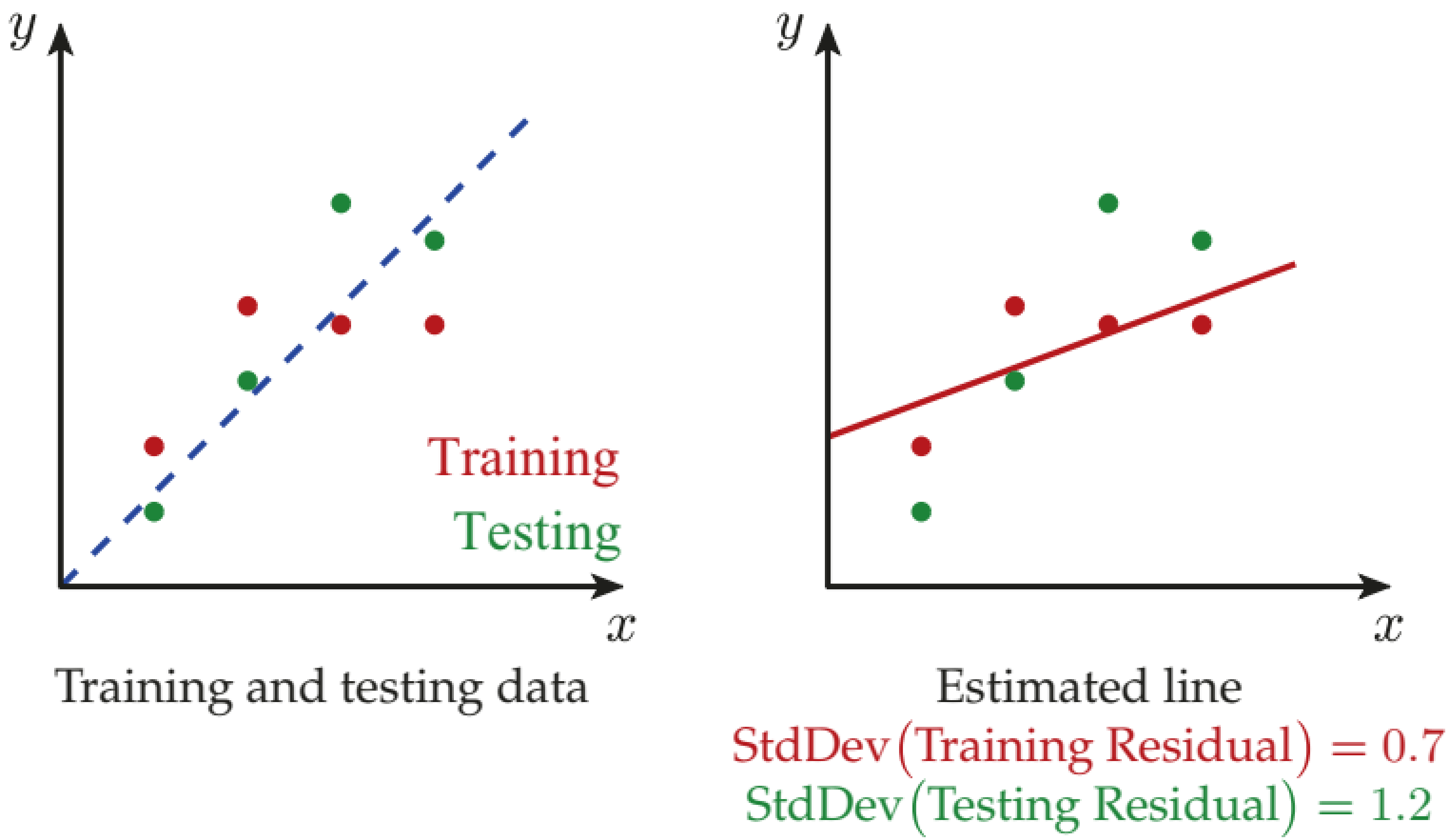
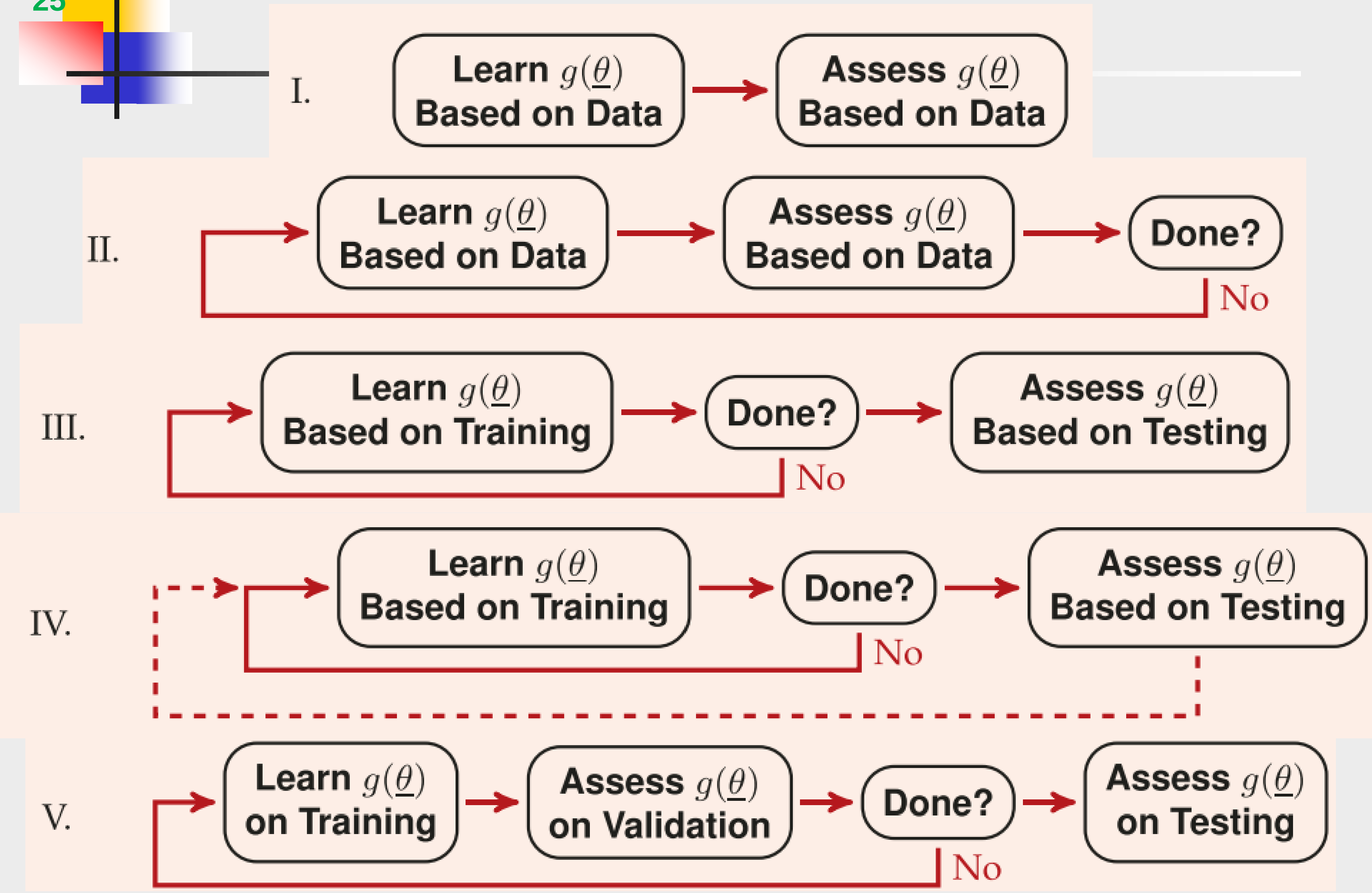


Fig. 3.7. We have separate training and testing data (left), such that the learned model (red line) is deduced from the training data, but assessed against the testing data. Observe the degree to which the estimated line fits to the noise, based on the difference between the fit to training data (overfit,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

Use of Data in Learning



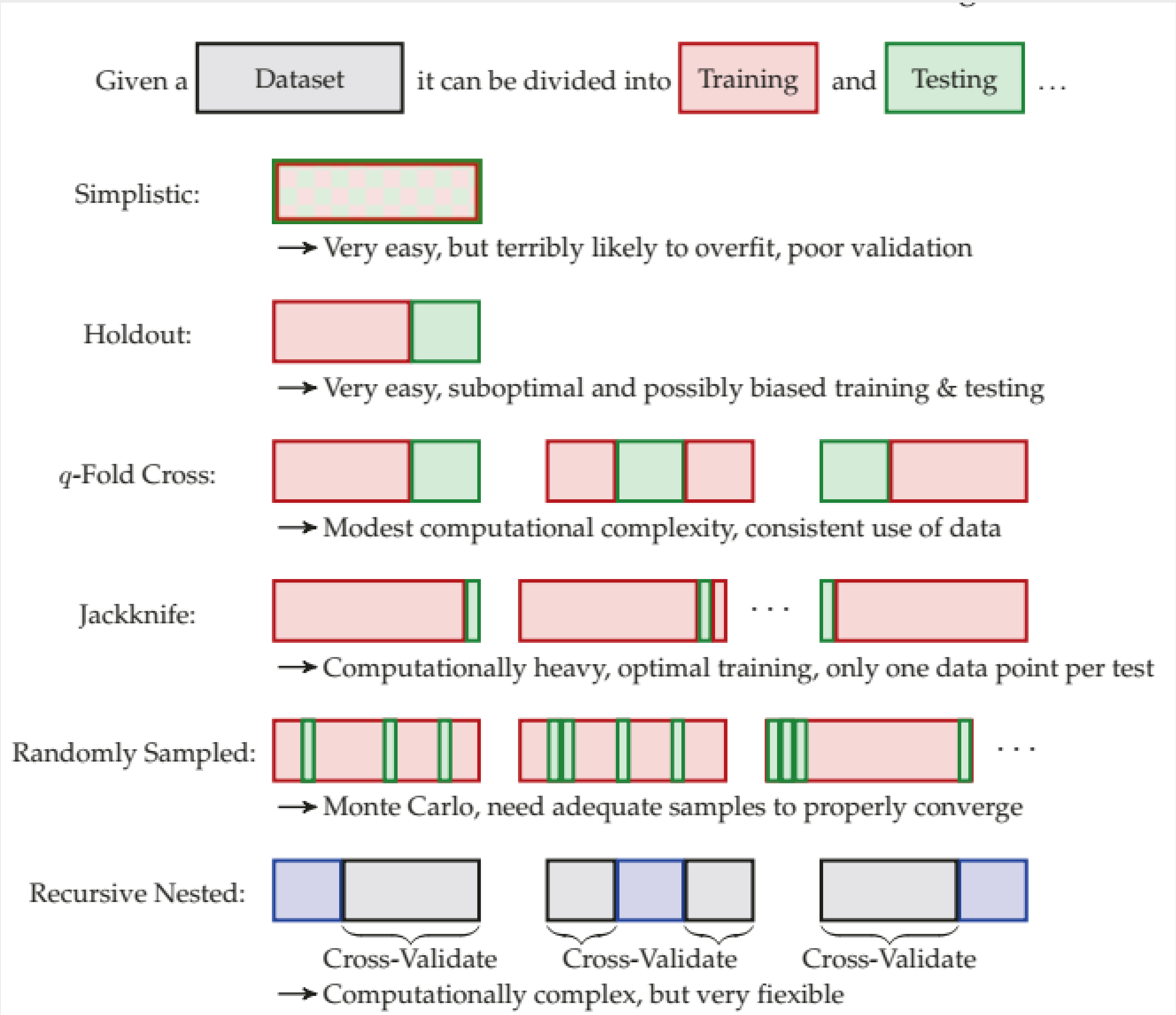
Classifier Evaluation / Performance Measure

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

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

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

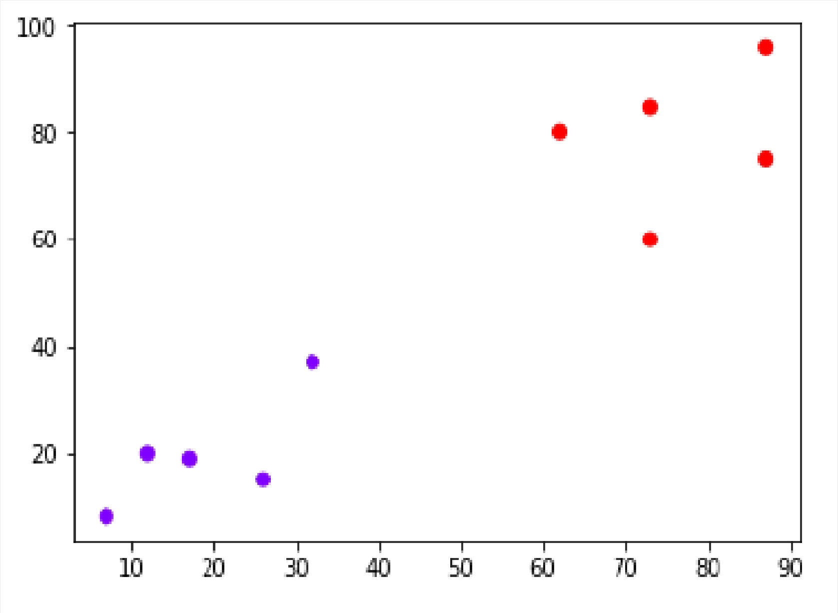
Classifier Validation

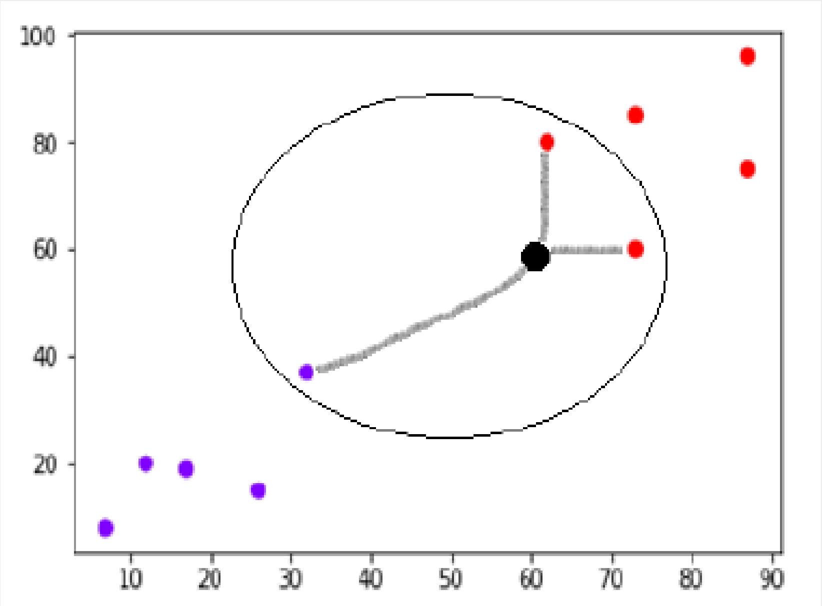


K-nearest neighbors (KNN) Classifier

•**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.



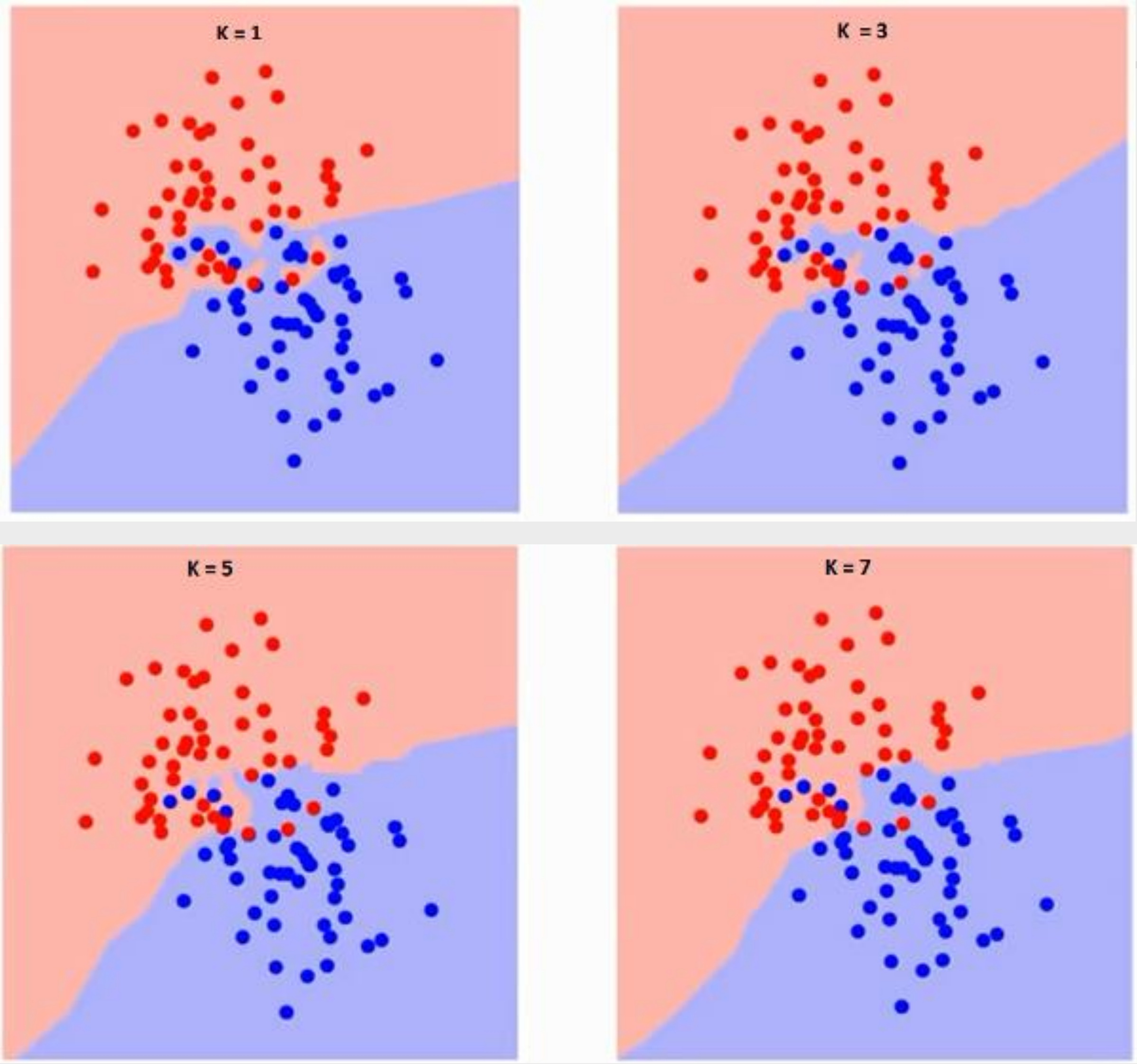


https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm

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