IMAGE DATA ANALYSIS (6CFU)

MODULE OF

REMOTE SENSING

(9 CFU)

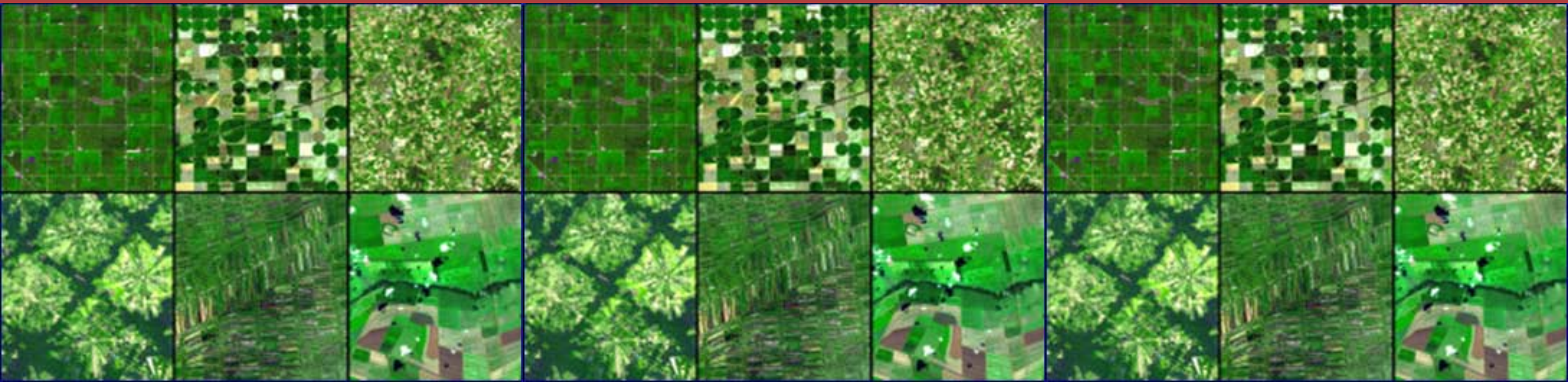
A.Y. 2022/23

MASTER OF SCIENCE IN COMMUNICATION TECHNOLOGIES AND MULTIMEDIA

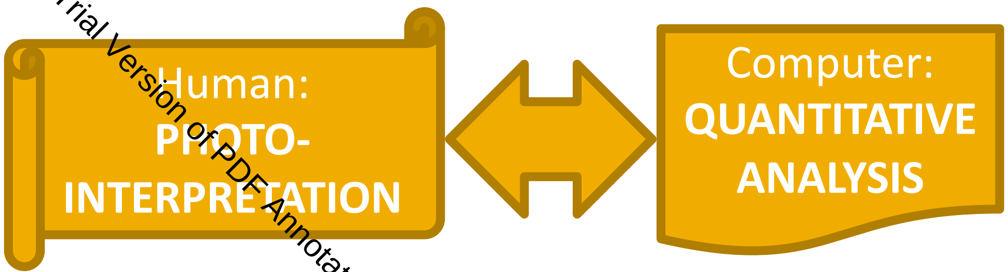
MASTER OF SCIENCE IN COMPUTER SCIENCE, LM INGEGNERIA INFORMATICA

PROF. ALBERTO SIGNORONI

THE INTERPRETATION OF DIGITAL IMAGE DATA



Approaches to Interpretation



Notice: most of contents referred to the Remote Sensing domain can be easily extended to other domains qualified by the presence of **visual interpretation professionals**, e.g. radiologists, pathologists, physicists, manufacturing quality expert, botanists,…

- **Photointerpretation (by human data analysts)**, involving direct human interaction and therefore high-level decisions, is good for spatial (shape, size, orientation and texture) assessment and recognition but poor in quantitative accuracy
  - Roads, coastlines and river systems, fracture patterns, and lineaments generally, are usually readily identified by their spatial disposition.
  - Temporal data, such as the change in a particular object or cover type in an image from one date to another can often be used by the photointerpreter as, for example, in discriminating deciduous or ephemeral vegetation from perennial types.
  - Spectral clues are utilised in photointerpretation based upon the analyst's foreknowledge of, and experience with, the spectral reflectance characteristics of typical ground cover types, and how those characteristics are sampled by the sensor on the platforms used to acquire the image data.

- **Quantitative analysis (by computers)**, possibly requiring little or no human interaction, has reduced spatial reasoning ability but high quantitative and discrimination accuracy.
  - Its high accuracy comes from the ability of a computer to process (possibly all) pixels in a given image and to take account of the full range of spectral, spatial and radiometric detail present.
    - Its poor spatial properties come from the relative difficulty with which decisions about shape, size, orientation and texture can be made using standard sequential computing techniques.
    - **NB** Improved spatial interpretation «capabilities» can originate from *deep learning* approaches.

Approaches to Interpretation: the classic view of complementarity

- Those two approaches to image interpretation have their own roles and often these are **complementary**:
  - photointerpretation is aided substantially if a degree of digital image processing is applied to the image data beforehand, while
  - quantitative analysis depends for its success on information provided at key stages by a visual data analyst. This information very often is drawn from photointerpretation.
    - Its possible exploitation spans from simple parameter/filter selection to large annotated data collection for training a supervised machine learning system

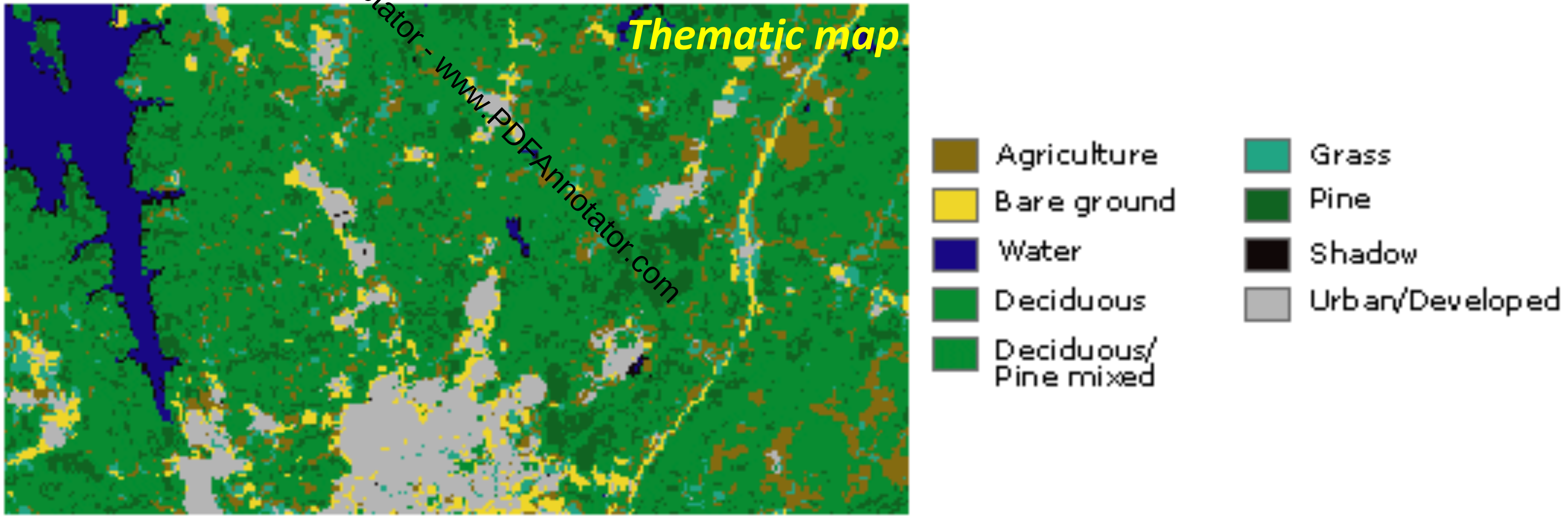
Photointerpretation (by a human analyst/interpreter)	Quantitative analysis (by computer)
On a scale large relative to pixel size	At individual pixel level
Inaccurate area estimates	Accurate area estimates possible
Only limited multispectral analysis	Can perform true multispectral (multidimensional) analysis
Can assimilate only a limited number of distinct brightness levels (say 16 levels in each feature)	Can make use quantitatively of all available brightness levels in all features (e.g. 256, 1024, 4096)
Shape determination is easy	~~Shape determination involves complex software decisions~~
Spatial information is easy to use in a qualitative sense	~~Limited techniques available for making use of spatial data~~

Not true anymore with today machine learning

Approaches to Interpretation: pixel-wise classification

- In computer-based quantitative analysis the attributes of each pixel (such as the spectral bands available) are examined in order to give the pixel a **label** identifying it as belonging to a particular *class of* ***pixels*** *of interest* to the user.

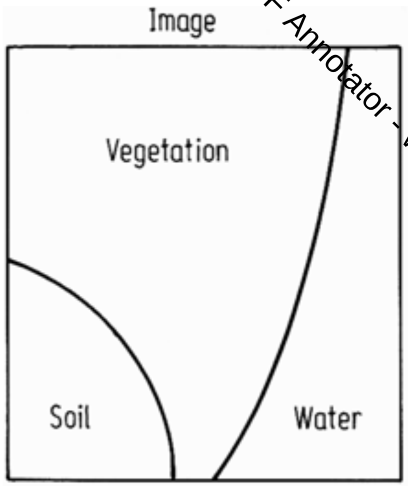
- As a result, the process is often also called ***classification***.



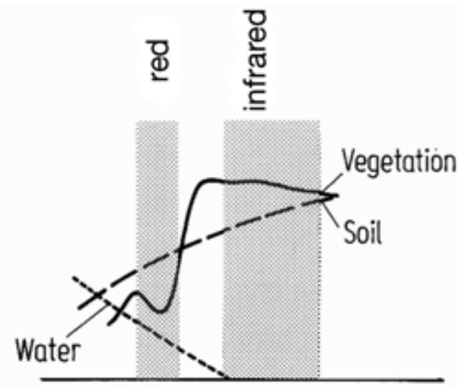
  - The human analyst is unable to apply photointerpretation at pixel level (unless only small groups of pixels are of interest) and to discriminate to the limit of the radiometric resolution generally available (only three or so of the complete set of spectral components of an image can be used readily, by color composite images).
  - When a computer can be used for analysis (data are in digital format), it can do so ***at the pixel level*** and can examine and identify as many pixels as required. In addition, computer-based analysis allow to ***take full account of the multidimensional aspect of the data*** including its full radiometric resolution.

Pixel-wise classification in the Pattern Space

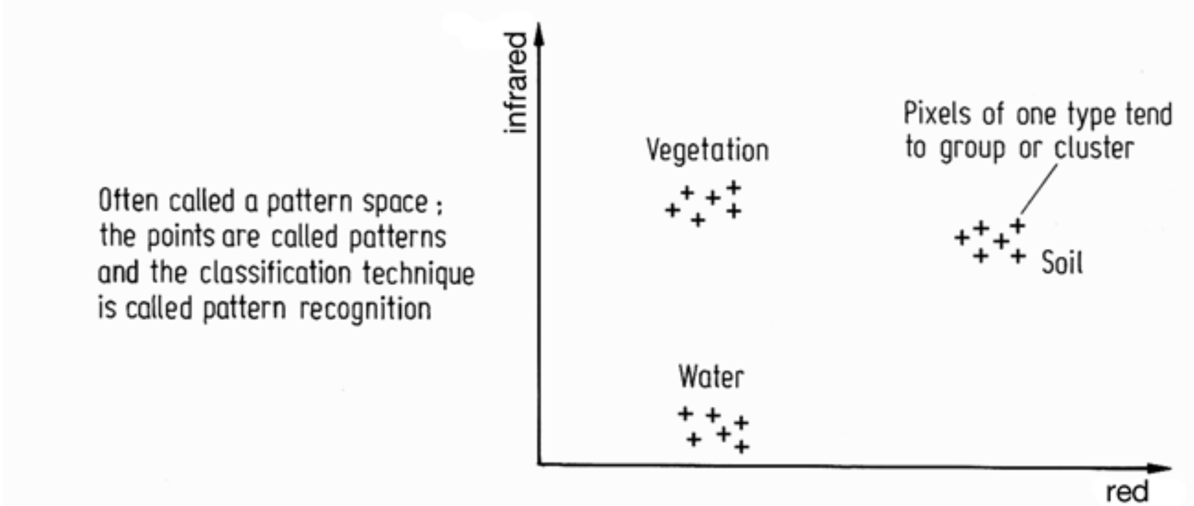
- An effective means by which multispectral data can be represented in order to formulate algorithms for quantitative analysis is to plot them in a **pattern space**, or multispectral vector space, with as many dimensions as there are spectral components.



Goal:
thematic map
(pixel-based
or object-based)
of spectral classes
(**classification**)



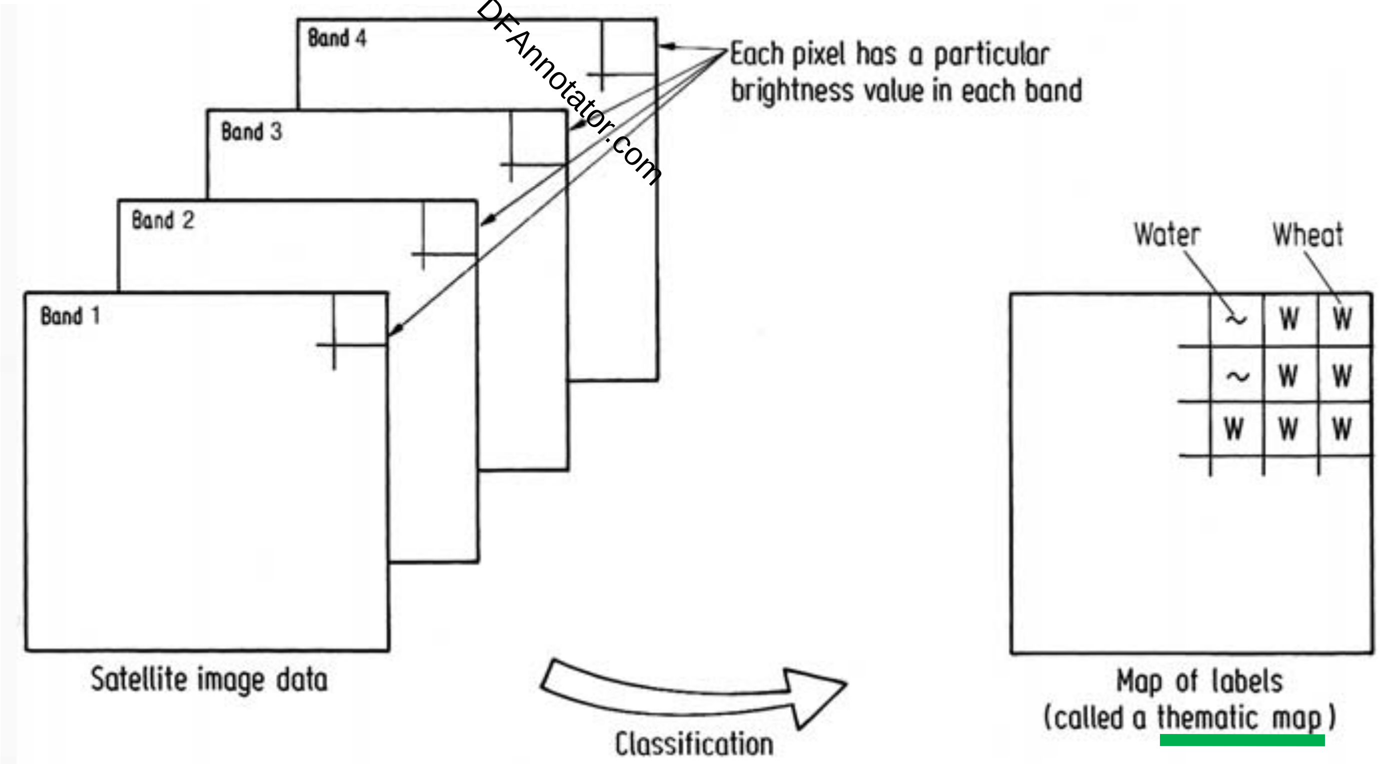
Spectral reflectance characteristics of different ground cover types



In pattern space, each pixel of an image plots as a point with **coordinates given by the brightness value of the pixel in each spectral component**.

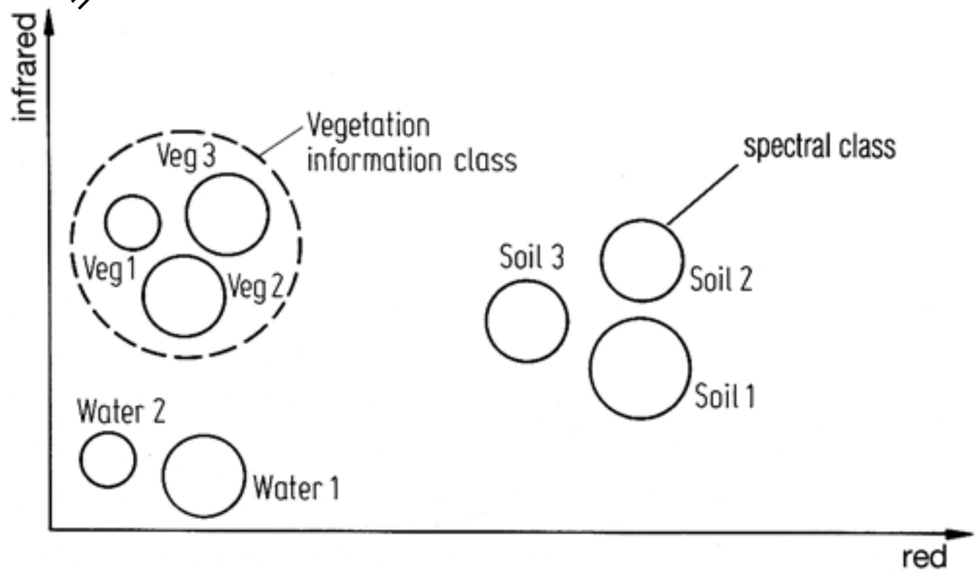
Pixel-wise classification in the Pattern Space

- Recognition that image data exists in sets of spectral classes, and identification of those classes as corresponding to specific ground cover types, is carried out using the techniques of mathematical **pattern recognition** or *pattern classification* and their more recent *machine learning* implementations.
  - The patterns are the multiband *pixel vectors* themselves, which contain the sets of brightness values *for* the pixels arranged in column form $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$



Multispectral Pattern Space: Information vs Spectral Classes

- Provided the spectral bands have been designed to provide good *discrimination* it is expected that pixels would form *groups* in multispectral space *corresponding to various ground cover types*, the sizes and shapes of the groups being dependent *upon varieties of cover type, systematic noise* and *topographic effects*.
  - The groups or clusters of pixel points related to **target cover types** are referred to as **information classes** since they are the actual classes of data which a computer will need to be able to recognize.
- In practice *the information class groupings may not be single clusters* as depicted in Figure. Instead, it is not unusual to find *several clusters for the same region in the pattern space* (e.g. of soil or vegetation), i.e. for the same apparent cover type we want to discriminate.
  - These are not only as a result of specific differences in types of cover but also result from differences in moisture content, soil types, underlying vegetation and topographic influences (inclination, shadows).
  - Consequently, a multispectral space is more likely to appear as shown in Figure in which each information class is seen to be composed of several so-called **spectral classes**.

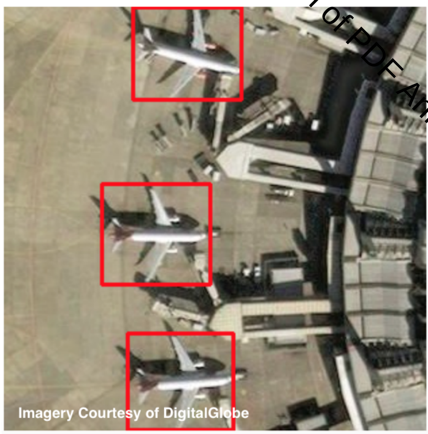


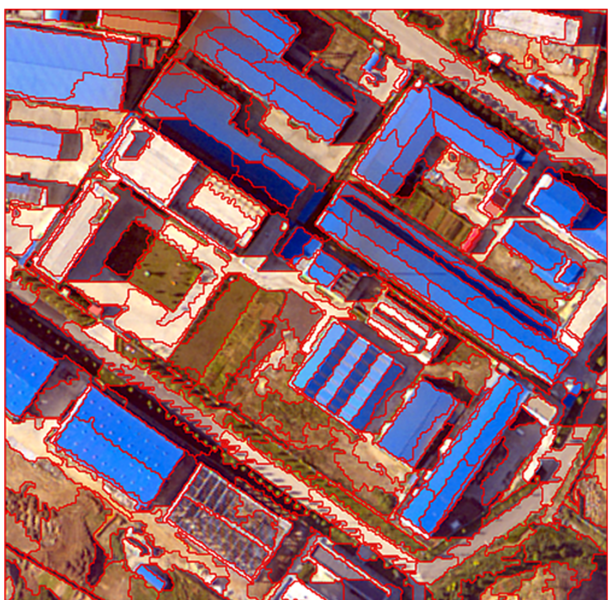
Multispectral Pattern Space: Information vs Spectral Classes

- In many cases the information classes of interest do not form distinct clusters or groups of clusters but rather are part of a *continuum of data* in the multispectral space.
  - This happens for example when, in a land systems exercise, there is a gradation of canopy closure with position so that satellite or aircraft sensors might see a gradual variation in the mixture of canopy and understory. The information classes here might correspond to nominated ***percentage mixtures*** rather than to sets of well defined subclasses as depicted in Figure.
  - It is necessary in situations such as this to determine appropriate sets of spectral classes that represent the information classes.
- In quantitative analysis it is the spectral classes that a computer will be typically asked to work with since they are the “natural" groupings or clusters in the data.
  - After quantitative analysis is complete it could happen that *the analyst (or the same learning system) simply associates* all the relevant spectral classes found with the one appropriate information class.
    - In the context of statistical models, widely adopted to classification, spectral classes will be seen to be unimodal probability distributions and information classes as possible multimodal distributions. The latter need to be resolved into sets of single modes for convenience and accuracy in analysis.
    - In the context of modern learning systems these aspects are inherently solved (information classes can be directly learned without the need to see them split in unimodal components) but this could risk to hinder a deeper understanding of data distributions.

Approaches to Interpretation: object-wise detection/classification

- Image classification can also be conceived as **object-driven** when the goal is to detect, segment or classify an object, or more objects in the image.



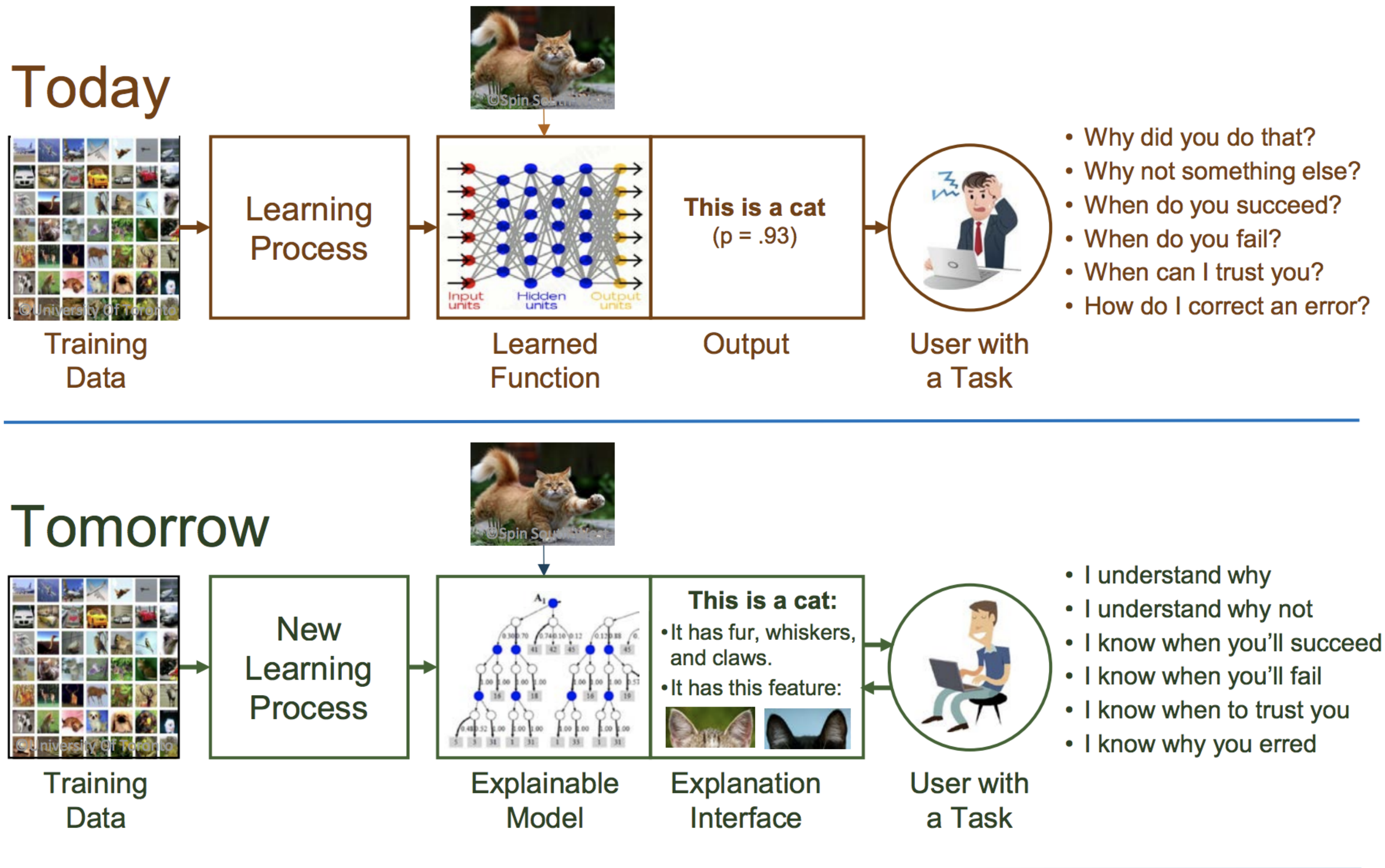


- In the case of object-level detection humans are good but modern analysis techniques outperform humans in rapid categorization tasks, when humans do not have enough time to complete them.
- Humans may remain superior to machines when a high level of contextual information/knowledge or of experience is needed to correctly interpret images.
- On the other hand, machines can be superior to even experienced humans for challenging or subtle interpretation tasks.

NB in this case the pattern space is made of space-related components (entire images or portions of them enter in the classifier) or a combination of spatial and spectral. Thus pattern space dimensionality tends to increase.

Approaches to Interpretation: the current "interpretability" perspective

- For some years now (Deep Learning era) human-machine interplay, within and beyond the context of image interpretation, has taken on new perspectives: **machine-based interpretation need to be** **human interpretable**



Definition and types of Machine learning

- **Machine learning**
  - Machine Learning is the ability to teach a computer without explicitly programming it
  - Examples are used to train computers to perform tasks that would be difficult to program
  - The computer can adopt a “model” among available ones to accomplish a certain task

- **Datasets** are usually divided in
  - ***Training set***: the data used to learn a model
  - ***Validation set***: is used to select a (statistical) model from a set of candidate ones, or to select the parameters of a model. If more than a model guarantee the same performance the simpler is selected (Ockham’s razor principle)
  - ***Test set***: like the validation set is independent from the training set and it is used to assess the performance (e.g. accuracy) of the model

- **Unsupervised Learning**: Training data is unlabeled - Goal is to (blindly) categorize the observations, e.g. to discover information or to help subsequent direct human interpretation or to ease data collection for supervised learning
- **Reinforcement Learning**: Training data is unlabeled - System receives feedback (rewards) for its actions - Goal is to perform better actions (typical of robotics or autonomous game play) - *not treated in this course*
- **Supervised Learning**: Training data are labeled - Goal is correctly label new data

Machine Learning overview

	Supervised	Unsupervised	Reinforcement
Training data	Data and correct output	Data	States, actions, and rewards
Learning target	Data-output relationship	Patterns in data	Policy
Evaluation	Statistics	Fitness	Reward value
Typical application	Classifiers	Clustering	Controllers

Applications of Machine Learning: imaging and beyond

- Language Translation
  - translate spoken and or written languages (e.g. Google Translate)
- Speech Recognition
  - convert voice snippets to text (e.g. Siri, Cortana, and Alexa)
- **Image Classification**
  - label images with appropriate categories (e.g. Multispectral images, Photo collections,...)
- Handwriting Recognition
  - convert written letters into digital letters
- Autonomous Driving
  - enable cars to drive
- Many other practical problems (spam detection, medical diagnosis, face and biometric recognition, anomaly detection, weather or financial forecast, fault prediction, automatic game playing, automatic image caption generation, industrial quality inspection...)

- **Features** are the observations that are used (and are relevant) to form predictions
  - ...**a very general concept** (from low-level to high-level semantic/structural content)
  - For image classification, the pixels (e.g. multispectral) are the features (patterns)
    - but other data extracted from images can become features (like for SIFT)
  - For voice recognition, the pitch and volume of the sound samples are the features
  - For autonomous cars, data extracted from the cameras, range sensors, and GPS are features or data from which to extract features.

Applications of Machine Learning: classification and beyond

- Main application in imaging
  - classification
  - detection (localization + classification)
  - segmentation (accurate localization)
  - others (style transfer, image generation, ...)

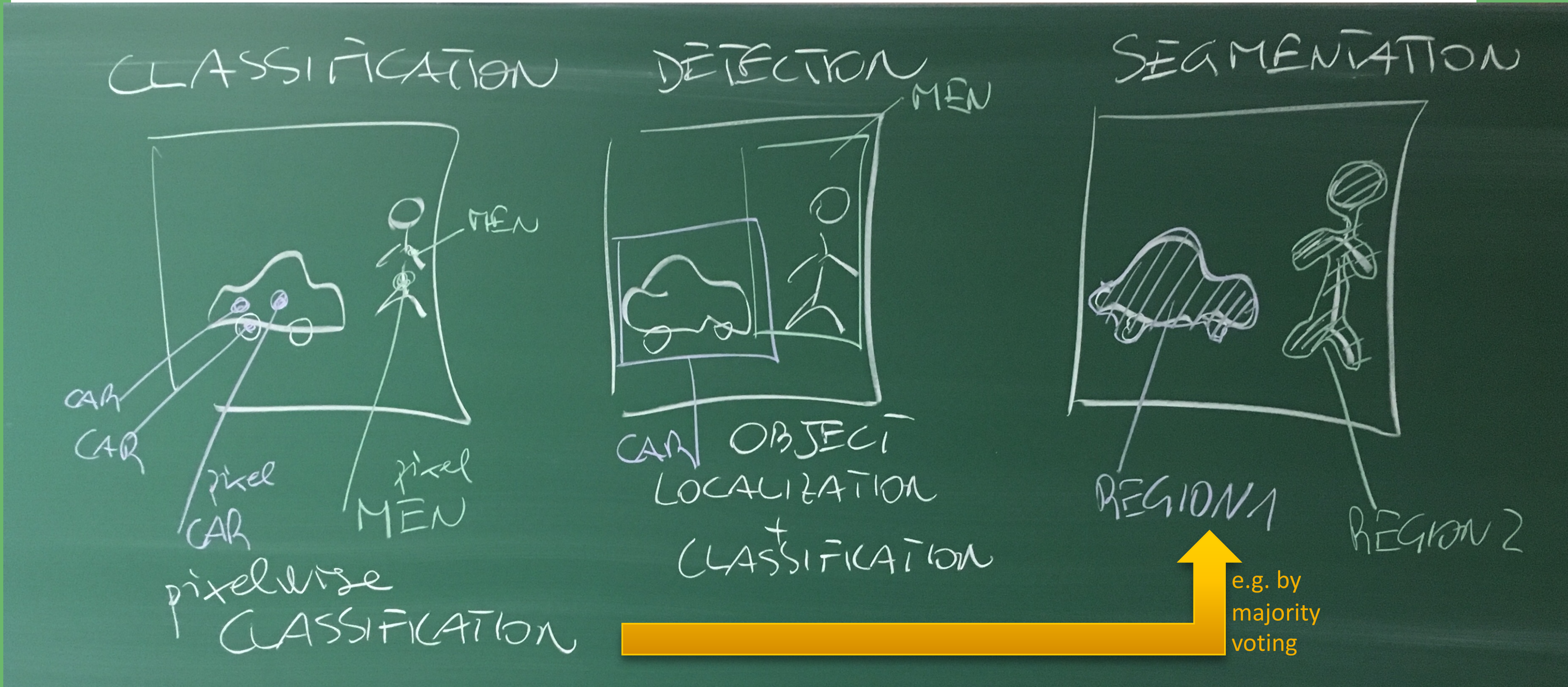
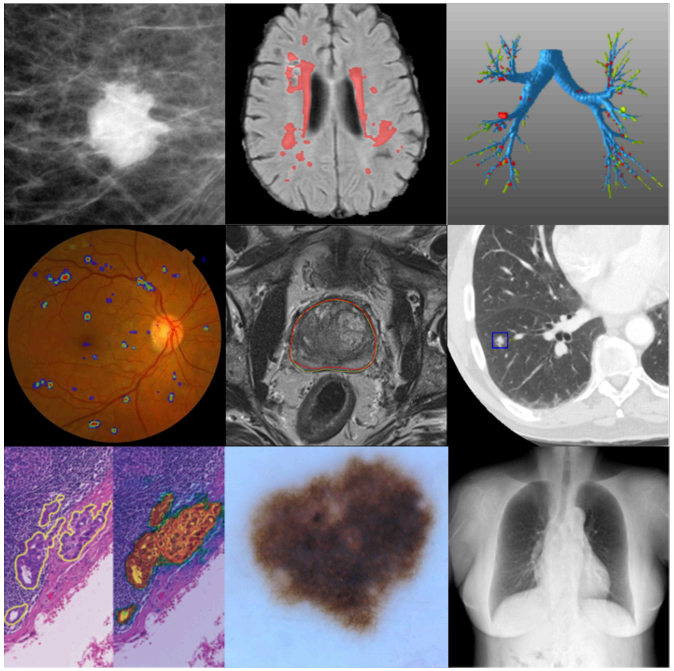
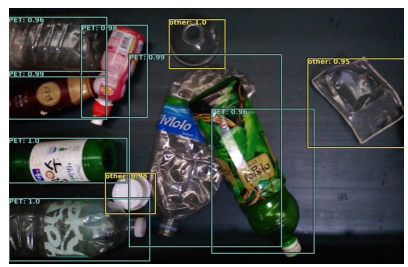


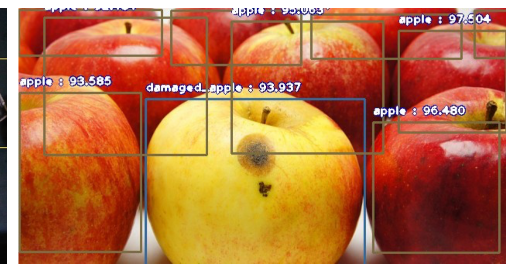
Image interpretation domains: beyond remote sensing

- Biomedical

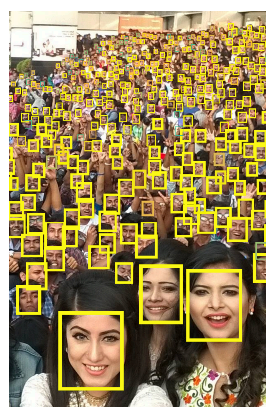


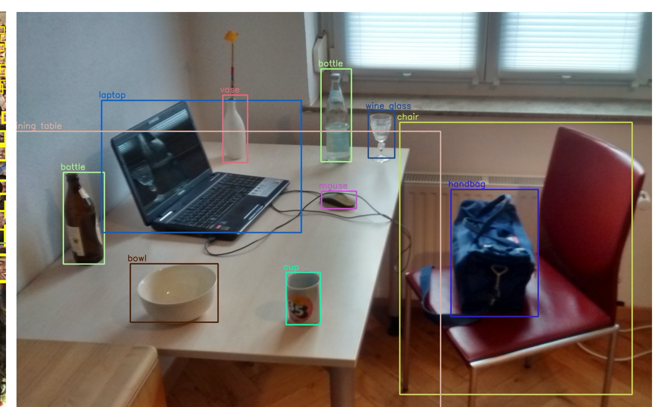
- Manufacturing/recycling/food&beverage





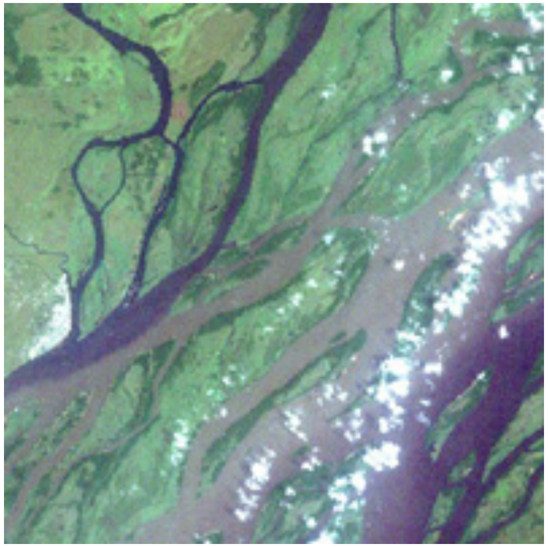
- Indoor/outdoor

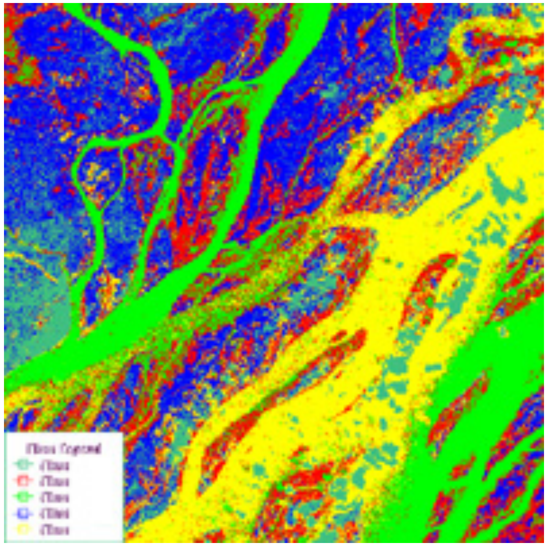




Unsupervised Classification

- **Unsupervised classification** is a mean by which pixels in an image are assigned to spectral classes <u>without the user having foreknowledge of the existence or names of those classes</u>.
- It is performed most often using **clustering methods**.
  - These procedures can be used to determine the number and location of the spectral classes into which the data falls and to determine the spectral class of each pixel.
  - The analyst then identifies those classes afterwards (by visual interpretation or by associating a sample of pixels in each class with available reference data, which could include maps and information from ground visits).
- Clustering procedures are generally *computationally expensive*, yet they are central to help professionals in the analysis of remote sensing imagery.
  - While the information classes for a particular exercise are known, the analyst is usually totally *unaware* of the spectral classes, or sub-classes as they are sometimes called.
- Unsupervised classification is therefore useful for *determining the spectral class composition of the data prior to detailed analysis by the methods of supervised classification.*





Supervised Classification

- A range of **supervised classification** procedures is possible.
  - ***Statistical*** methodologies have been the mainstay of quantitative analysis since the 1970s.
  - Other methods are based on non-statistical, ***geometric*** techniques that seek to place separating surfaces between the classes.
  - We will see both classic statistical and geometric supervised classification in some detail before approaching evolutions that lead to the current ***deep learning*** predominance.

- Here follows a brief introduction to the concepts involved in statistical classification.
  - An important assumption in statistical supervised classification usually adopted in remote sensing is that *each spectral class can be described by a probability (multivariable) distribution in multispectral space*.
    - This is not unreasonable since it would be imagined that most pixels in a distinct cluster or spectral class would lie towards the centre and would decrease in density for positions away from the class centre, thereby *resembling* a probability distribution.
    - The distribution found to be of most value is the normal or Gaussian distribution. It gives rise to *tractable mathematical descriptions of the supervised classification process*, and is *robust* in the sense that *classification accuracy is not overly sensitive to violations of the assumptions* that the classes are normal.
  - A two dimensional multispectral space with the spectral classes so modeled is depicted in the next Figure.
    - The decision boundaries shown in the figure represent those points in multispectral space where a pixel has *equal chance of belonging to two classes*. The boundaries therefore partition the space into regions associated with each class.
    - *This is not new to students with telecommunications background…*

Supervised Classification

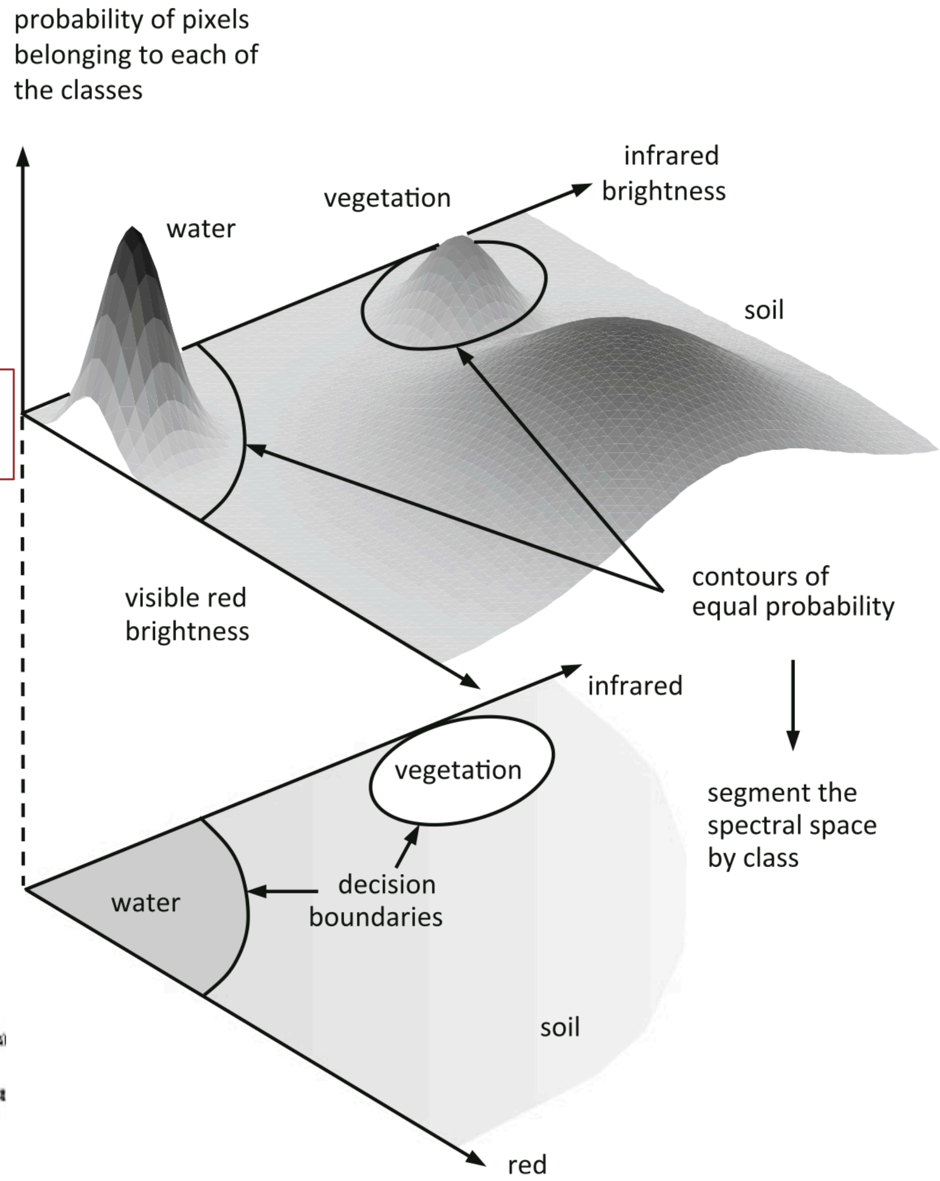
The idea of statistical (parametric) modeling for supervised learning

$$p(\mathbf{x}|\omega_i) \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$$

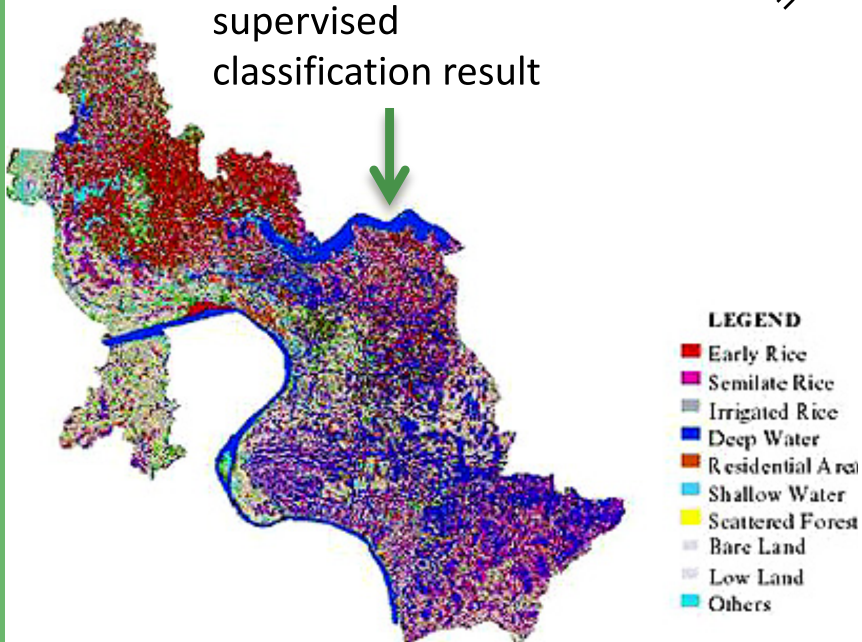
Prob of vector $\mathbf{x}$ given a class ω_i

mean vector

covariance matrix



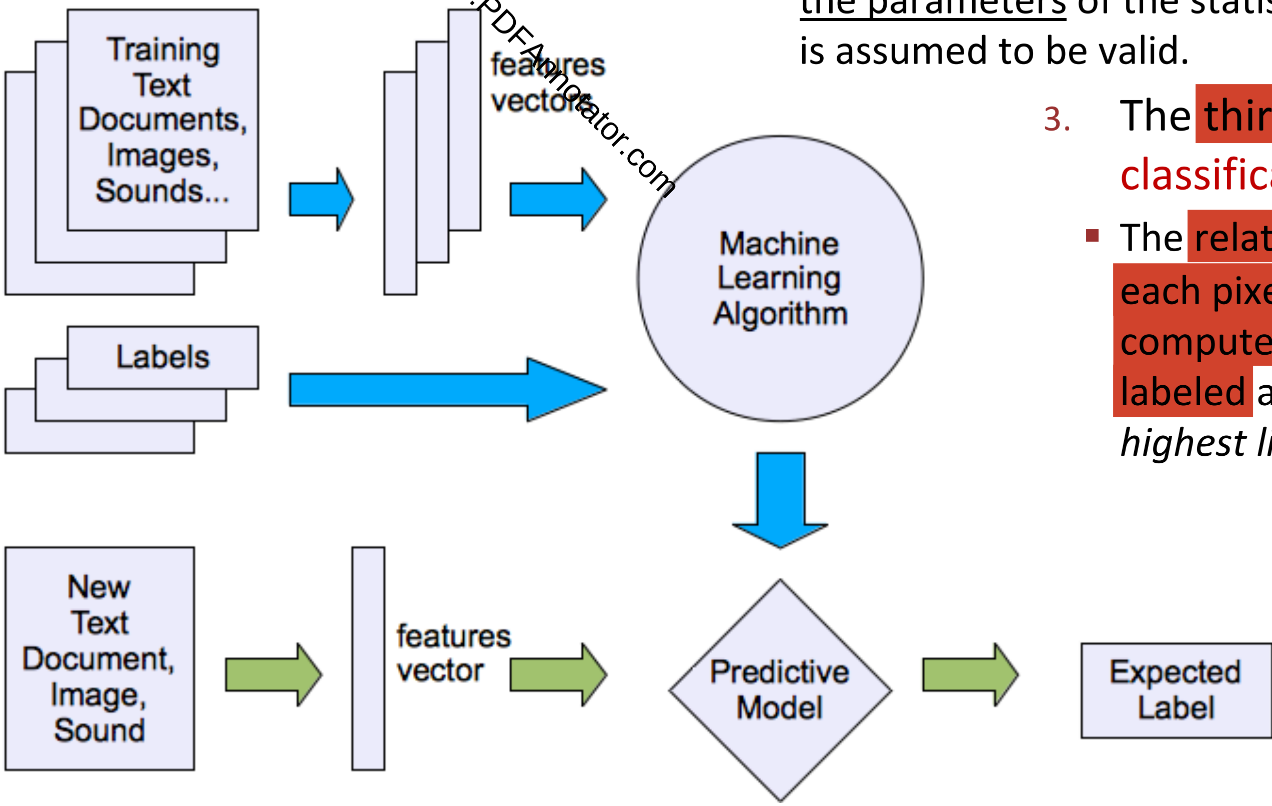
Example of supervised classification result



Supervised Classification

- Supervised classification takes place in three ~~moments~~ Stages:

1. First a set of training pixels is selected/labeled for each spectral class.
   - This may be done using information from ground surveys, aerial photography, topographic maps or any other source of reference data.
2. The second step is the learning phase.
   - In the case of statistical methods, this lead to the parameters of the statistical model which is assumed to be valid.
3. The third step is the classification phase
   - The relative likelihoods for each pixel in the image are computed and the pixel labeled according to the *highest likelihood*.



Supervised Classification

- The view of **statistical supervised classification** we will see first has been based upon an assumption that the classes can be modeled by probability distributions and, as a consequence, are described by the parameters of those distributions.
  - As a result it is also referred to as a **parametric supervised method** (fixed set of parameters).
- Other supervised techniques will be considered, in which neither distribution models nor a predetermined set of parameters to estimate are relevant.
  - These are referred to as **non-parametric methods**.
  - For example ***k-nearest neighbor*** and ***support vector machine*** are often called "non-parametric" classification methods that have been shown to offer good results also in remote sensing applications.
- The clear distinction between parametric and non-parametric methods is debatable
  - see e.g. https://machinelearningmastery.com/parametric-and-nonparametric-machine-learning-algorithms/
  - for neural networks and deep learning methods the question is even more fuzzy

Steps in Supervised Classification

- Supervised classification is the procedure most often used for quantitative analysis of image data.
  - It rests upon using suitable algorithms to label the pixels in an image as representing particular ground cover types, or classes.
  - A variety of algorithms is available for this.
  - Irrespective of the particular method chosen, the essential practical steps usually include:

1. Decide the set of **classes** into which the image is to be segmented.
   - These are the information classes: for example in RS they are *ground cover types* (water, urban regions, croplands, rangelands, etc.) in medical imaging they can be various kinds of healty/patholgical tissues, in manufacturing different materials or surface finishing…
2. Choose representative or prototype pixels from each of the desired set of classes (pixel **annotations**).
   - These pixels are said to form *training data*. In RS training sets for each class can be established using site visits, maps, air photographs or even photointerpretation of a colour composite product formed from the image data.
   - Often the training pixels for a given class will lie in a common region enclosed by a border (a cultivation field, an artefact/anatomical part). That region can be called a *training field*.
3. Use the training data to estimate the "parameters" of the particular classifier algorithm to be used.
   - These parameters will be the properties of the *probability model* used **or** will be equations that define *partitions* in the multispectral space.
   - The set of parameters for a given class is sometimes called the *signature* of that class.

Steps in Supervised Classification (of Remote Sensing Imaging Data)

4. Using the trained classifier, label or classify every pixel in the image into one of the desired (information) classes.
   - Here the whole image segment of interest is typically classified.
   - Whereas training in Step 2 may have required the user to identify perhaps 1% of the image pixels by other means, the computer will label the rest by classification.
5. Produce tabular summaries or class (thematic) maps which summarize the results of the classification.
6. Assess the accuracy of the final product using a labeled testing data set.
   - In practice it might be necessary to decide, on the basis of the results obtained at Step 6, to refine the training process in order to improve classification accuracy.

- It is our objective now to consider the range of algorithms that could be used in 3 and 4. In so doing it will be often assumed (especially for the parametric methods) that the information classes each consists of only one spectral class, so that the two names will be used synonymously.
- We will often use examples from the Remote Sensing domain, but the very same steps can be used in many application domains and domain-specific classes.

Measuring the success and performance for classification

- True Positive **TP** : Correctly identified as “relevant” (typical term from diagnostic contexts)
- True Negative **TN** : Correctly identified as not relevant
- False Positive **FP** : Incorrectly labeled as relevant
- False Negative **FN** : Incorrectly labeled as not relevant

Prediction:
(find cats)

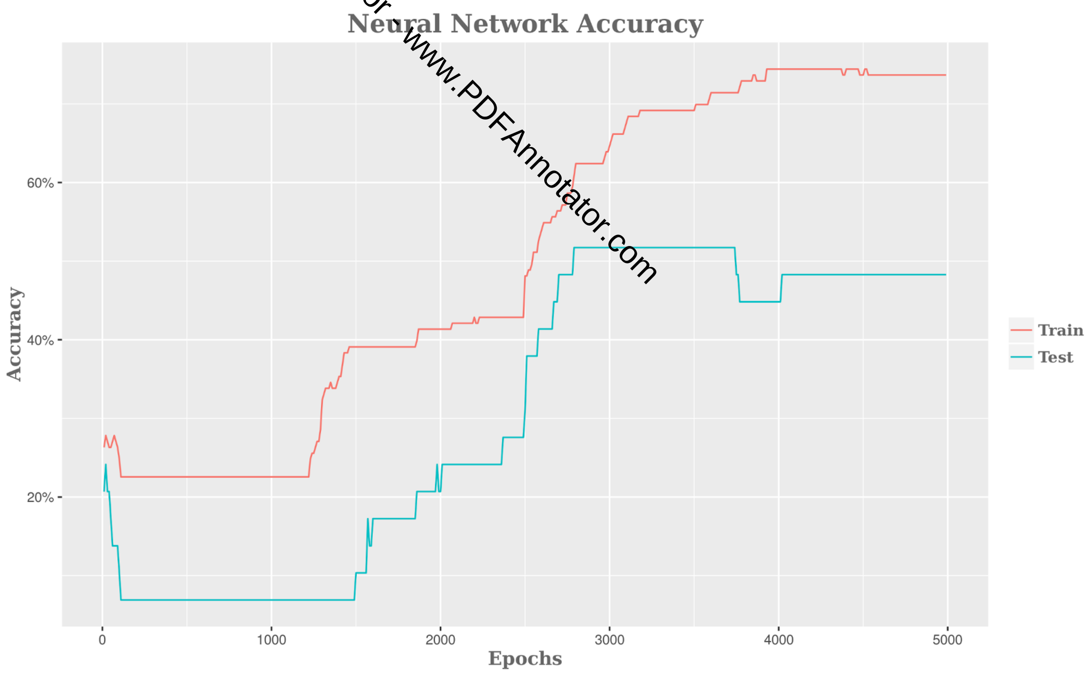
Image dataset:

- **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)

Why Precision and Recall ignore True Negatives?
Typically, there are a lot more negatives than positives (target to find). Counting TN would skew the statistics and favour a system that classifies everything as negatives

Typical issues in Machine Learning: overfitting and underfitting

- **Overfitting**: in case model performs well on training data but poorly on test data
  - Overfitting is a common problem that happens when a model learns the training examples well, but is unable to generalize to new data (e.g. since it learns from irrelevant features)
  - Overfitting can be avoided by using techniques like **regularization** and **cross-validation**

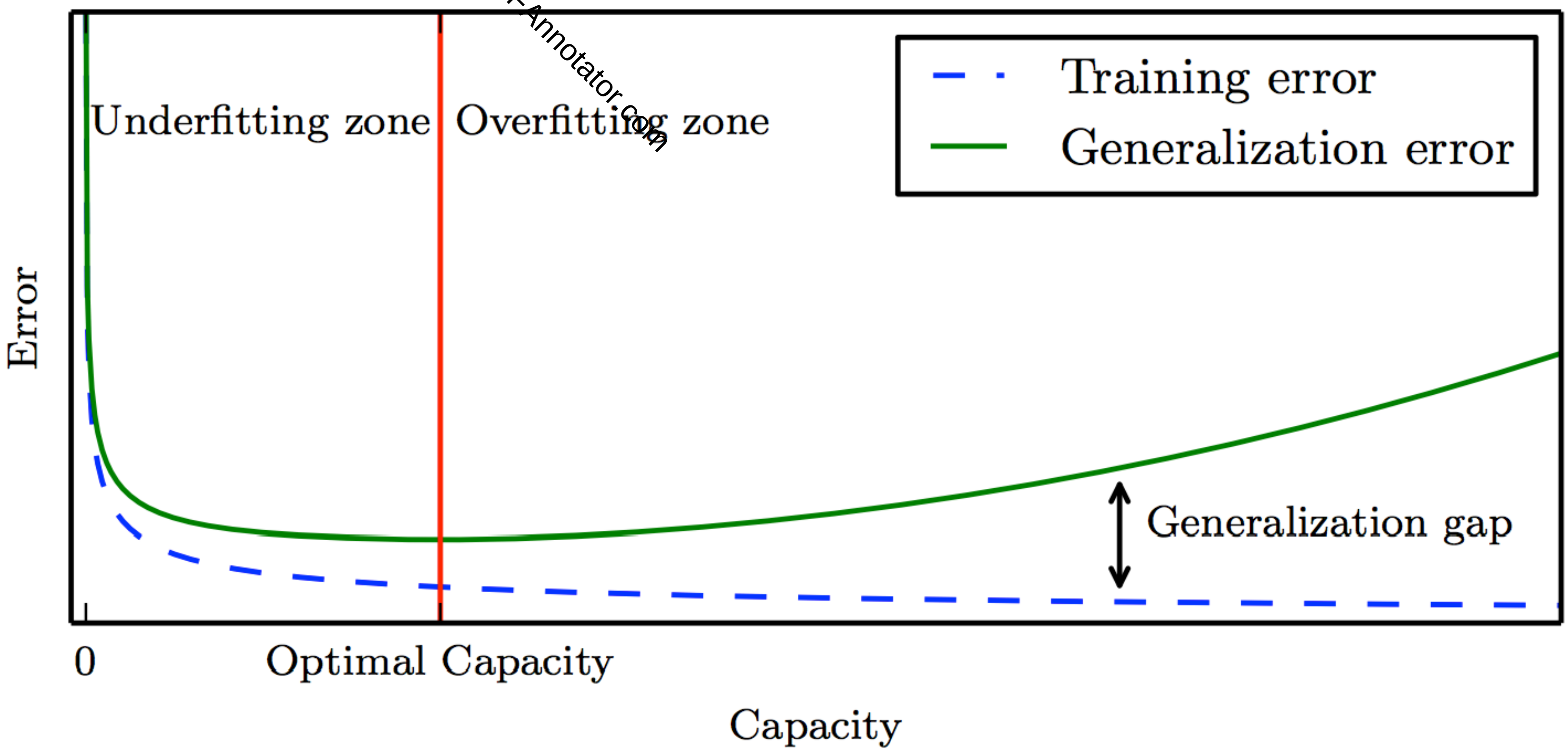


The diagram shows an example of accuracy performance on both training and test set during learning iterations (epochs). Overfitting can be clearly seen.

- **Underfitting**: model that can neither model the training data nor generalize to new data (it can be seen the specular issue with respect to overfitting, however, to solve it one needs to move on and try alternate machine learning techniques)

Typical issues in Machine Learning: generalization and capacity

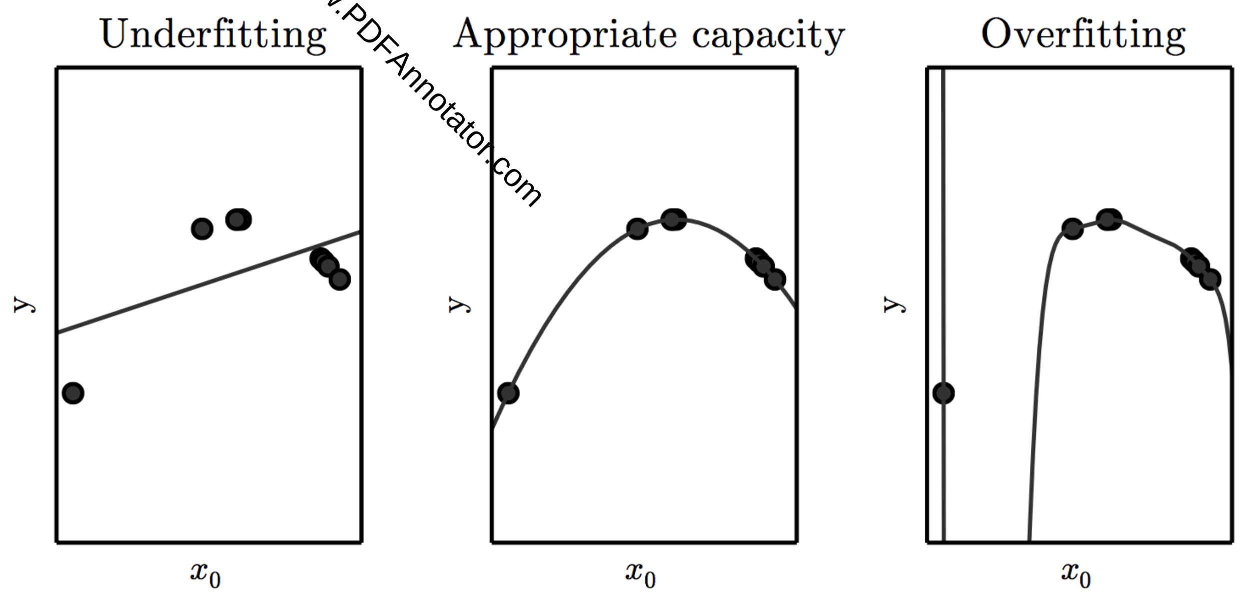
- **Generalization**: the ability to perform well on previously unobserved inputs
  - **training error**: error measure on the training set
  - **test error** or **generalization error**: same error measure on a test set; more formally the expected value of the error on a new input.
    - We want this error to be low as well and this differentiate machine learning from pure optimization or minimization of a loss function typically related to a training set



Typical issues in Machine Learning: model capacity

- **Model capacity**: informally is the ability to fit a wide variety of functions
  - Models with low capacity may struggle to fit the training set
  - Models with high capacity might overfit by memorizing properties of the training set that do not serve them well on the test set

- **Example: Polynomial fitting (capacity = degree of fitting polynomial)**



- More in general capacity control is operated working on the **hypothesis space** characterizing a specific learning algorithm, i.e. the set of functions that the learning algorithm is allowed to select as being the solution.

Typical issues in Machine Learning: bias and variance

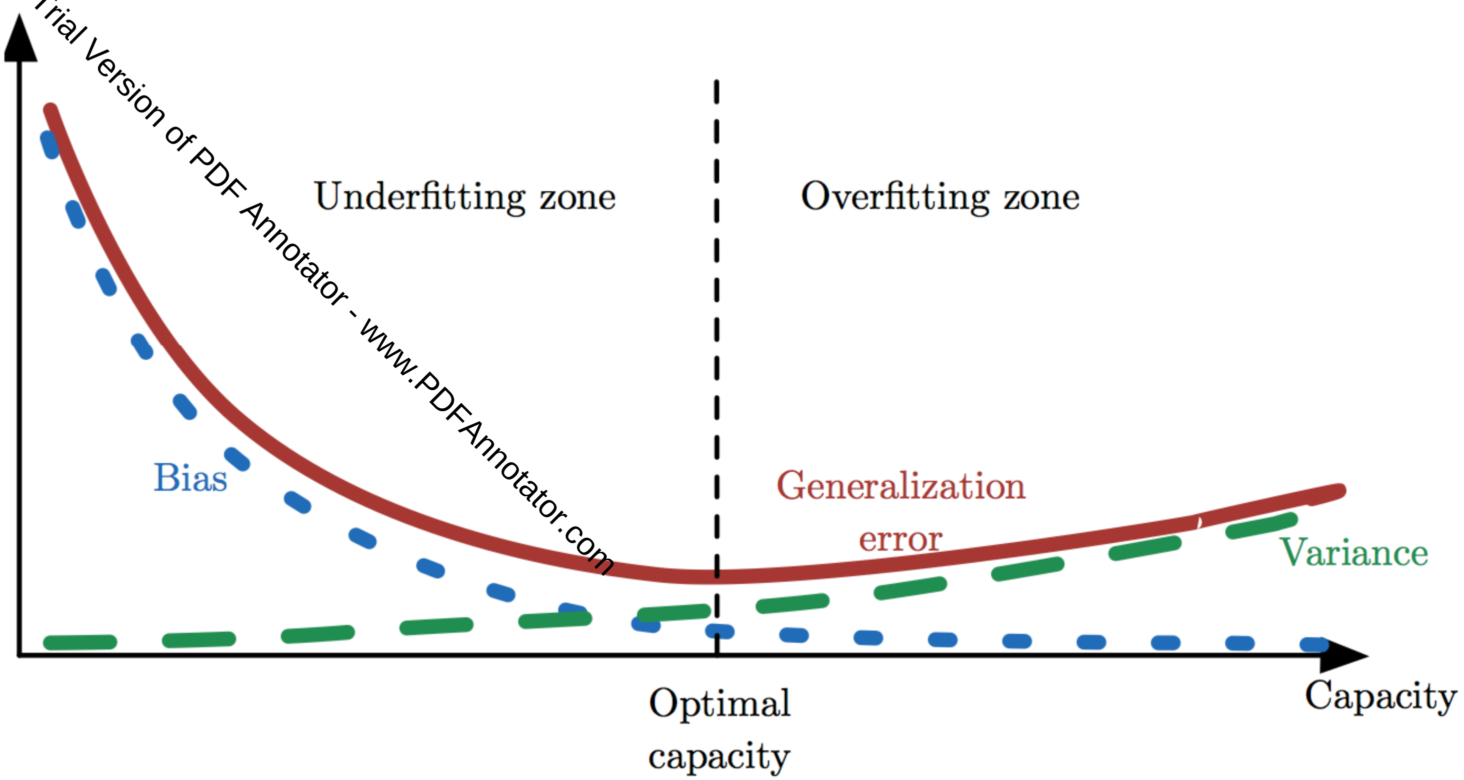
- **Bias**: expected difference between model's prediction and truth
- **Variance**: how much the (expected) model differs among (variations of) training sets

- **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

- Bias and Variance are typically inversely related.
  - An example of a high bias and low variance model is a model that always predicts "cat"
    - It's bias is high because it frequently misses to classify the image, but its variance is low because its prediction results do not depend on the training data used
  - An example of a low bias and high variance model is a model that predicts "cat" if the image matches pixel-to-pixel with a cat in the training data
    - It's bias is low in the training data because it has 100% correct labels, but its variance is high because it's predictions will differ depending on what images are in the training data

- **Cross-validation** is used to find a good balance between bias and variance
  - repeat training and test computation on different randomly chosen subsets of the original one

Typical issues in Machine Learning: overall picture



- Typical activities characterizing machine learning
  - Make the training error small → avoid underfitting
  - Make the gap between training and generalization (test) error small → avoid overfitting
  - Trade-off under- with over-fitting → select optimal capacity (+ Occam's razor principle)
  - Trade-off bias with variance → use cross-validation
  - **Regularization: model selection to reduce generalization error but not training error**