Introduction to Advanced Pattern Recognition

Marco Loog

ASCI A1 : Advanced Pattern Recognition

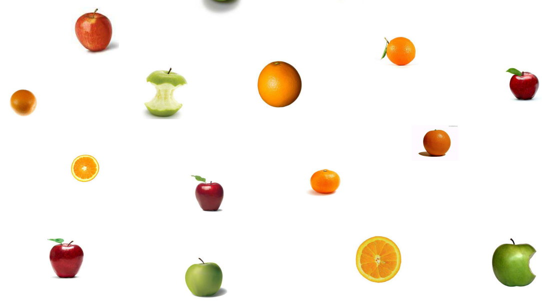
TU Delft
Delft University of Technology

Supervised Learning

- Aims to find solutions to difficult decision, classification, and prediction problems
- Automation often, possibly implicit, an issue
  - Have computers and robots do task that are dangerous, tedious, boring, etc.
- Part of the point : Humans often severely biased in judgment, very inaccurate

TU Delft

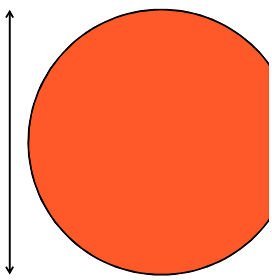
Comparing  and  ...



TU Delft

Supervised Learning = Not Modeling

- Fruits orange, reddish-green to yellowish-green, round, 4-12 cm, consist of a leathery peel 6 mm thick, tightly adherent, protecting the juicy inner pulp, which is divided into segments that may not contain seeds, depending on the cultivar…



TU Delft

Orange Modeling

- Difficult, hassle, overly ambitious, inaccurate,…
- Captures typicalities

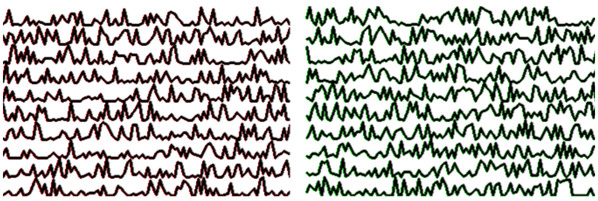
TU Delft

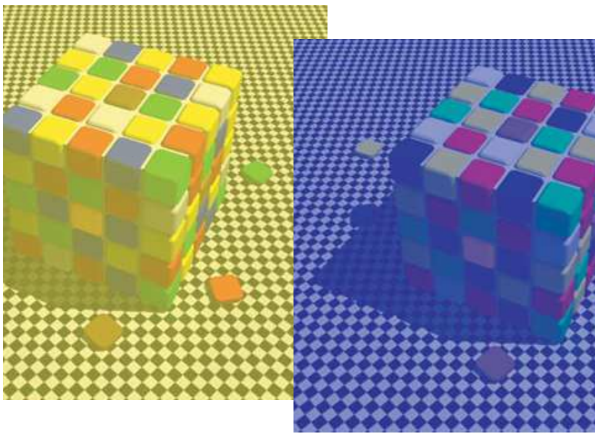
Supervised Learning

- …is learning by example
- Given input and associated output, determine input-output mapping
- Mapping should be able to **generalize** to new and previously unseen examples

Some Remarks

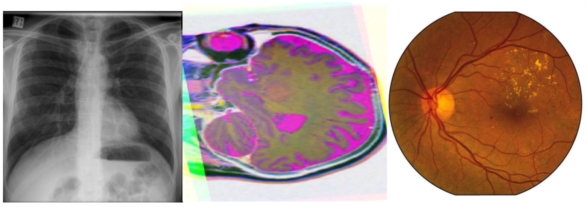
- In the end, one never relies solely on extremes of pure modeling or pure model-free learning
- Learning does use models, though they are weak and nonspecific
- Part of the point : Humans often severely biased in judgment, very inaccurate
  - Cannot even trust your own eyes



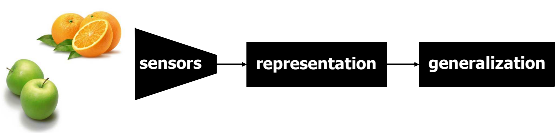


Some Problem Types

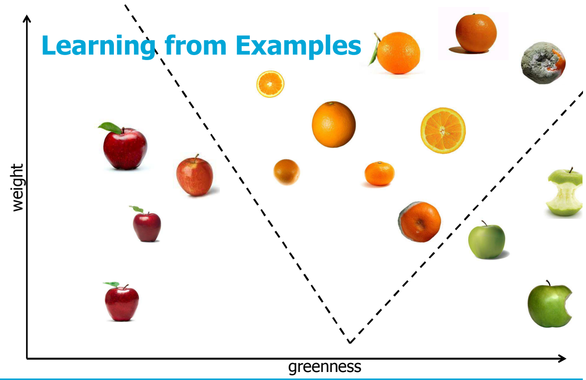
- Classification, segmentation, detection



Basic Pattern Recognition Machine



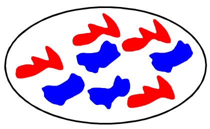
Learning from Examples



On Features

- How do we get to features?
- Features are based on knowledge
  - Good knowledge ⇒ good features ⇒ [almost] separable groups
  - Lack of knowledge ⇒ [too many] bad features ⇒ hardly separable groups
  - Many features ≈ lack of knowledge

Pattern Recognition

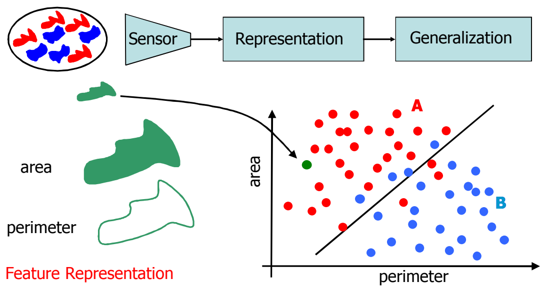


Examples of objects for different classes

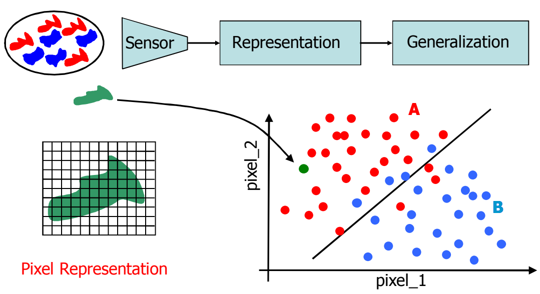
Object of unknown class to be classified

A ? B

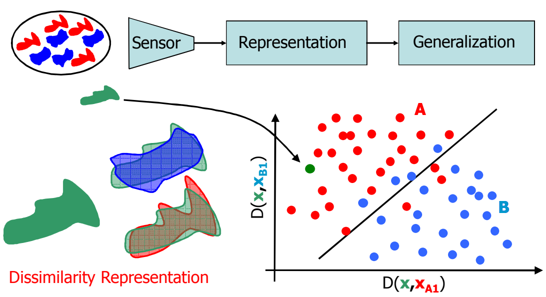
Pattern Recognition System



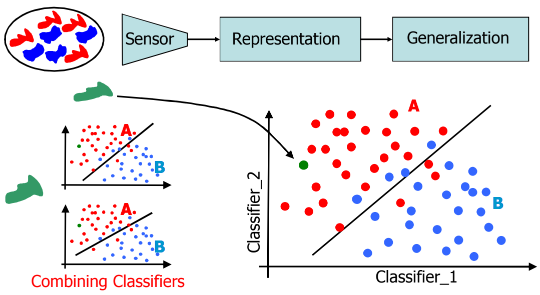
Pattern Recognition System



Pattern Recognition System



Pattern Recognition System

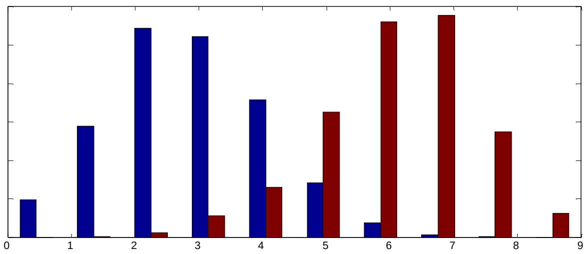


Features Reduce!

- Classes in feature space overlap
- Features only tell you part of the story

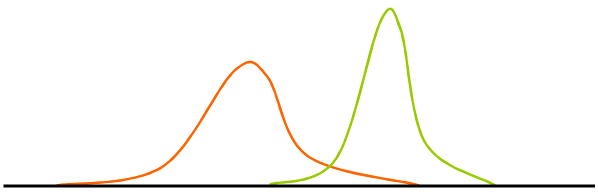
Probabilistic Generalization

- Probabilities good for overlapping classes



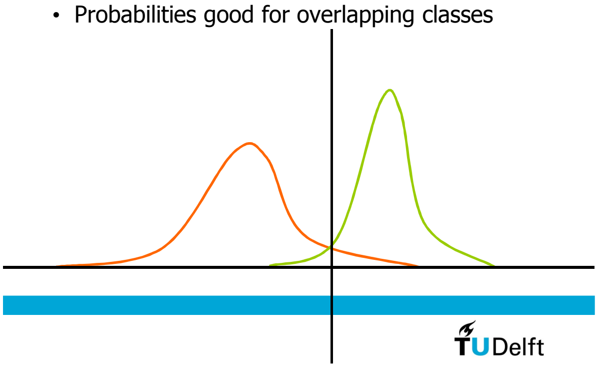
Probabilistic Generalization

- Probabilities good for overlapping classes

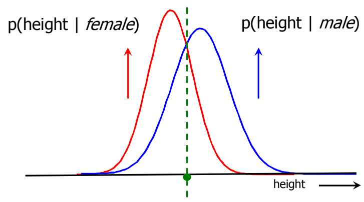


Probabilistic Generalization

- Probabilities good for overlapping classes

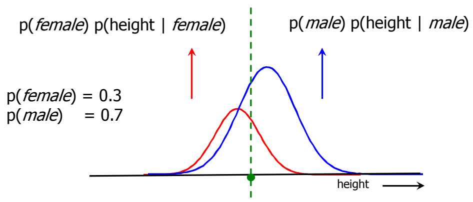


So How to?



Guess what is the gender of somebody with certain height...

Decisions Based on Densities



Guess what is the gender of somebody with certain height...

Bayes Decision Rule

p(*female* | height) > p(*male* | height) → *female* else *male*

Bayes :

$$\frac{p(\text{height} \mid female)\, p(female)}{p(\text{height})} > \frac{p(\text{height} \mid male)\, p(male)}{p(\text{height})}$$

p(height | *female*) p(*female*) > p(height | *male*) p(*male*) → *female* else *male*

pdf estimated from training set

class prior probabilities known, guessed or estimated

Bayes Decision Rule

p(A|x) > p(B|x) → A else B

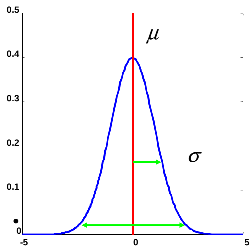
$$\frac{p(x|A)\,p(A)}{p(x)} > \frac{p(x|B)\,p(B)}{p(x)} \rightarrow \text{A else B}$$

p(x|A) p(A) > p(x|B) p(B) → A else B

2-class problems: S(x) = p(x|A) p(A) - p(x|B) p(B) > 0 → A else B

n-class problems: Class(x) = argmax$_\omega$(p(x|ω) p(ω))

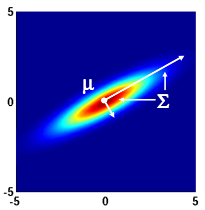
Gaussian Distribution



- Normal distribution = Gaussian distribution
- Standard normal distribution: $\mu = 0$, $\sigma^2 = 1$
- 95% of data between [$\mu - 2\sigma$, $\mu + 2\sigma$] [in 1D!]

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

Multivariate Gaussians

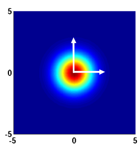


$$\mathbf{\Sigma} = \begin{bmatrix} 3 & 1½ \\ 1½ & 2 \end{bmatrix}$$

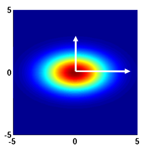
- k - dimensional density :

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k \det(\mathbf{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

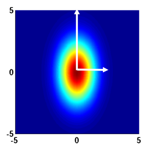
Multivariate Gaussians



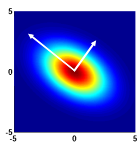
$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mathbf{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$



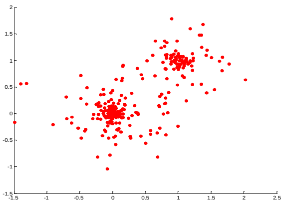
$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$



$$\mathbf{\Sigma} = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}$$

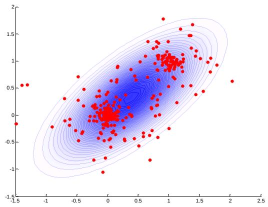
Parametric Estimation

- Assume Gaussian model
- Estimate mean and covariance from data



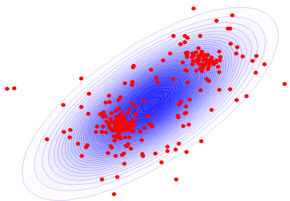
Parametric Estimation

- Assume Gaussian model
- Estimate mean and covariance from data



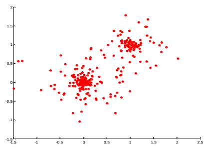
Parametric Drawback

- Poor model fit resulting in invalid conclusions
- Model not flexible enough



Mixture Modelling

- Often used parametric model
- Comes closer to nonparametric formulations
- Basic idea is to model complex pdf as weighted sum of simple pdfs



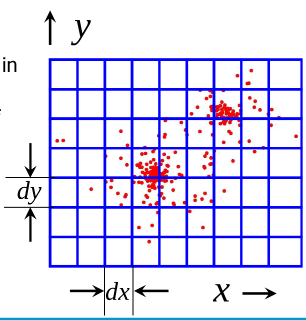
$$p(x) = \sum_{i=1}^{N} P_i p(x|\theta_i)$$

Nonparametric Estimation

Histogram method

1. Divide feature space in N^2 bins
2. Count the number of objects in each bin
3. Normalize :

$$\hat{p}(x) = \frac{n_{ij}}{\sum_{i,j=1}^{N} n_{ij} dx dy}$$



Parzen Density Estimation

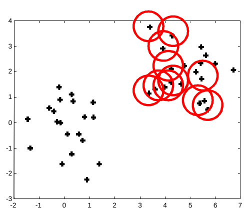
- Fix volume of bin, vary positions of bins, add contribution of each bin
- Define 'bin'-shape [kernel] :

$$K(\mathbf{r}) > 0$$

$$\int K(\mathbf{r})\, d\mathbf{r} = 1$$

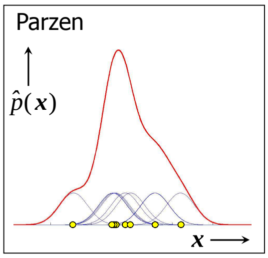
- For test object $\mathbf{z}$ sum all bins

$$p(\mathbf{z}) = \frac{1}{hn} \sum_i K\left(\frac{\mathbf{z}-\mathbf{x}_i}{h}\right)$$



Parzen Density Estimation

- With Gaussian kernel : $K(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{x^2}{2h^2}\right)$



TUDelft

Conclusions for This First Hour

- Learning can be considered the opposite of modeling
- Learning from examples is based on features
- The class overlap causes need for probabilities
- Correct probability densities $\Rightarrow$ optimal classification
- Typically, need to resort to some form of estimation

TUDelft