These are the slides of the lecture

Pattern Recognition

Winter term 2020/21

Friedrich-Alexander University of Erlangen-Nuremberg.

Please feel free to reuse any of the figures and slides, as long as you keep a reference to the source of these slides at `https://lme.tf.fau.de/teaching/` acknowledging the authors Niemann, Hornegger, Hahn, Steidl, Nöth, Seitz, Rodriguez, Das and Maier.

Erlangen, January 8, 2021
Prof. Dr.-Ing. Andreas Maier

Pattern Recognition (PR)

Prof. Dr.-Ing. Andreas Maier
Pattern Recognition Lab (CS 5), Friedrich-Alexander-Universität Erlangen-Nürnberg
Winter Term 2020/21

Introduction

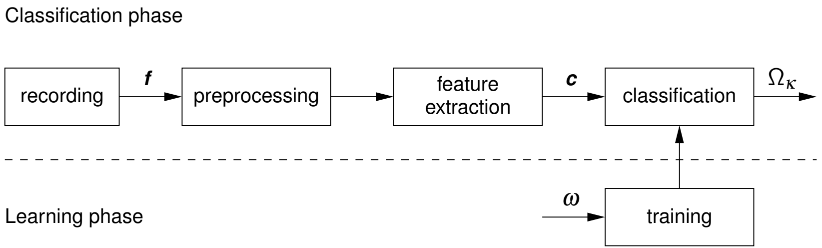
Pattern Recognition in Erlangen

Prof. Dr. Heinrich Niemann

Prof. Dr. Joachim Hornegger

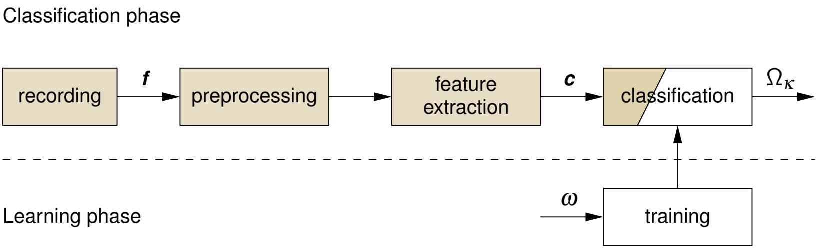
PD Dr. Stefan Steidl

Lecture Topics

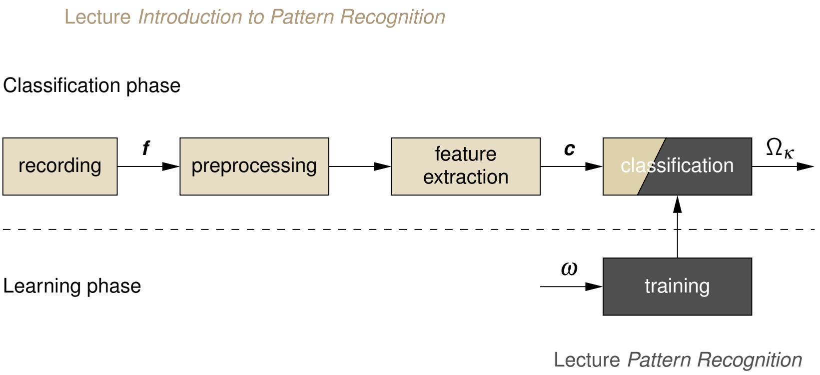


Lecture Topics

Lecture *Introduction to Pattern Recognition*



Lecture Topics



Lecture Topics

- Bayes
- Naïve Bayes
- Logistic Regression
- Discriminant Analysis
- Perceptrons
- Support Vector Machines
- Norms
- Optimization
- Kernel Methods
- Expectation Maximization
- Ada Boost

Pattern Recognition – What For?

- Speech recognition
- Image processing
- Fingerprint identification
- Optical character recognition (OCR)
- Industrial workflows
  - Quality control
  - Sorting
- ...

Example

- “Determining Iris flower according to species using optical sensing”

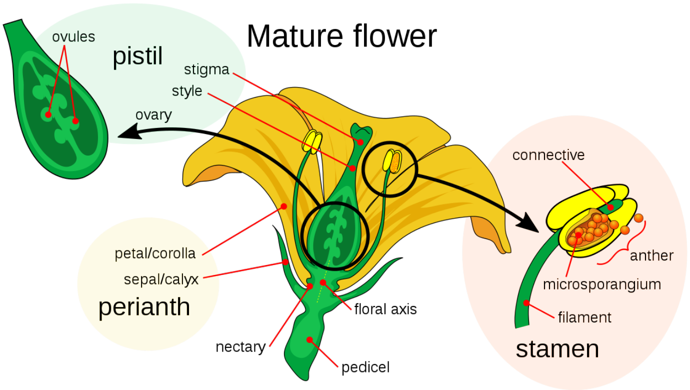
Species → Iris Versicolor

Species → Iris Virginica

1. Fisher 36, "The use of multiple measurements in taxonomic problems".
2. Anderson 36, "The species problem in Iris". Annals of the Missouri Botanical Garden.

- Risk: They may be sold interchangeably which annoys customers!

Example (cont.)



LadyofHats, Public domain, via Wikimedia Commons

Example (cont.)

Problem analysis

- Set up a camera and take some sample images
- Extract characteristics that make distinction between species possible
  - Sepal Length
  - Sepal Width
  - Petal Length
  - Petal Width
  - Color, etc.
- The is the set of all suggested features to explore for use in our classifier!

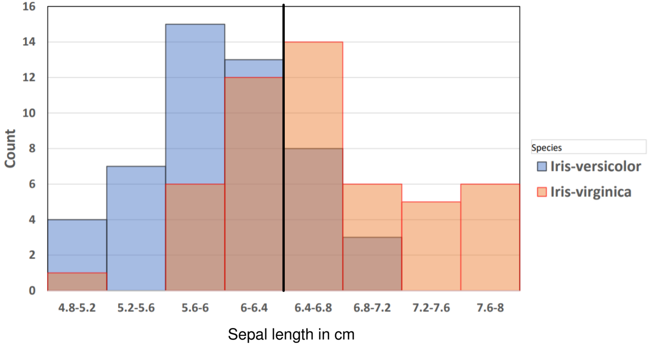
Example (cont.)

- Preprocessing
  - Segmentation operation: isolate flowers from one another and from the background
- Feature extraction: extract best features from single flower image
  - Data reduction
  - Curse of dimensionality
- Classification Classify features with a trained classifier

Example (cont.)

Classification

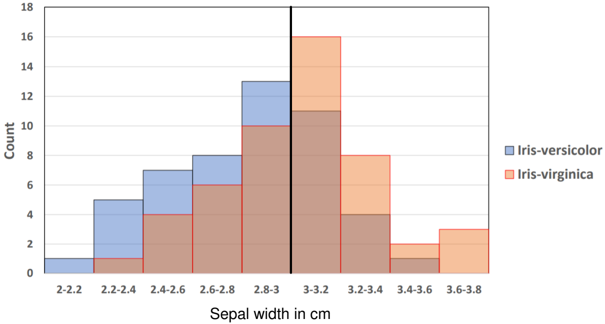
- Possible feature for discrimination: Sepal Length



- The sepal length alone is a poor feature for classification!

Example (cont.)

- Another possible feature: Sepal Width

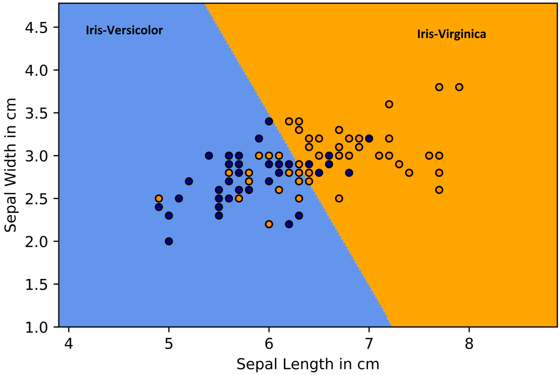


- Relationship between decision boundary and costs!
- What is worse?
  - Red flowers in the blue bucket? $\Rightarrow$ Lower decision boundary!
  - Blue flowers in the red bucket? $\Rightarrow$ Raise decision boundary!
- Optimum decision always depends on definition of cost function

Example (cont.)

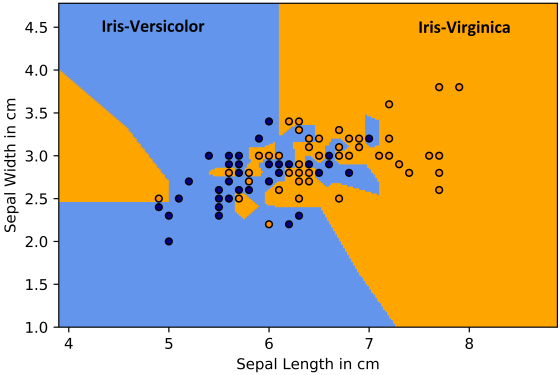
- Adopt Length (x_1) and add the Width (x_2) of the Sepal of the flower

$$\text{Flower} \longrightarrow \boldsymbol{x}^T = [x_1, x_2]$$



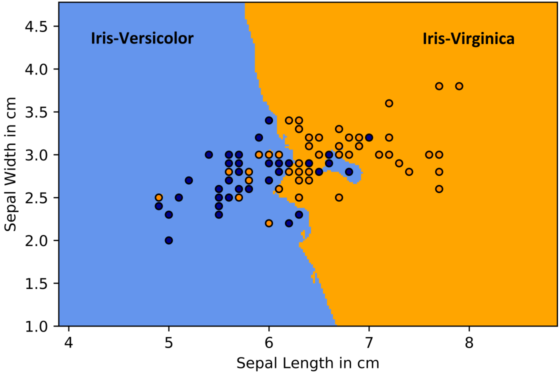
Example (cont.)

- We might add other features that are not correlated with the ones we already have. A precaution should be taken not to reduce the performance by adding such “noisy features”.
- Ideally, the best decision boundary should be the one which provides an optimal performance such as in the following figure:



Example (cont.)

- However, our satisfaction is premature because the central aim of designing a classifier is to correctly classify novel input.
- Issue of generalization!



Feature Extraction

- Recorded input signal
  - Camera, microphone, x-ray signal, etc.

- Digitization: sampling and quantization
- Preprocessing
- Computation of features

Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Next Time in

Pattern Recognition

Postulates for Pattern Recognition

6 Postulates:

1. Availability of a representative sample ω of patterns ${}^{i}\boldsymbol{f}(\boldsymbol{x})$ for the given field of problems Ω

$$\omega = \{{}^{1}\boldsymbol{f}(\boldsymbol{x}), \ldots, {}^{N}\boldsymbol{f}(\boldsymbol{x})\} \subseteq \Omega$$

Postulates for Pattern Recognition

6 Postulates:

1. Availability of a representative sample ω of patterns ${}^{i}\boldsymbol{f}(\boldsymbol{x})$ for the given field of problems Ω

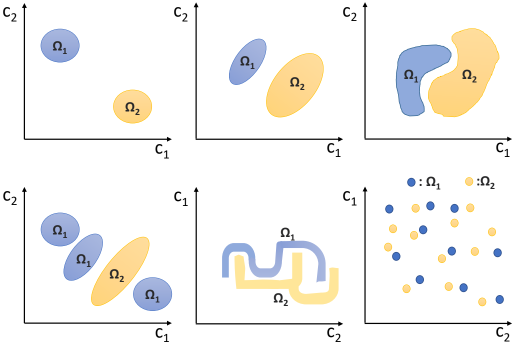
$$\omega = \{{}^{1}\boldsymbol{f}(\boldsymbol{x}), \ldots, {}^{N}\boldsymbol{f}(\boldsymbol{x})\} \subseteq \Omega$$

2. A (simple) pattern has features, which characterize its membership in a certain class Ω_κ.

Postulates for Pattern Recognition (cont.)

3. Compact domain in the feature space of features of the same class; domains of different classes are (reasonably) separable.
    - small intra-class distance
    - high inter-class distance

Example of an increasingly less compact domain in the feature space:



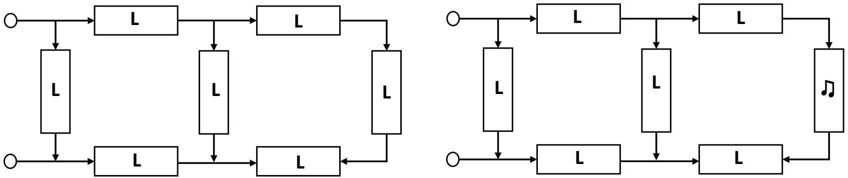
Postulates for Pattern Recognition (cont.)

4. A (complex) pattern consists of simpler constituents, which have certain relations to each other. A pattern may be decomposed into these constituents.

Postulates for Pattern Recognition (cont.)

4. A (complex) pattern consists of simpler constituents, which have certain relations to each other. A pattern may be decomposed into these constituents.

5. A (complex) pattern $\boldsymbol{f}(\boldsymbol{x}) \in \Omega$ has a certain structure. Not any arrangement of simple constituents is a valid pattern. Many patterns may be represented with relatively few constituents.



Postulates for Pattern Recognition (cont.)

6. Two patterns are similar if their features or simpler constituents differ only lightly.

Performance Evaluation

- Feature vectors are used as input for the classifier
- Classifaction results in a discrete class index
- Confusion matrix:

		hypothesis					
		Ω_1	Ω_2	Ω_3	$\dots$	Ω_K	$\sum$
reference	Ω_1	n_{11}	n_{12}	n_{13}	$\dots$	n_{1K}	N_1
	Ω_2	n_{21}	n_{22}	n_{23}	$\dots$	n_{2K}	N_2
	Ω_3	n_{31}	n_{32}	n_{33}	$\dots$	n_{3K}	N_3
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	Ω_K	n_{K1}	n_{K2}	n_{K3}	$\dots$	n_{KK}	N_K
	$\sum$						N

Tab.: Confusion matrix with absolute frequencies for a K-class problem

Performance Evaluation (cont.)

Evaluation of classifiers

- Accuracy / Recognition Rate

$$\mathrm{RR} := \frac{1}{N} \sum_{\kappa=1}^{K} n_{\kappa\kappa} \cdot 100\,\%$$

Performance Evaluation (cont.)

Evaluation of classifiers

- Accuracy / Recognition Rate

$$\mathrm{RR} := \frac{1}{N} \sum_{\kappa=1}^{K} n_{\kappa\kappa} \cdot 100\,\%$$

- Recall and Precision

$$\begin{aligned} \text{recall}_{\kappa} &= \frac{n_{\kappa\kappa}}{\sum_{i=1}^{K} n_{\kappa i}} = \frac{n_{\kappa\kappa}}{N_{\kappa}} \\ \text{precision}_{\kappa} &= \frac{n_{\kappa\kappa}}{\sum_{i=1}^{K} n_{i\kappa}} \end{aligned}$$

Performance Evaluation (cont.)

Evaluation of classifiers

- Accuracy / Recognition Rate

$$\mathrm{RR} := \frac{1}{N} \sum_{\kappa=1}^{K} n_{\kappa\kappa} \cdot 100\,\%$$

- Recall and Precision

$$\begin{aligned}\mathrm{recall}_{\kappa} &= \frac{n_{\kappa\kappa}}{\sum_{i=1}^{K} n_{\kappa i}} = \frac{n_{\kappa\kappa}}{N_{\kappa}}\\ \mathrm{precision}_{\kappa} &= \frac{n_{\kappa\kappa}}{\sum_{i=1}^{K} n_{i\kappa}}\end{aligned}$$

- (Unweighted) Average Recall

$$\mathrm{UAR} := \frac{1}{K} \sum_{\kappa=1}^{K} \frac{n_{\kappa\kappa}}{N_{\kappa}} \cdot 100\,\%$$

Performance Evaluation (cont.)

Special case: only 2 classes

		Reference	
		Positive	Negative
Hyp.	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Various performance measures:

- True positive rate (hit rate, recall, sensitivity): $\frac{\#TP}{\#TP + \#FN}$
- False positive rate (false alarm rate, fall-out): $\frac{\#FP}{\#FP + \#TN}$
- Positive predictive value (precision): $\frac{\#TP}{\#TP + \#FP}$
- Negative predictive value: $\frac{\#TN}{\#TN + \#FN}$
- True negative rate (specifity): $\frac{\#TN}{\#FP + \#TN} = 1 -$ false positive rate

Performance Evaluation (cont.)

More performance measures:

- Accuracy:

$$\mathrm{ACC} = \frac{\#\mathrm{TP} + \#\mathrm{TN}}{\#\mathrm{TP} + \#\mathrm{FP} + \#\mathrm{FN} + \#\mathrm{TN}}$$

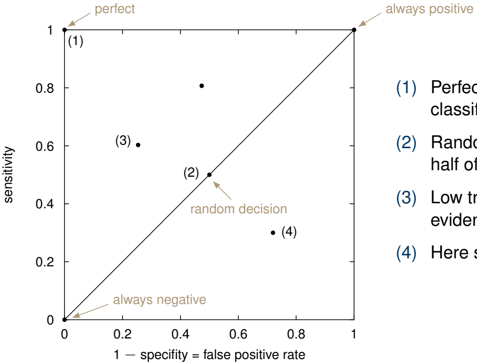
- F-measure: harmonic mean of recall and precision:

$$F = \frac{2 \cdot \mathrm{recall} \cdot \mathrm{precision}}{\mathrm{recall} + \mathrm{precision}}$$

Performance Evaluation (cont.)

Receiver Operating Characteristic (ROC) Curves

- A classifier defines a single 2-d point in the ROC graph.



(1) Perfect classifier: no false positives, all negatives are classified as negatives

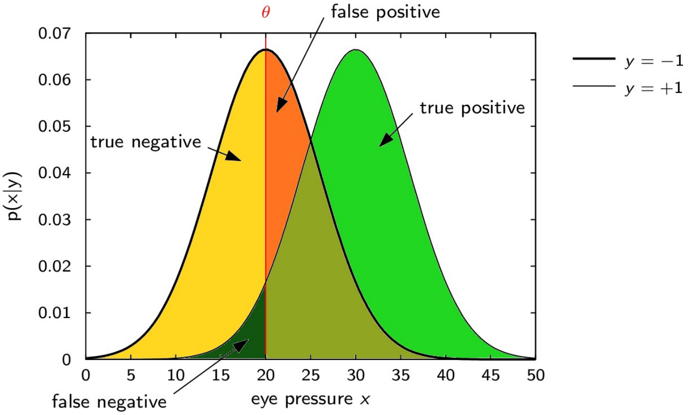
(2) Random decision: half of positives are classified correctly, half of negatives are classified correctly

(3) Low true positive rate, but lower false positive rate; strong evidence for positive classification

(4) Here something goes really wrong!

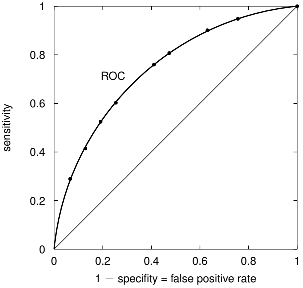
Performance Evaluation (cont.)

Example: Classification of glaucoma based on eye pressure



Performance Evaluation (cont.)

- Classification is a threshold decision (in general $\theta = 0.5$)
- The true positive rate and the false positive rate can be computed for different thresholds $\theta \in [0, 40]$
- Higher true positive rates for lower thresholds, but then higher false positive rates, too!
- The result is the ROC curve:



Performance Evaluation (cont.)

- Performance measure: area under curve (AUC)
- The true positive rate and the false positive rate do not depend on the total number of samples.
- Hence, the ROC curve is independent of the priors of both classes (as opposed to recall-precision-curves).

Types of Classifiers

Different types of classifiers:

- Statistical
- Parametric
- Nonparametric
- Linear
- Nonlinear
- etc.

Learning Phase

- Classifier is only as good as the training samples
- The more training samples the better!

- Distinction between supervised and unsupervised learning

- Computational complexity of the classifier may depend on the size of the training set

Pattern Recognition Lab

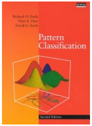
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Next Time in

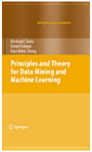
Pattern Recognition

Literature

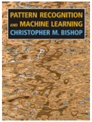
- Richard O. Duda, Peter E. Hart, David G. Stock:
  Pattern Classification, 2nd edition,
  John Wiley & Sons, New York, 2001, EUR 114.45



- Trevor Hastie, Robert Tibshirani, Jerome Friedman:
  The Elements of Statistical Learning –
  Data Mining, Inference, and Prediction,
  2nd edition, Springer, New York, 2009, EUR 71.05
  `http://www-stat.stanford.edu/~tibs/ElemStatLearn/`



- Christopher M. Bishop:
  Pattern Recognition and Machine Learning,
  Springer, New York, 2006, EUR 75.05



Further Readings

- Richard O. Duda, Peter E. Hart, David G. Stock:
  Pattern Classification, 2nd edition,
  John Wiley & Sons, New York, 2001, EUR 114.45

- H. Niemann:
  Klassifikation von Mustern
  2. überarbeitete Auflage, 2003
  `http://www5.informatik.uni-erlangen.de/fileadmin/Persons/NiemannHeinrich/klassifikation-von-mustern/m00links.html`

Comprehensive Questions

- What are features vectors?

Comprehensive Questions

- What are features vectors?
- What are the postulates of pattern recognition?

Comprehensive Questions

- What are features vectors?
- What are the postulates of pattern recognition?
- Why do we need training samples?

Comprehensive Questions

- What are features vectors?
- What are the postulates of pattern recognition?
- Why do we need training samples?
- How can we test the performance of our classifier?