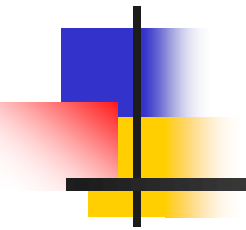
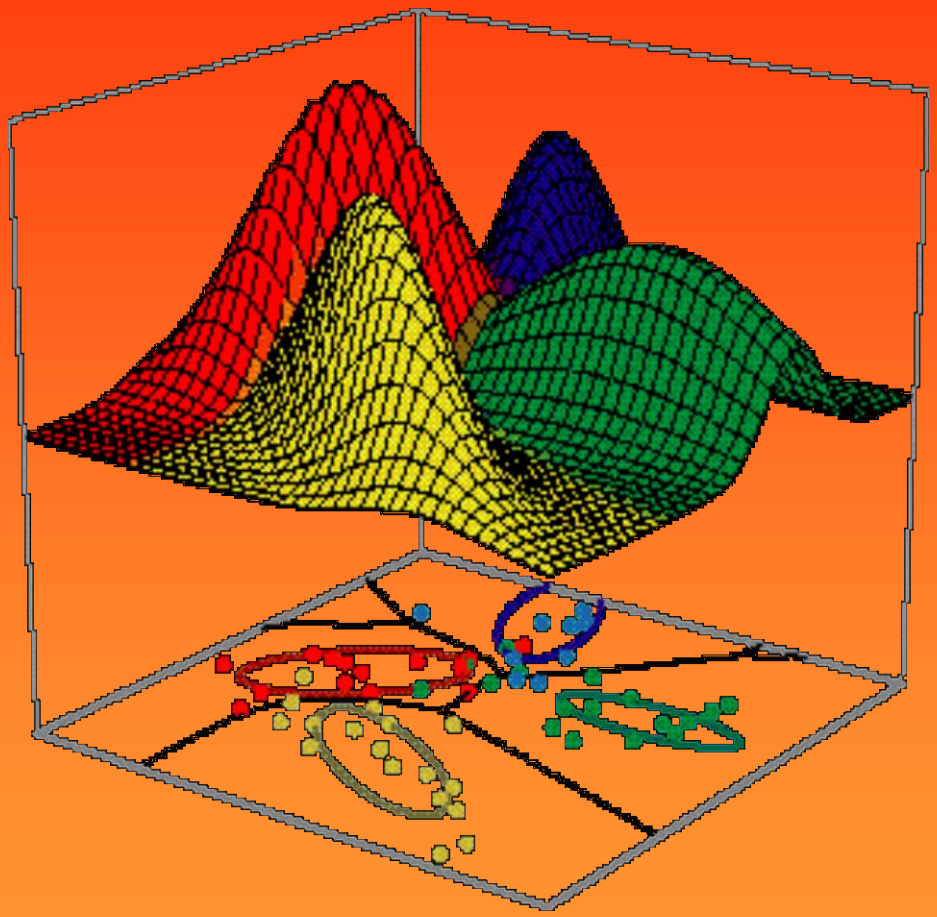


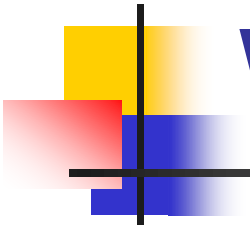
Pattern Recognition Basics



Pattern Classification

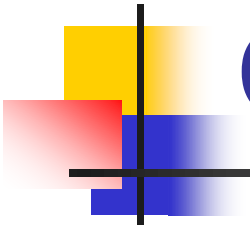
Most of the material in these slides are taken from the figures in

Pattern Classification (2nd ed) by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2001



What is pattern recognition?

- Definition from Duda, et al.
  Pattern recognition is the act of taking in raw data and taking an action based on the "category" of the pattern.
- A **pattern** is an object, process or event that can be given a name.
- A **pattern class** (or category) is a set of patterns sharing common attributes and usually originating from the same source.

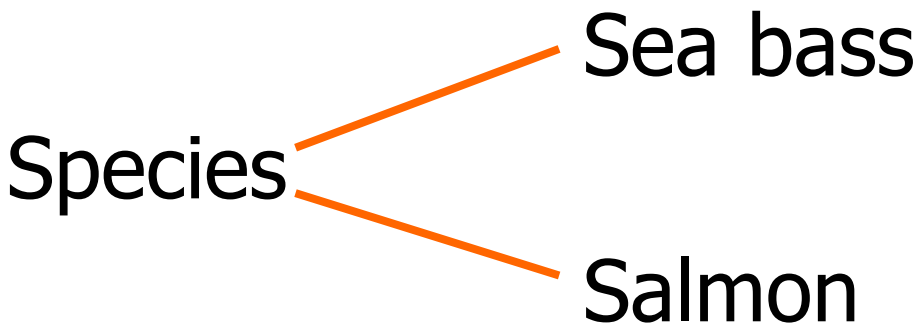


Contd...

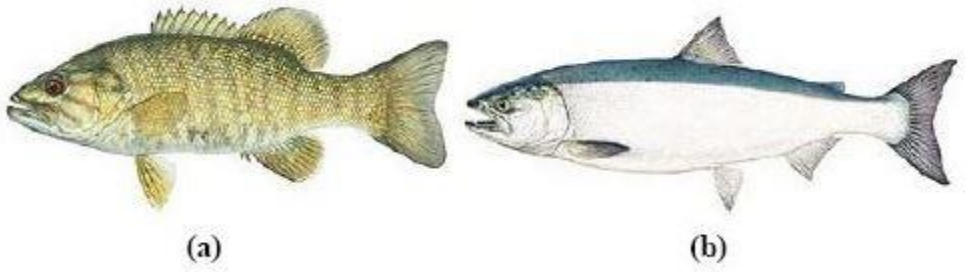
- During **recognition** (or **classification**) given objects are assigned to prescribed classes.
- A **classifier** is a machine which performs classification.
- We gain an understanding and appreciation for pattern recognition in the real world – visual scenes, noises, etc.
  - Human senses: sight, hearing, taste, smell, touch

An Introductory Example

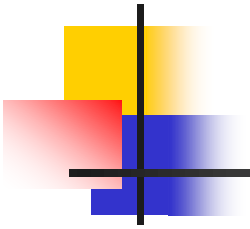
- "Sorting incoming Fish on a conveyor according to species using optical sensing"



An Introductory Example

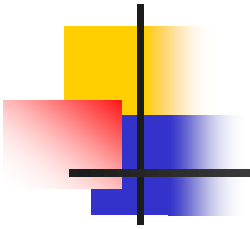


The objects to be classified: a) Salmon b) Sea bass



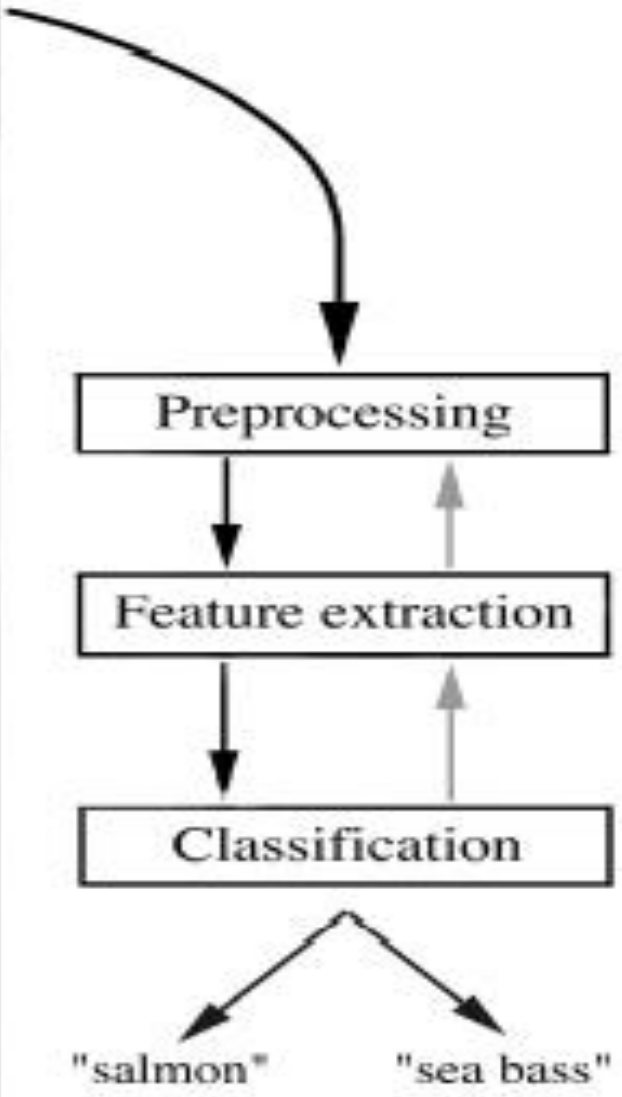
Problem Analysis

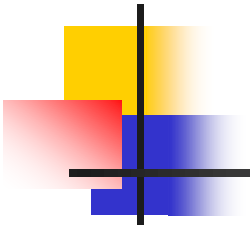
- Set up a camera and take some sample images to extract features
  - Length
  - Lightness
  - Width
  - Number and shape of fins
  - Position of the mouth, etc…



Pattern Classification System

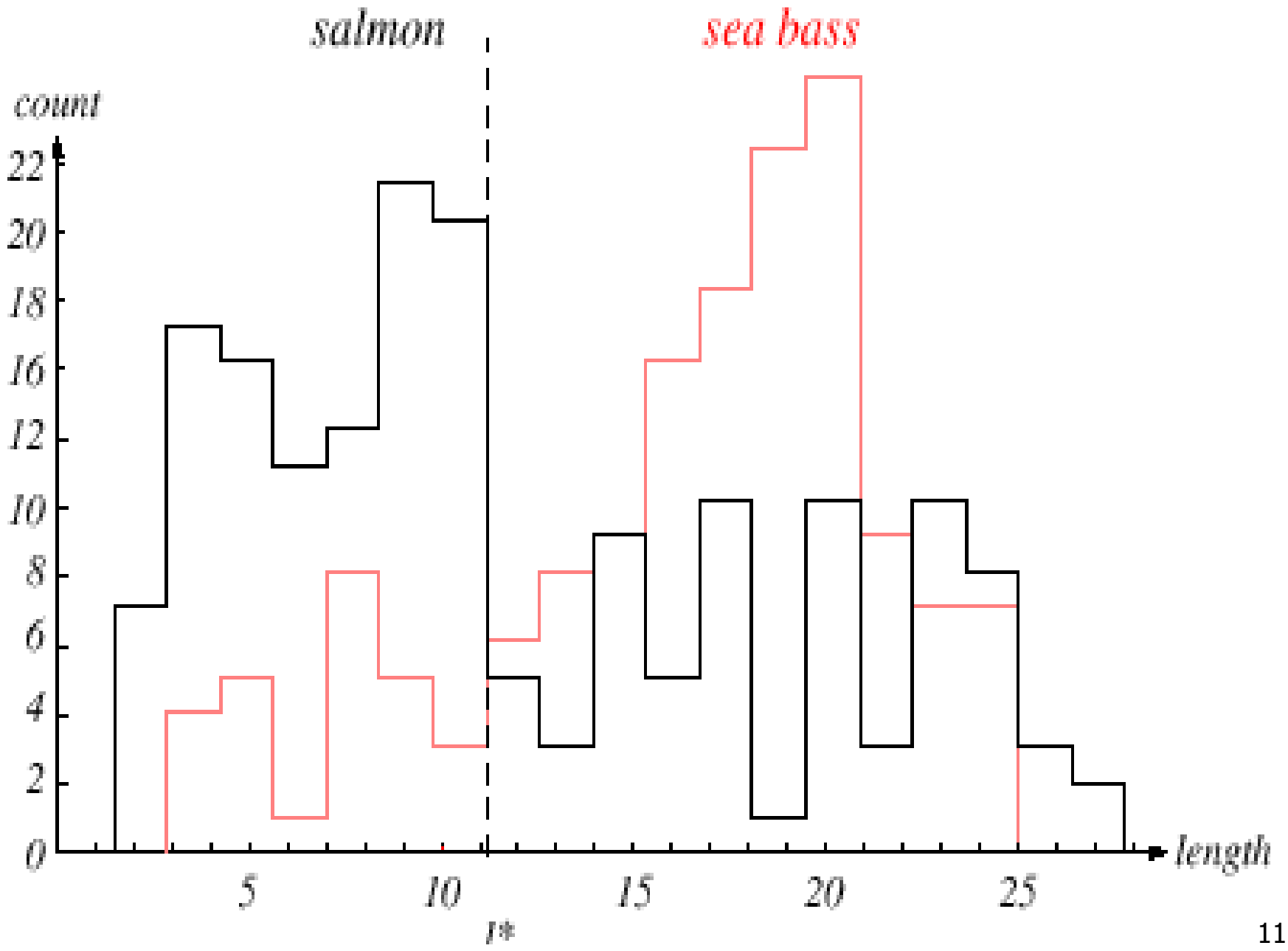
- Preprocessing
  - Segment (isolate) fishes from one another and from the background
- Feature Extraction
  - Reduce the data by measuring certain features
- Classification
  - Divide the feature space into decision regions

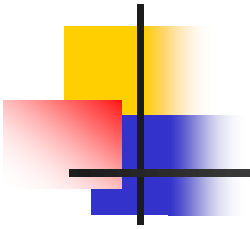

Preprocessing
Feature extraction
Classification
"salmon"
"sea bass"



Classification

- Initially use the length of the fish as a possible feature for discrimination

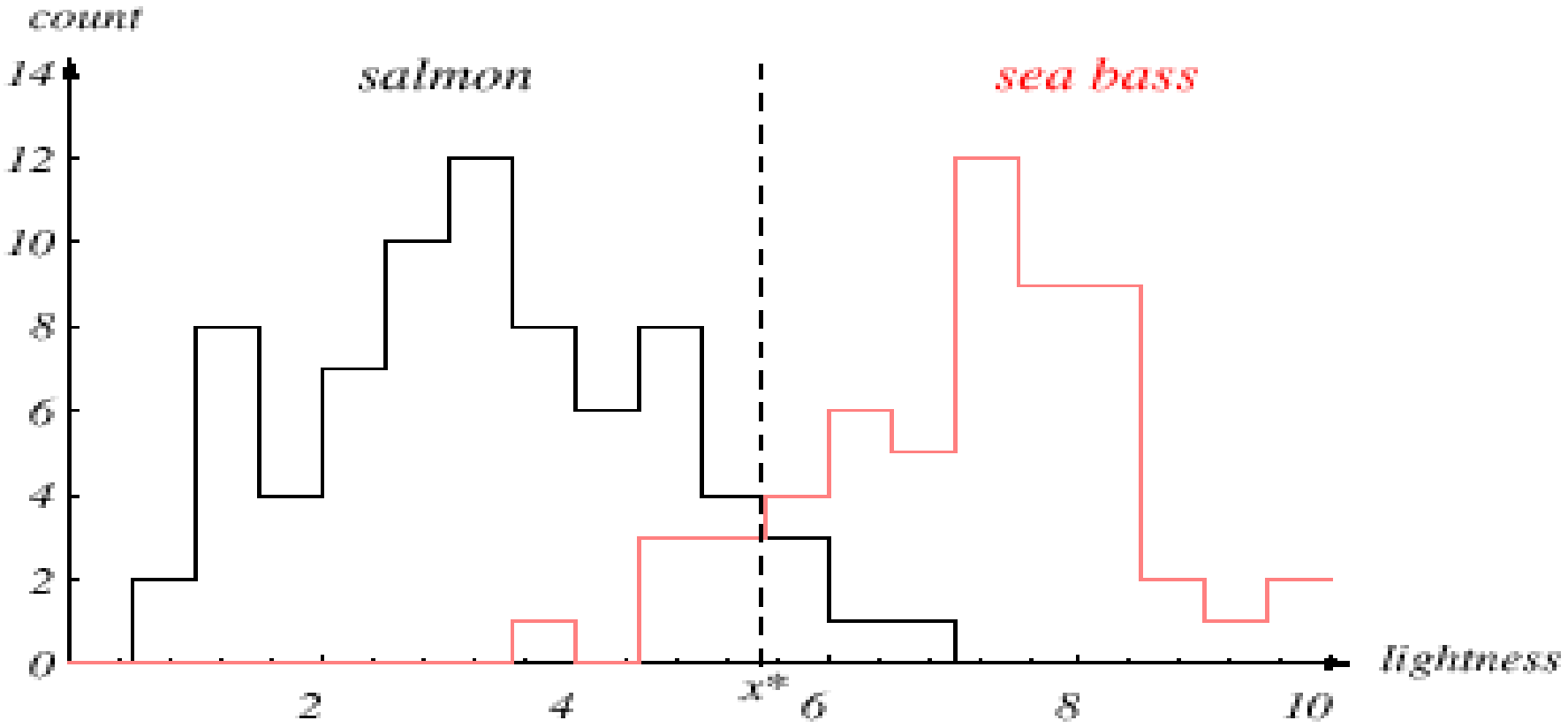

salmon
sea bass
count
22
20
18
16
12
10
8
6
4
2
0
5
10
15
20
25
length
l*



Feature Selection

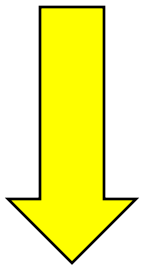
The length is a poor feature alone!

Select the lightness as a possible feature


count
salmon
sea bass
14
12
10
8
6
4
2
0
2
4
x*
6
8
10
lightness

Threshold decision boundary and cost relationship

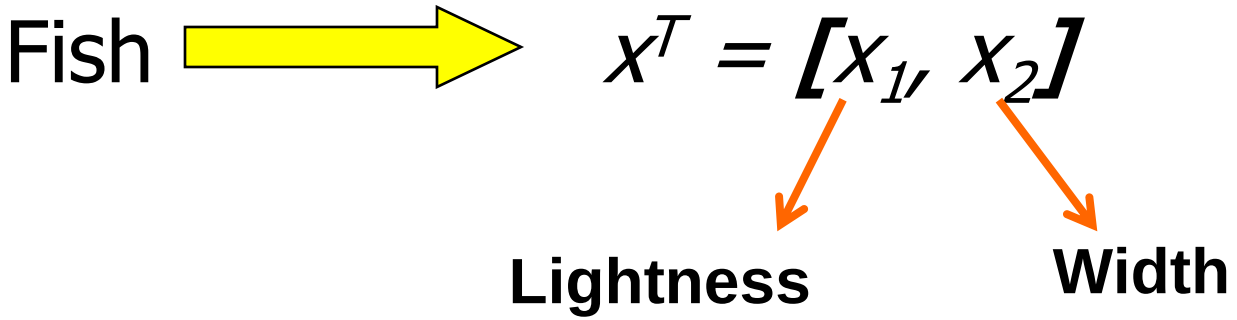
- Move decision boundary toward smaller values of lightness in order to minimize the cost (reduce the number of sea bass that are classified salmon!)



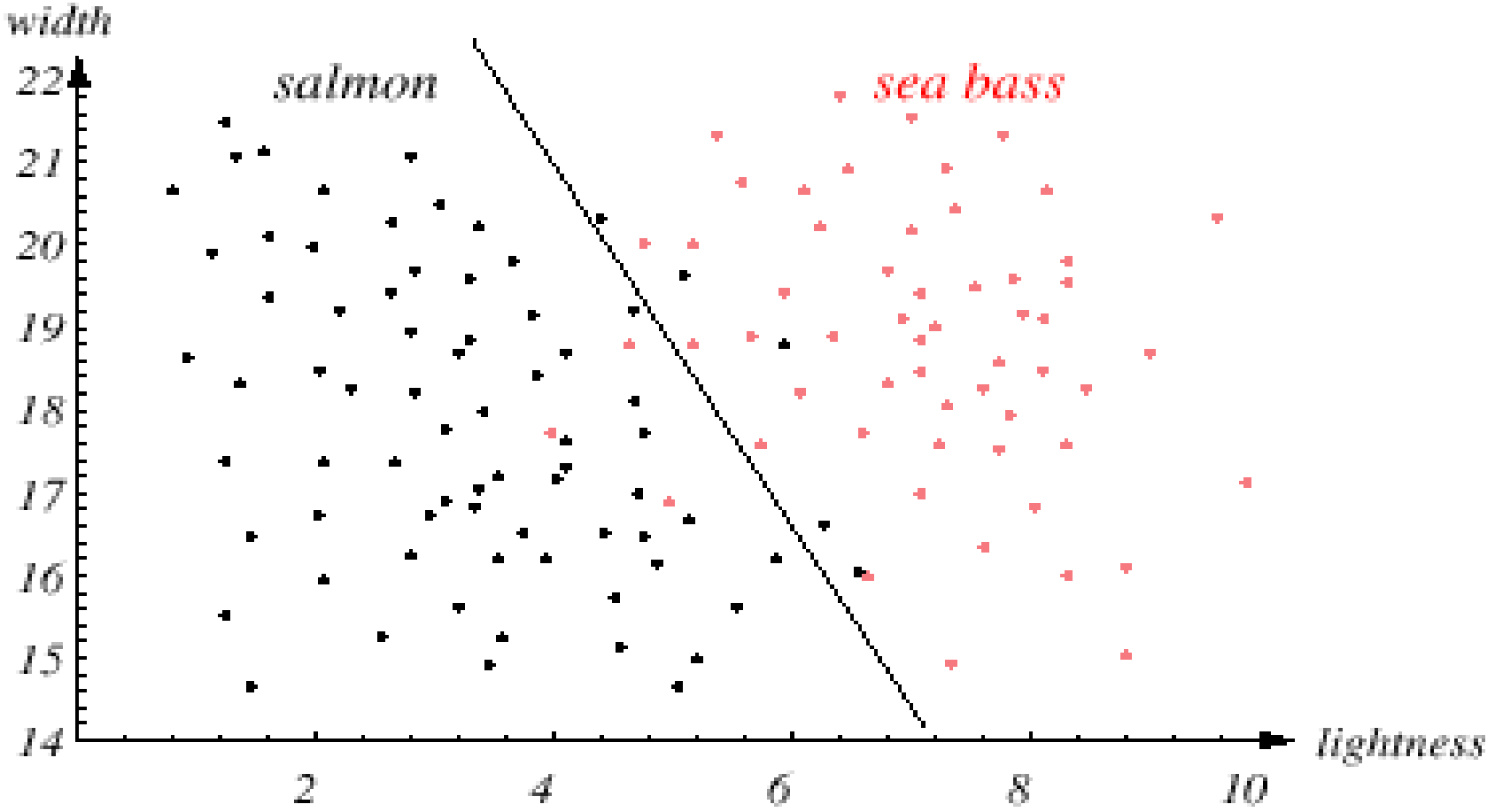
Task of decision theory

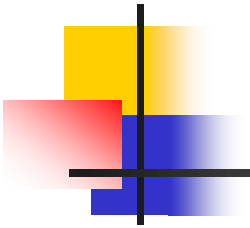
Feature Vector

- Adopt the lightness and add the width of the fish to the feature vector



Straight line decision boundary

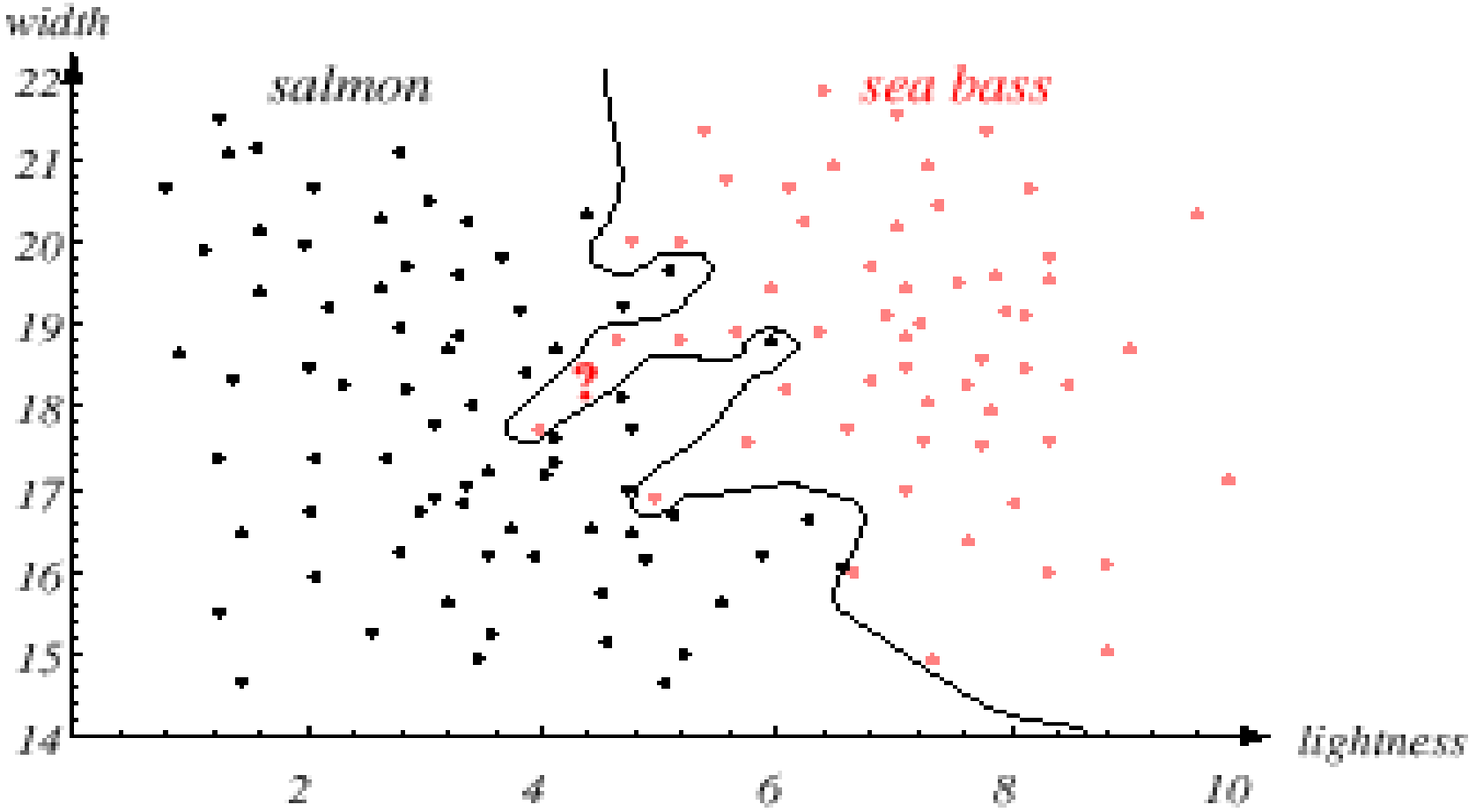




Features

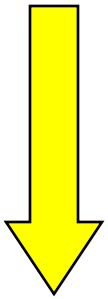
- We might add other features that are not highly correlated with the ones we already have. Be sure not to reduce the performance by adding “noisy features”
- Ideally, you might think the best decision boundary is the one that provides optimal performance on the training data (see the following figure)

Is this a good decision boundary?



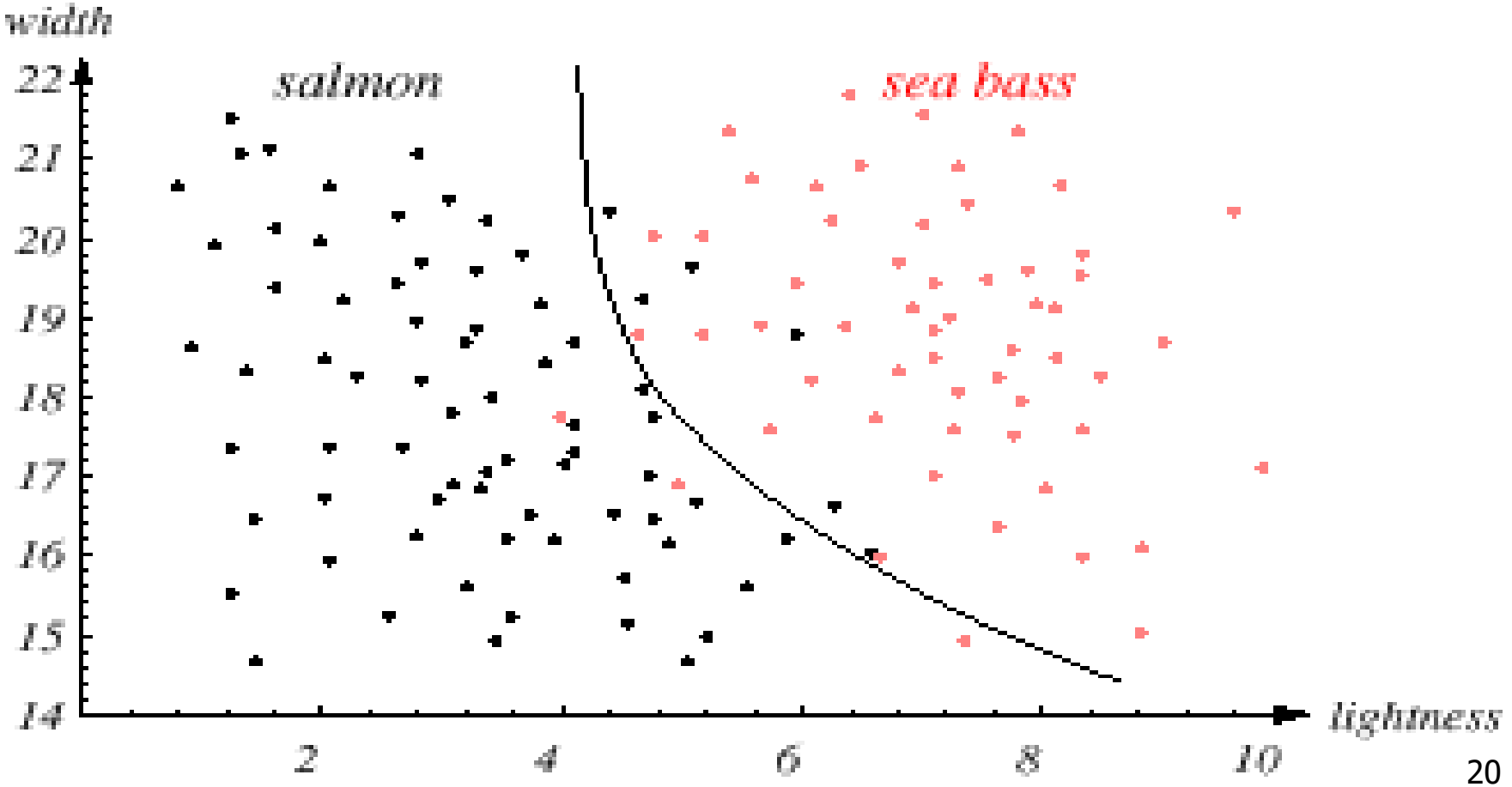
Decision Boundary Choice

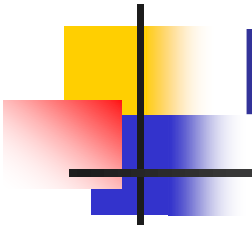
- Our satisfaction is premature because the central aim of designing a classifier is to correctly classify new (test) input



Issue of generalization!

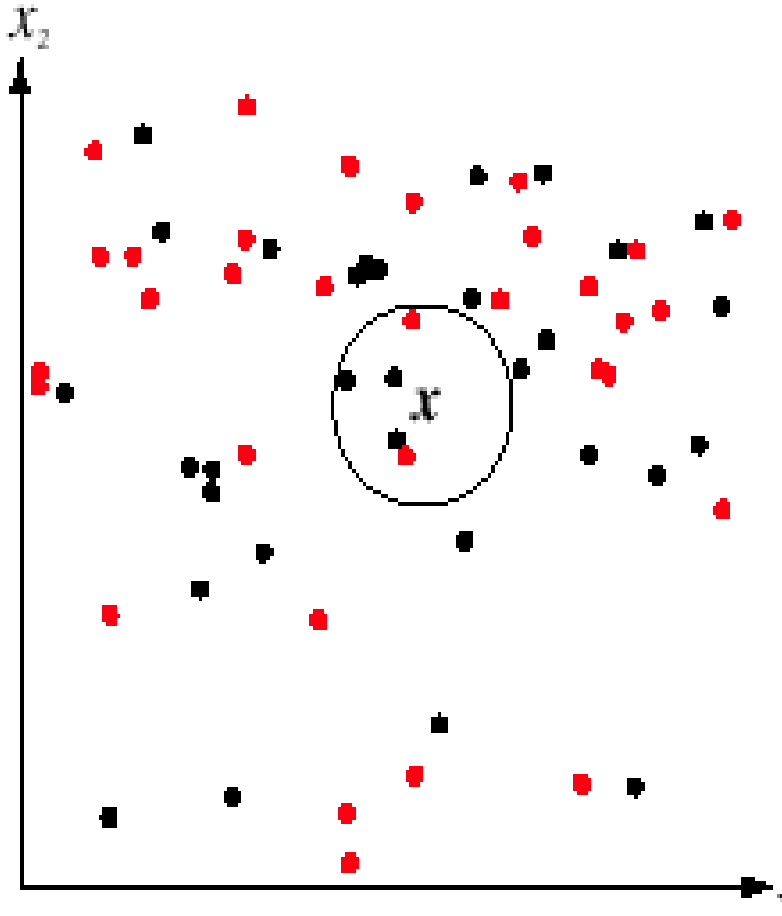
Better decision boundary





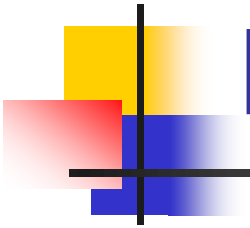
Baysian Decision Theory

- Pure statistical approach – parametric
- Assumes the underlying probability structures are known perfectly
- Makes theoretically optimal decisions



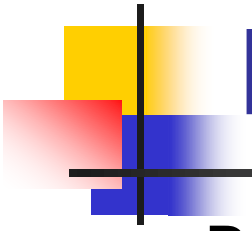
Non-parametric algorithm

FIGURE 4.15. The k-nearest-neighbor query starts at the test point $\mathbf{x}$ and grows a spherical region until it encloses k training samples, and it labels the test point by a majority vote of these samples. In this $k = 5$ case, the test point $\mathbf{x}$ would be labeled the category of the black points. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.



Pattern Recognition Stages

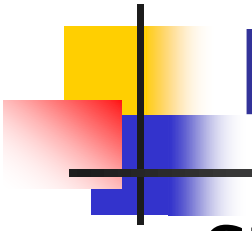
- Sensing
  - Use of a transducer (camera or microphone)
  - PR system depends on the bandwidth, the resolution sensitivity distortion of the transducer
- Preprocessing
  - Segmentation and grouping - patterns should be well separated and not overlap



Pattern Recognition Stages

Reasons for Preprocessing Analysis

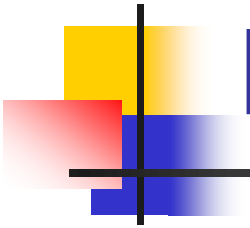
- Gathered data is not in the right format (eg. SQL Database, JSON, SCV etc…)
- Missing values and Outliers
- Scaling and Normalization
- Reduce Intrinsic Noise present in the dataset (part of the stored data might be corrupted)
- Some features in the dataset might not gather any information to our analysis



Pattern Recognition Stages

Steps of Preprocessing

- Data quality assessment
  - Missing values
    - Eliminate rows with missing data
    - Estimate missing values
  - Inconsistent values
  - Duplicate values
- Feature aggregation
  - Reduction of memory consumption and processing time
  - Aggregates is more stable than individual data objects



Pattern Recognition Stages

- Feature Sampling
  - Reduce the size of the dataset
- Dimensionality Reduction
  - Data analysis algorithms works better
  - Models are more understandable and explainable
  - Data may now also get easier to visualize
- Feature Encoding

  For continuous variables
  - Nominal
  - Ordinal

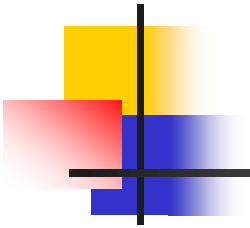
Pattern Recognition Stages

For numeric variables

- Interval
- Ratio

- Feature extraction

  Combining existing features to produce a more useful one.



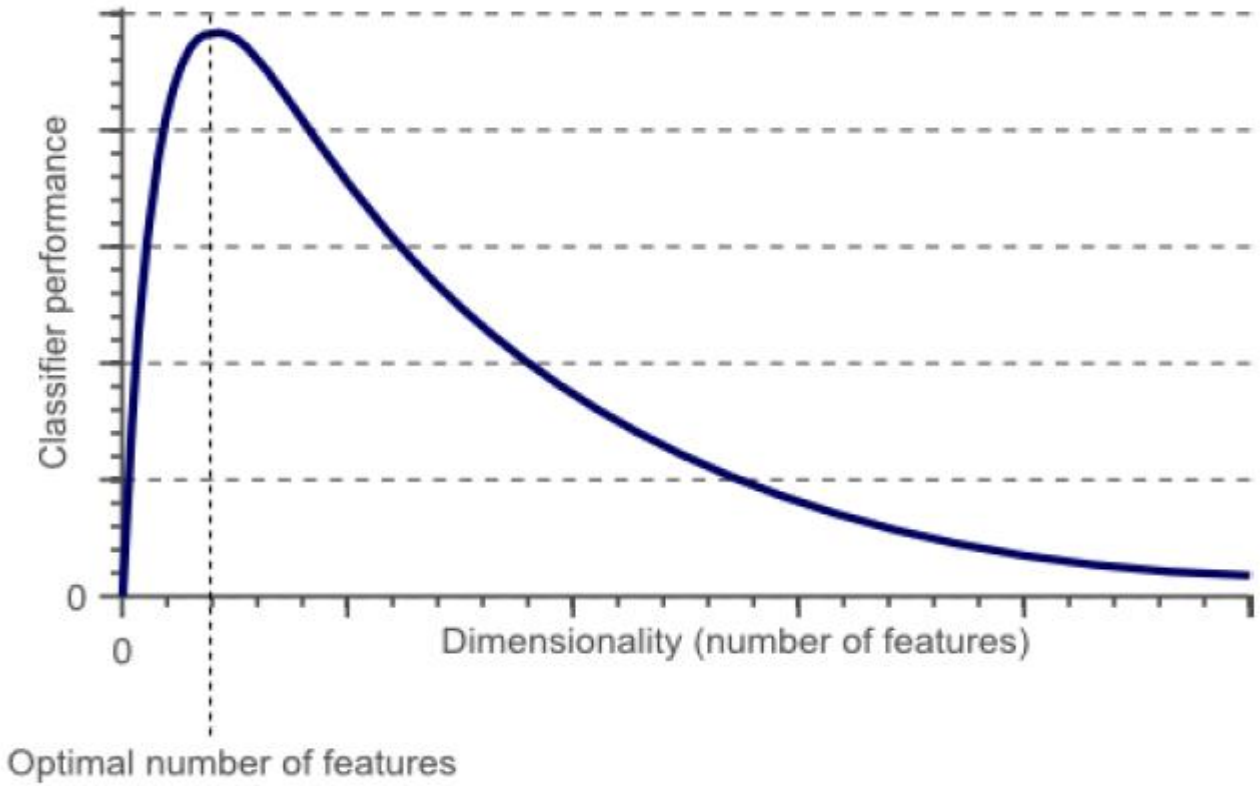
Pattern Recognition Stages (cont)

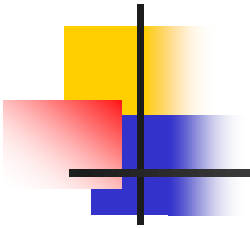
- Techniques
  - Principle Components Analysis (PCA)
  - Independent Components Analysis (ICA)
  - Linear Discriminant Analysis (LDA)
  - Locally Linear Embedding (LLE)
  - t-Distributed Stochastic Neighbor Embedding (t-SNE)
  - Auto Encoders (AE)

Pattern Recognition Stages (cont)

Pattern Recognition Stages (cont)

Relationship between Classifier Performance and Dimensionality



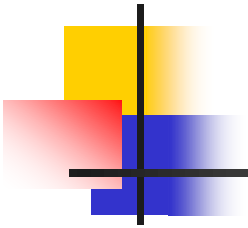


Pattern Recognition Stages (cont)

- Feature Selection

  Reducing number of features have several benefits:

  - Accuracy improvements
  - Overfitting risk reduction
  - Speedup in training
  - Improved data visualization
  - Increase in explainability of our model



Pattern Recognition Stages (cont)

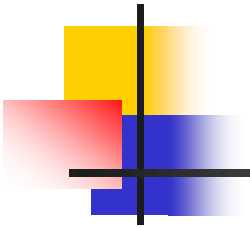
- Feature Selection Methods
  - Filter Method
  - Wrapper Method
  - Embedded Method
  - Recursive Feature Elimination
  - SelecFromModel
  - Correlation Matrix Analysis
  - Univariate Selection
  - Lasso Regression

Pattern Recognition Stages (cont)

- Train / Validation / Test Split

  After feature encoding is done, it is always advised to split the dataset into two, sometimes three parts.

Data Split into parts



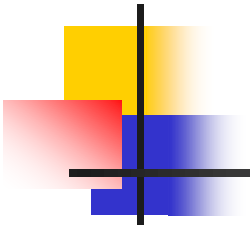
Pattern Recognition Stages (cont)

- Classification

  Classification is a supervised learning approach in which the computer program learns from the input data and then uses this learning to classify new observations.

  Types of classification algorithms in machine learning:

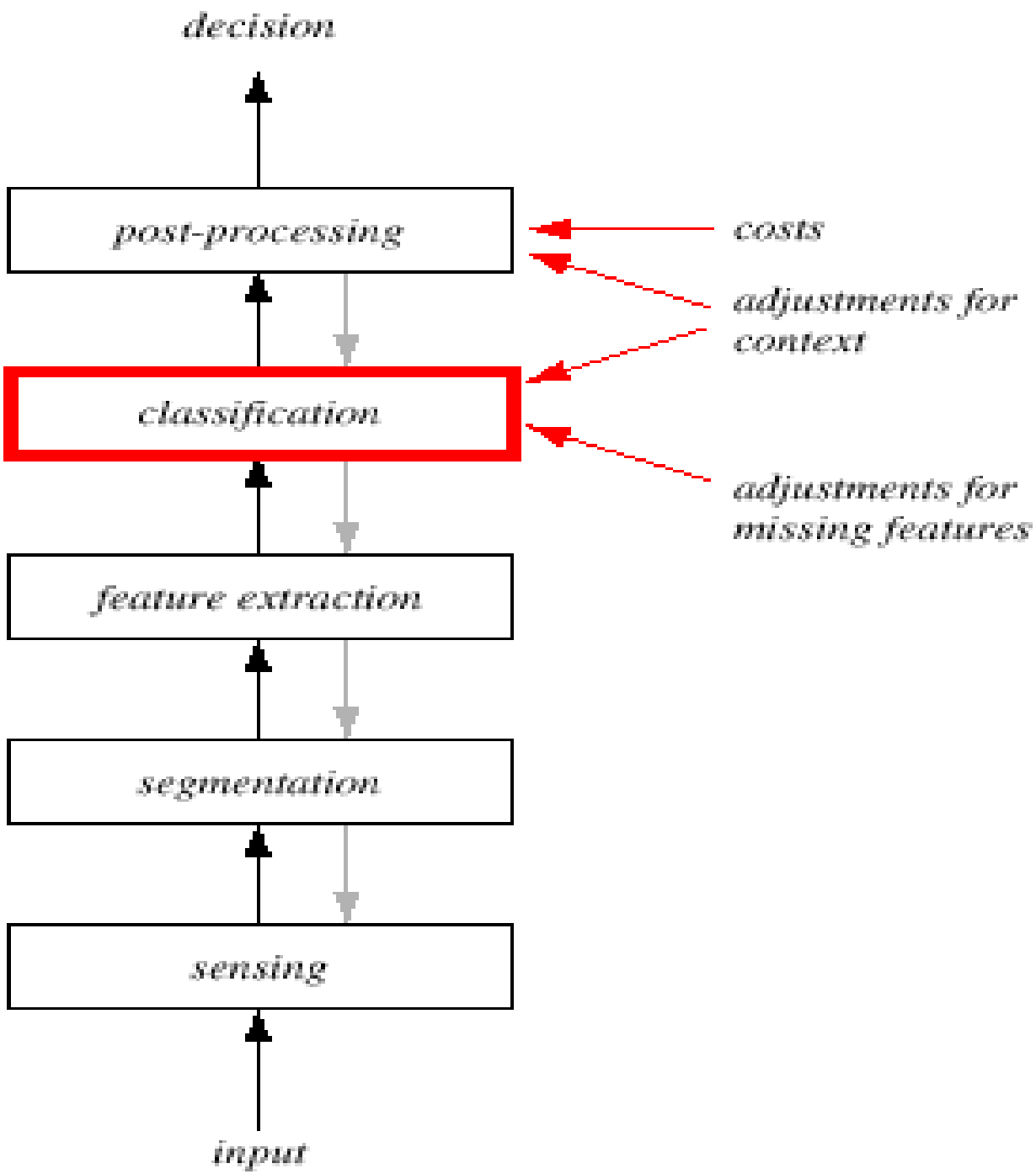
  - Linear Classifiers: Logistic Regression, Naïve Bayes Classifier
  - Nearest Neighbor
  - Support Vector Machines
  - Decision Trees

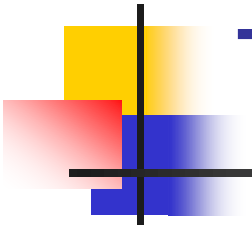


Pattern Recognition Stages (cont)

- Boosted Trees
- Random Forest
- Neural Networks

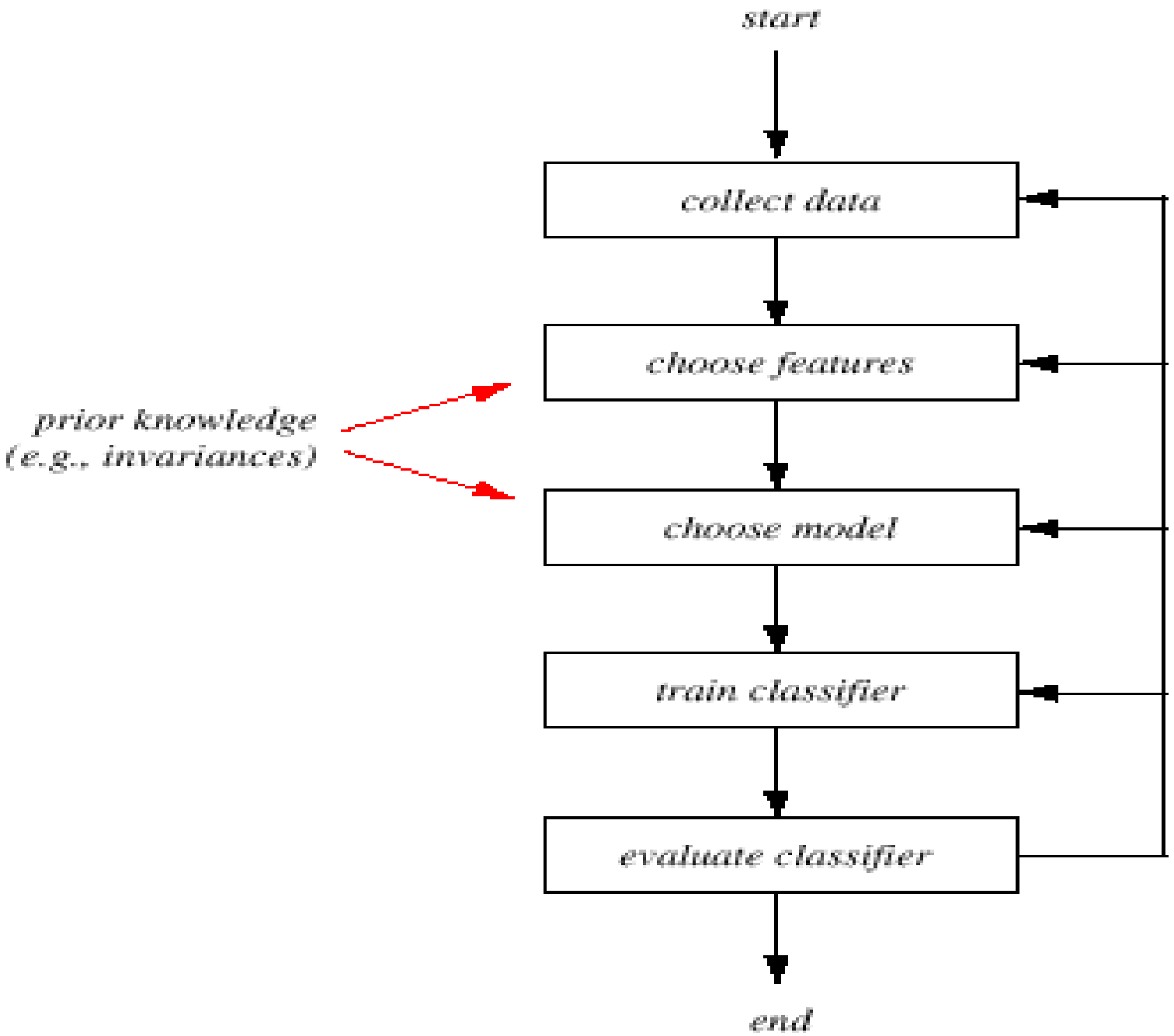
- Post Processing
  - Exploit context-dependent information to improve performance

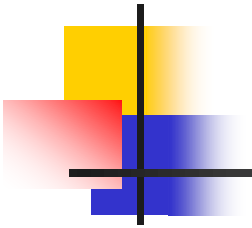

decision
post-processing
classification
feature extraction
segmentation
sensing
input
costs
adjustments for context
adjustments for missing features



The Design Cycle

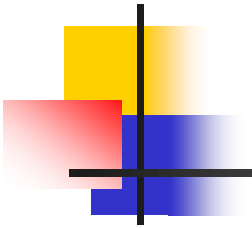
- Data collection
- Feature Choice
- Model Choice
- Training
- Evaluation
- Computational Complexity


start
collect data
choose features
prior knowledge
(e.g., invariances)
choose model
train classifier
evaluate classifier
end



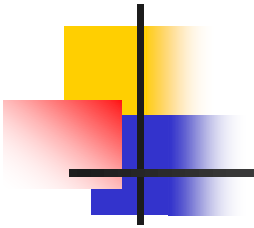
Data Collection

- How do we know when we have collected an adequately large and representative set of examples for training and testing the system?



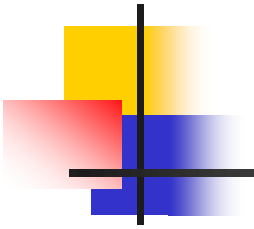
Choice of Features

- Depends on the characteristics of the problem domain
- Simple to extract, invariant to irrelevant transformations, insensitive to noise



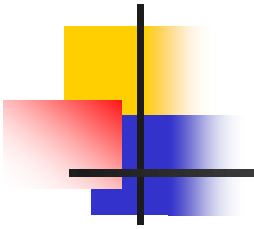
Model Choice

- Unsatisfied with the performance of our fish classifier and want to jump to another class of model



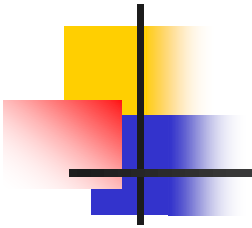
Training

- Use data to determine the classifier
- (Many different procedures for training classifiers and choosing models)



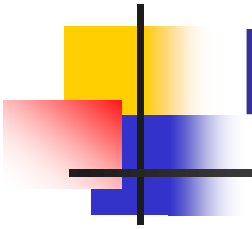
Evaluation

- Measure the error rate (or performance)
- Possibly switch from one set of features to another one



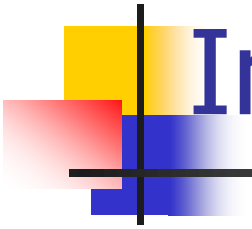
Computational Complexity

- What is the trade-off between computational ease and performance?
- How does an algorithm scale as a function of the number of features, patterns, or categories?



Learning and Adaptation

- Supervised learning
  - A teacher provides a category label for each pattern in the training set
- Unsupervised learning
  - The system forms clusters or “natural groupings” of the unlabeled input patterns



Introductory example conclusion

- Reader may be overwhelmed by the number, complexity, and magnitude of the sub-problems of Pattern Recognition
- Many of these sub-problems can indeed be solved
- Many fascinating unsolved problems still remain