Pattern Recognition

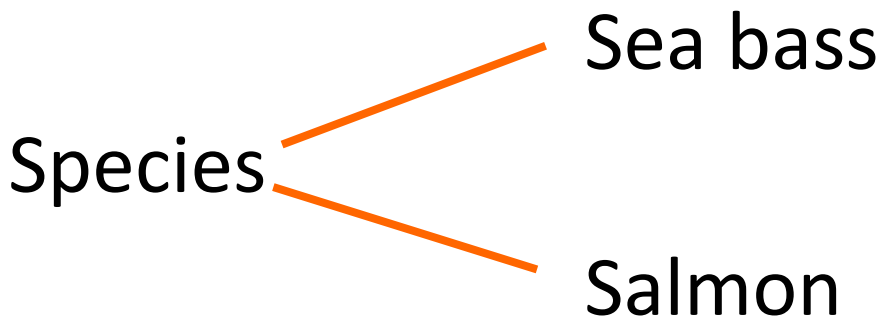
Dr. R. K. Chaurasiya

Dept. Electronics and Communication Engineering

MANIT Bhopal

An Example

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



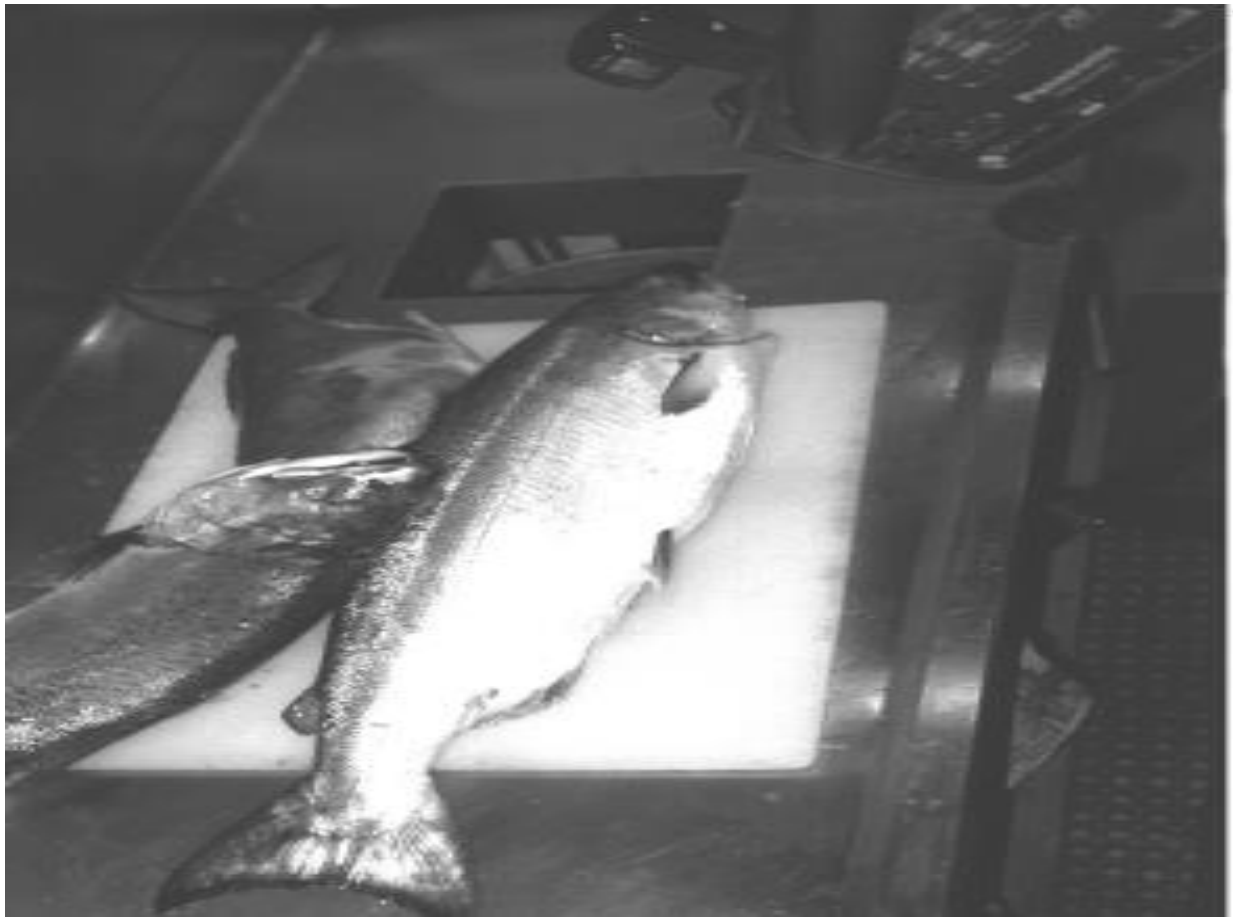
An Example

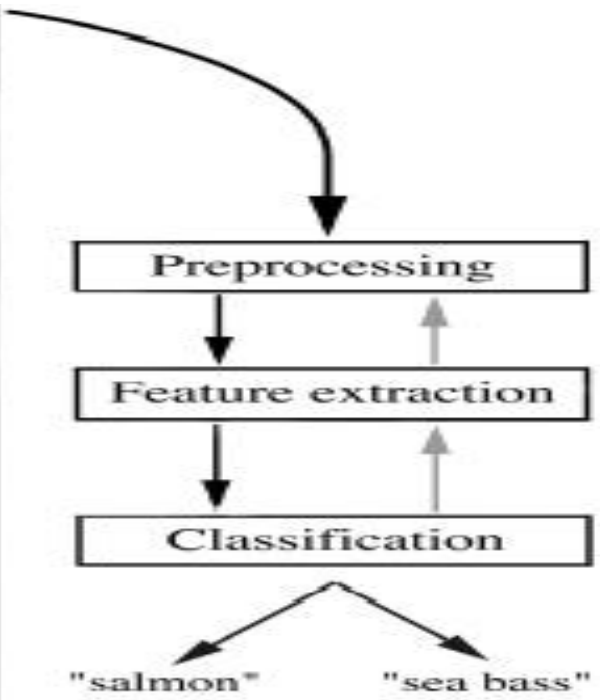
- Problem Analysis
  - Set up a camera and take some sample images to extract features
    - Length
    - Lightness
    - Width
    - Number and shape of fins
    - Position of the mouth, etc…

    - This is the set of all suggested features to explore for use in our classifier!

An Example

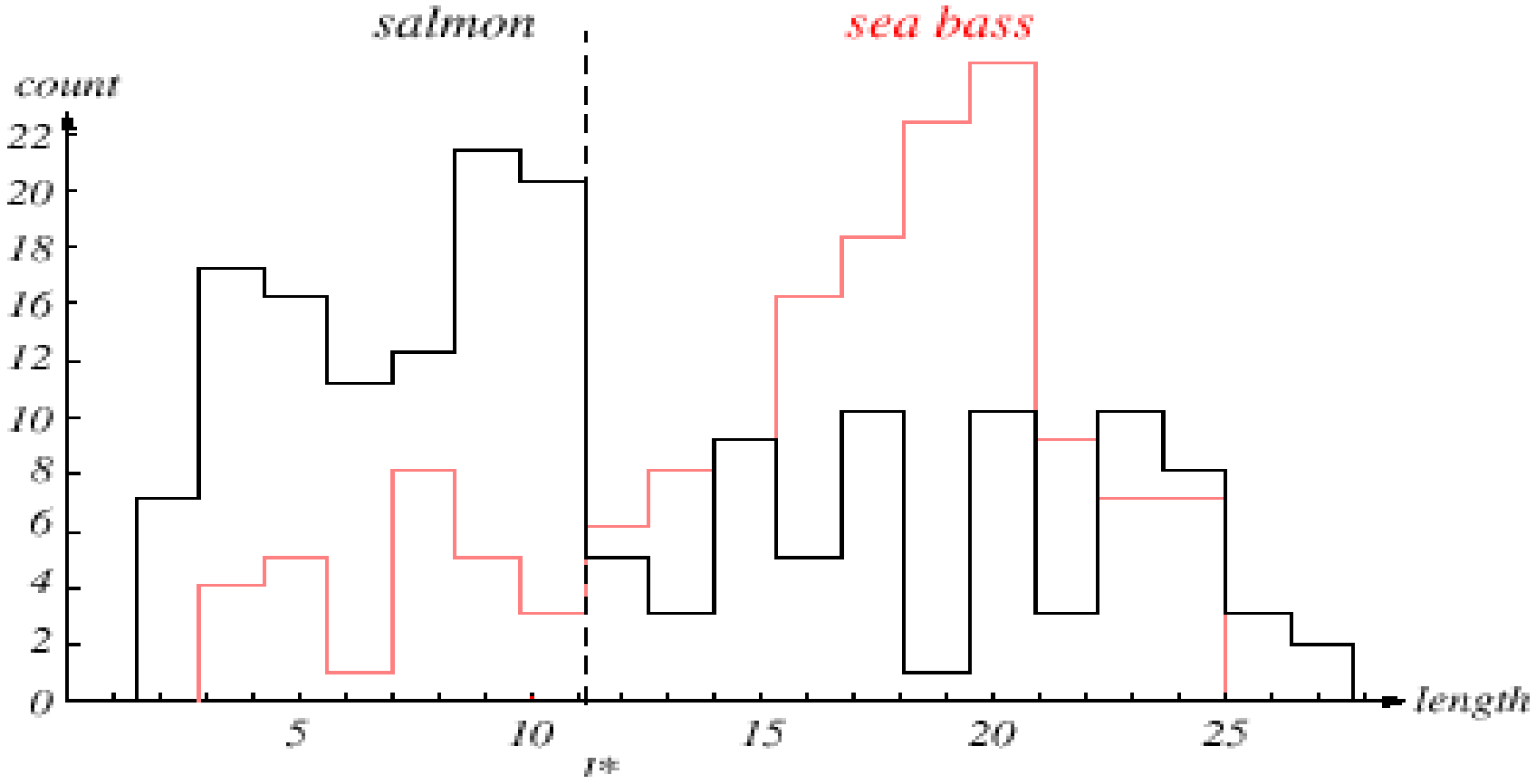
- Preprocessing
  - Use a segmentation operation to isolate fishes from one another and from the background
- Information from a single fish is sent to a feature extractor whose purpose is to reduce the data by measuring certain features
- The features are passed to a classifier




Preprocessing
Feature extraction
Classification
"salmon"
"sea bass"

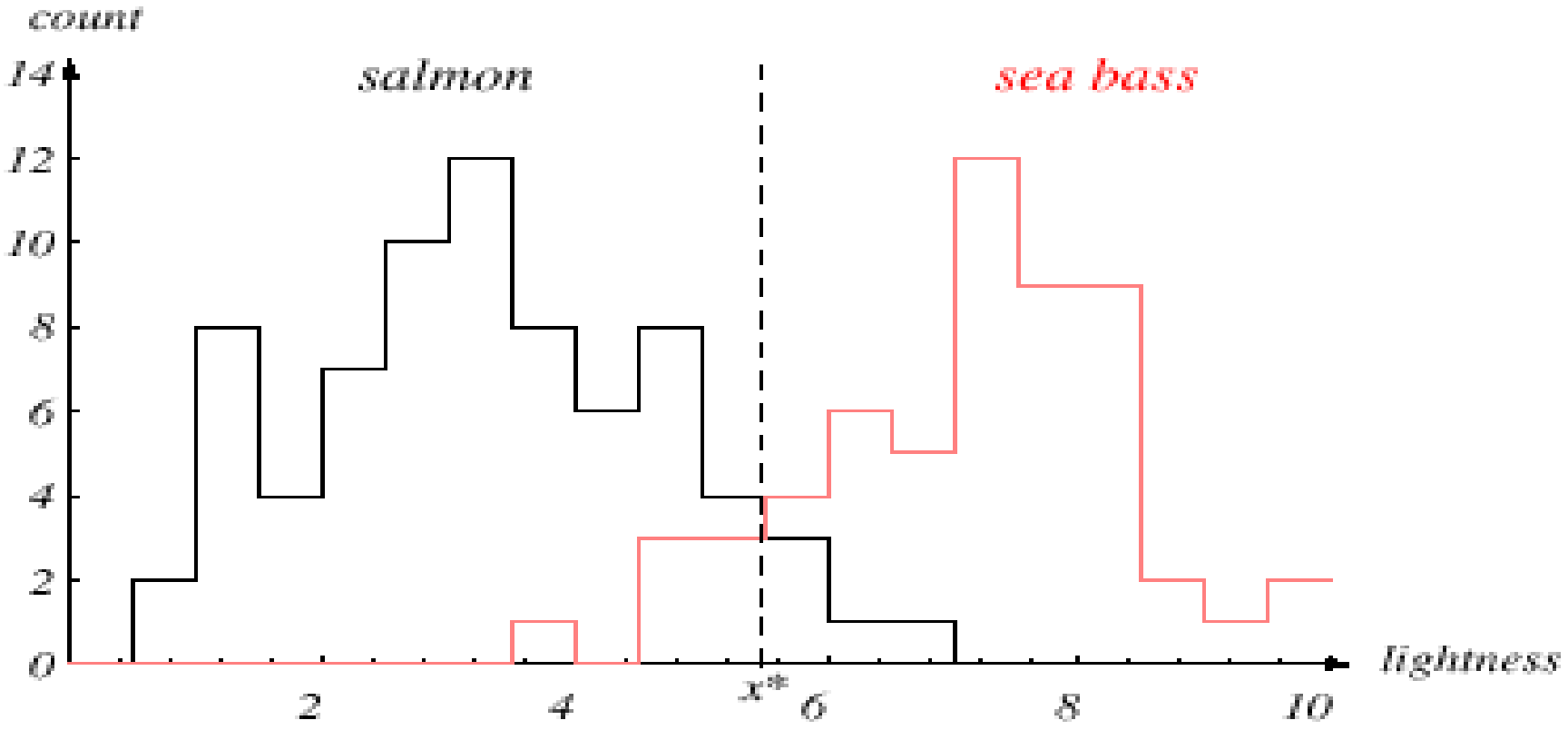
Classification

– Select the length of the fish as a possible feature for discrimination

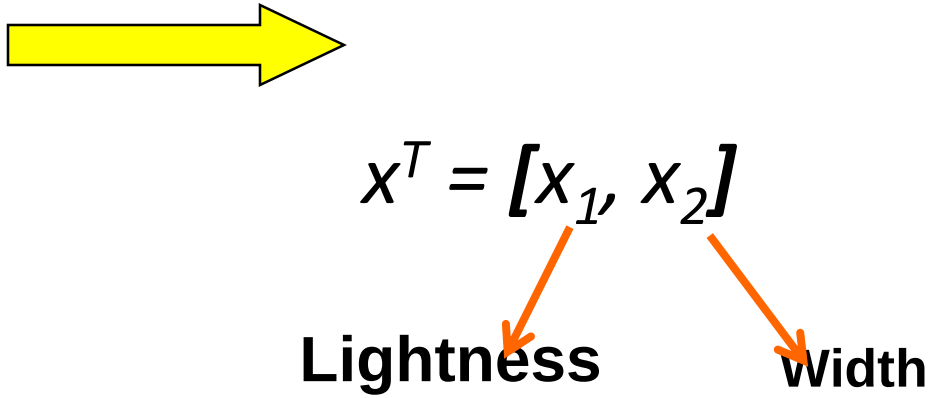

salmon
sea bass
count
22
20
18
16
12
10
8
6
4
2
0
5
10
15
20
25
length
l*

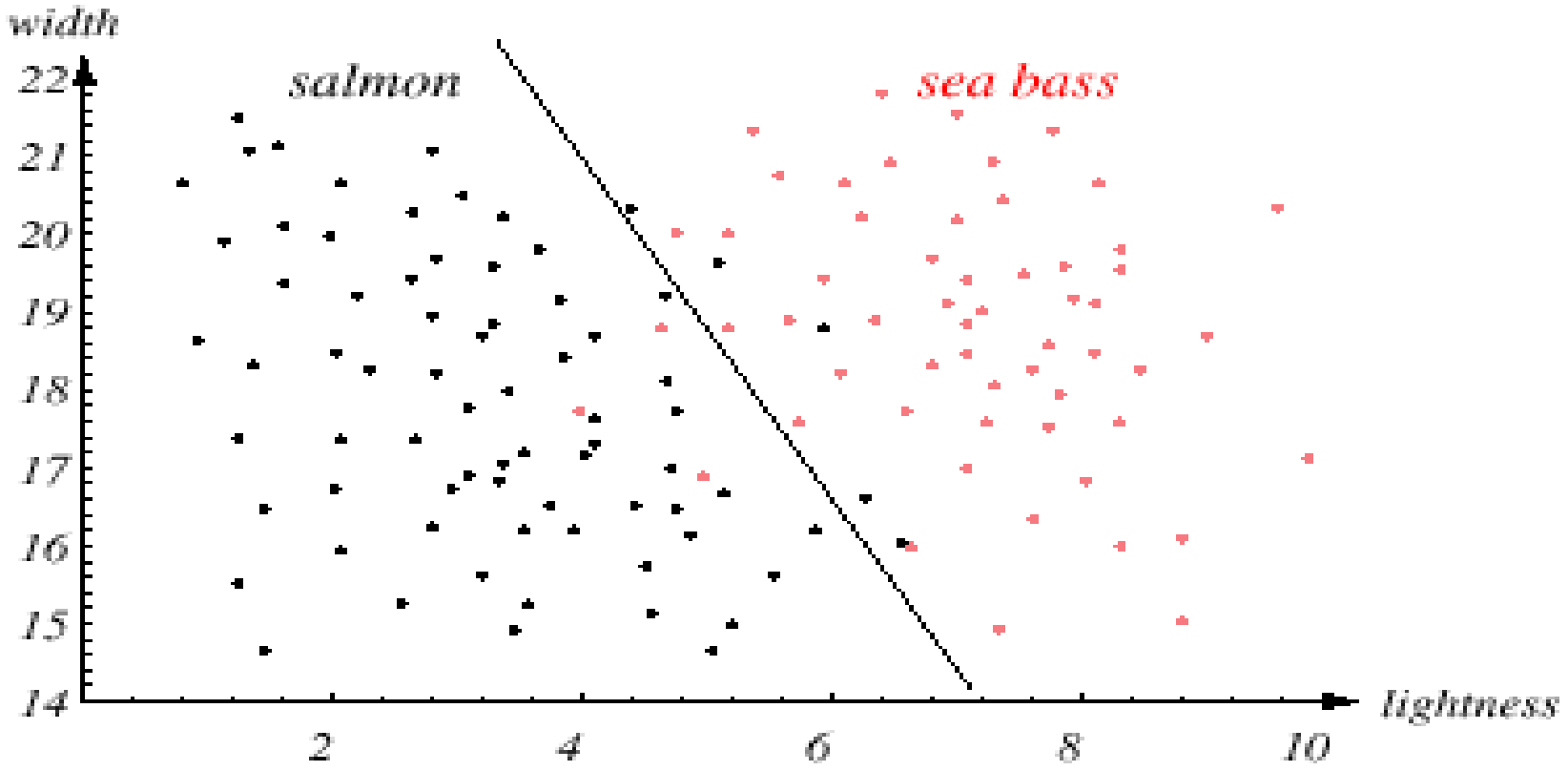
The **length** is a poor feature alone!

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

- Adopt the lightness and add the width of the fish

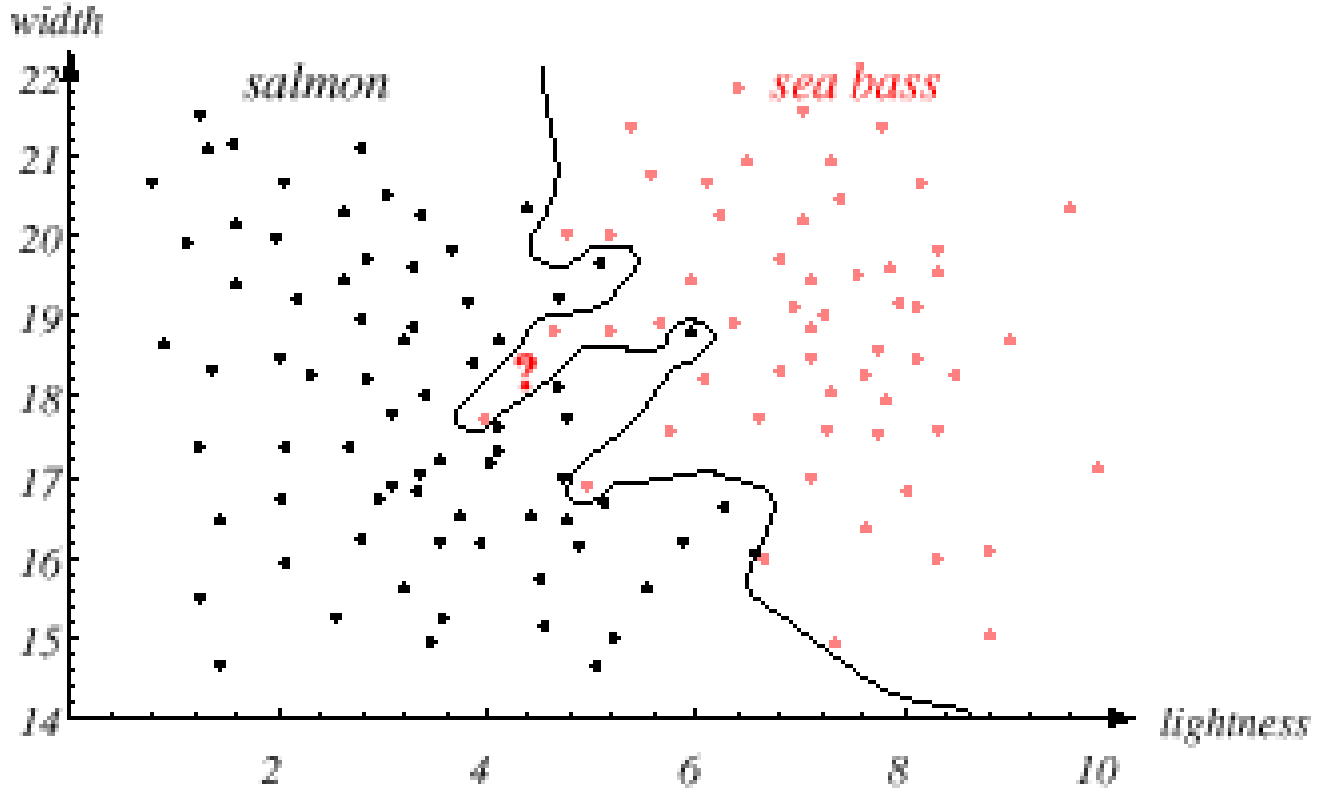



width
22
21
20
19
18
17
16
15
14
salmon
sea bass
2
4
6
8
10
lightness

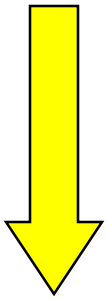
- We might add other features that are not correlated with the ones we already have. A precaution should be taken not to reduce the performance by adding such “noisy features”

- Ideally, the best decision boundary should be the one which provides an **optimal performance** such as in the following figure:

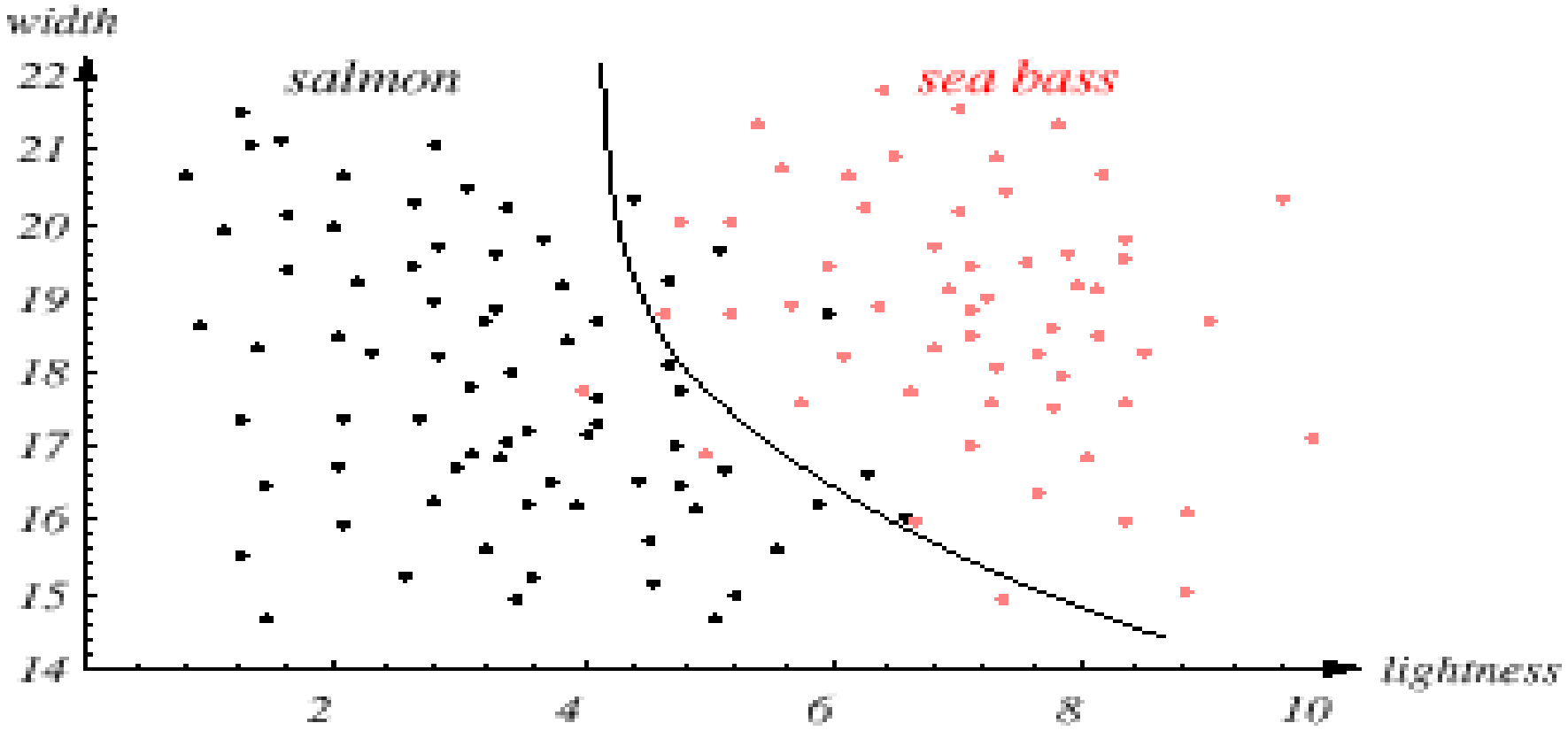
Zero Training Error Classifier-Do we need this?



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



Issue of generalization!


width
22
21
20
19
18
17
16
15
14
salmon
sea bass
2
4
6
8
10
lightness

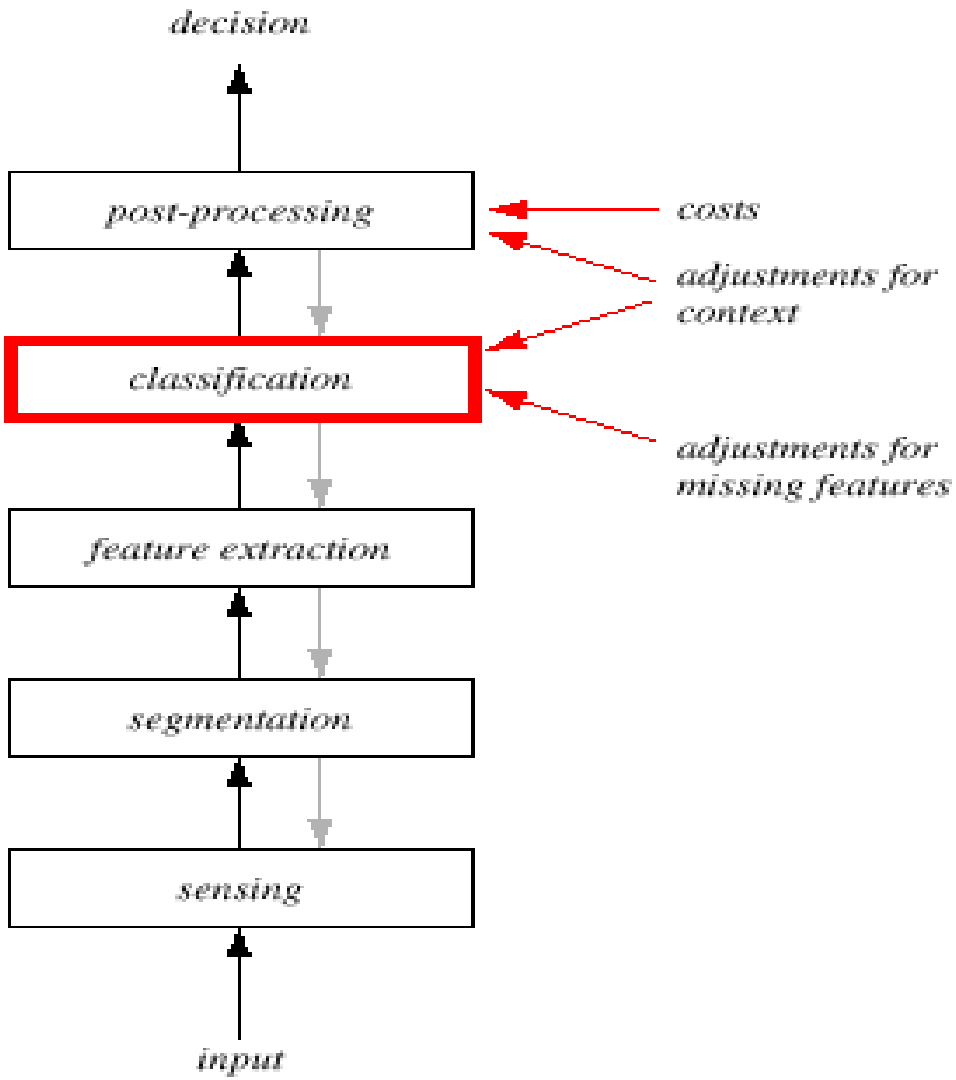
Pattern Recognition Systems

- Sensing
  - Use of a transducer (camera or microphone)
  - PR system depends on the bandwidth, the resolution, sensitivity, distortion, S/N ratio, latency of the transducer
- Segmentation and grouping
  - Patterns should be well separated and should not overlap

- Feature extraction
  - Discriminative features
  - Invariant features with respect to translation, rotation and scale.
- Classification
  - Use a feature vector provided by a feature extractor to assign the object to a category
- Post Processing
  - Exploit context input dependent information other than from the target pattern itself to improve performance

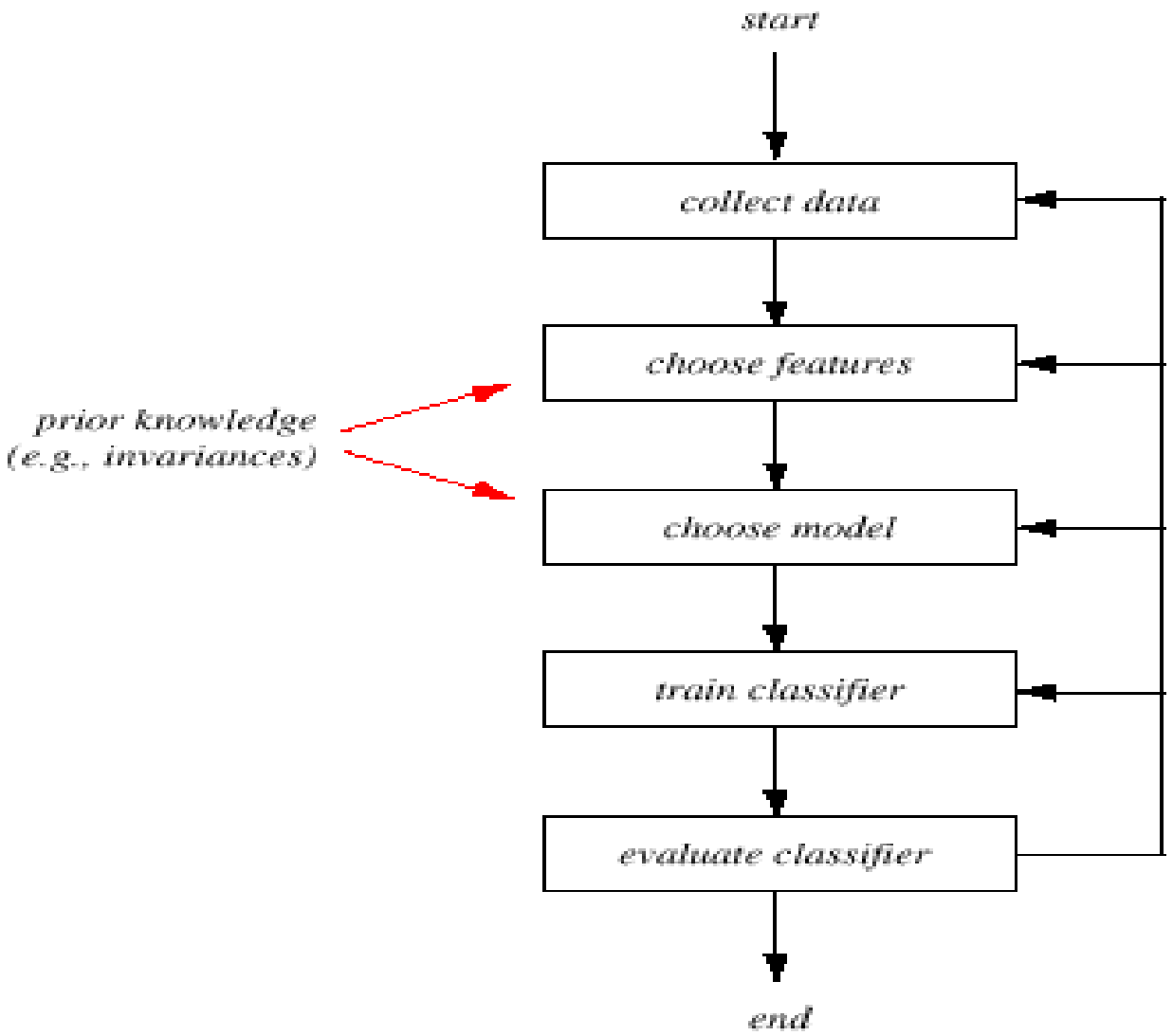
Post Processing

- Error rate- the percentage of new patterns that are assigned to the wrong category (seek min error rate)
- Risk- recommend actions that will minimize the total expected cost
- Context- exploited to improve system performance (input-independent info other than from the target itself)
- Multiple classifiers- each classifier operating on different aspects of the input (super classifier)


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

The PR 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?

- Feature Choice
  - Depends on the characteristics of the problem domain.
  - Find features that are: Simple to extract, invariant to irrelevant transformations, insensitive to noise, and useful in discriminating patterns in different categories.
  - Prior knowledge plays a major role- how to combine prior knowledge and empirical data to find relevant and effective features?

- Model Choice

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

    - Based on some function of the number and position of the fins, the color of the eyes, the weight, shape of the mouth …

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

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

Other Performance Measurement Criterions

True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

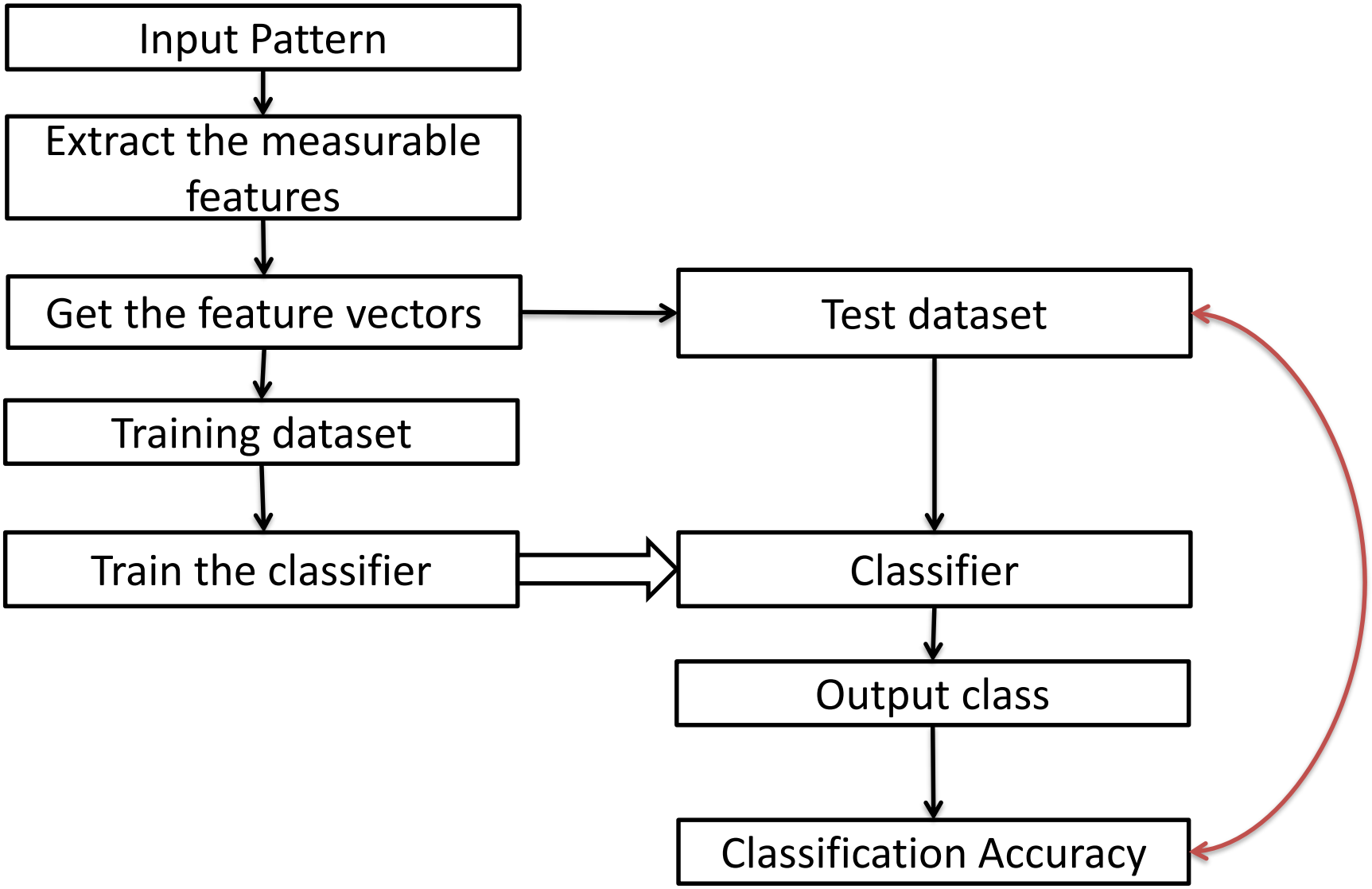
	p' (Predicted)	n' (Predicted)
p (Actual)	True Positive	False Negative
n (Actual)	False Positive	True Negative

$$Sensitivity = TPR = \frac{TP}{TP+FN} \times 100\%$$

$$Specificity = TNR = \frac{TN}{TN+FP} \times 100\%$$

$$Accuracy = \frac{Sensitivity+Specificity}{2} \times 100\%$$

PR/ML System Flowchart



- Computational Complexity
  - What is the trade-off between computational ease and performance?
  - (How an algorithm scales as a function of the number of features, patterns or categories?)

Learning and Adaptation

- Supervised learning
  - A teacher provides a category label or cost for each pattern in the training set, and seeks to reduce the sum of the costs of these patterns
- Unsupervised learning
  - The system forms clusters or “natural groupings” of the input patterns