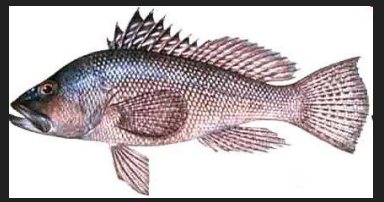
Classification

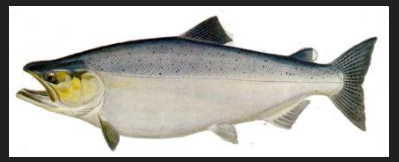
Example

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

Species
- Sea bass

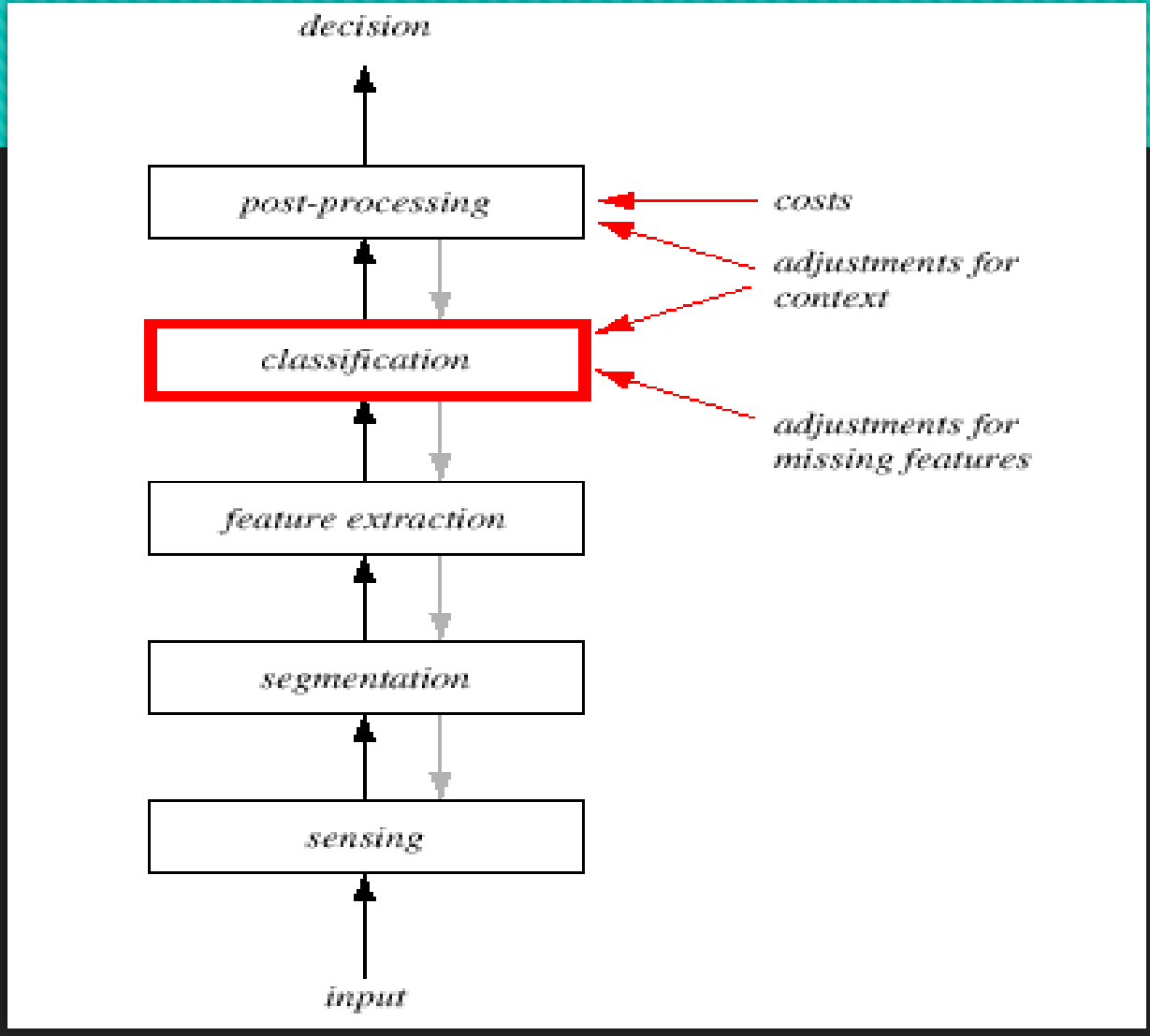


- Salmon



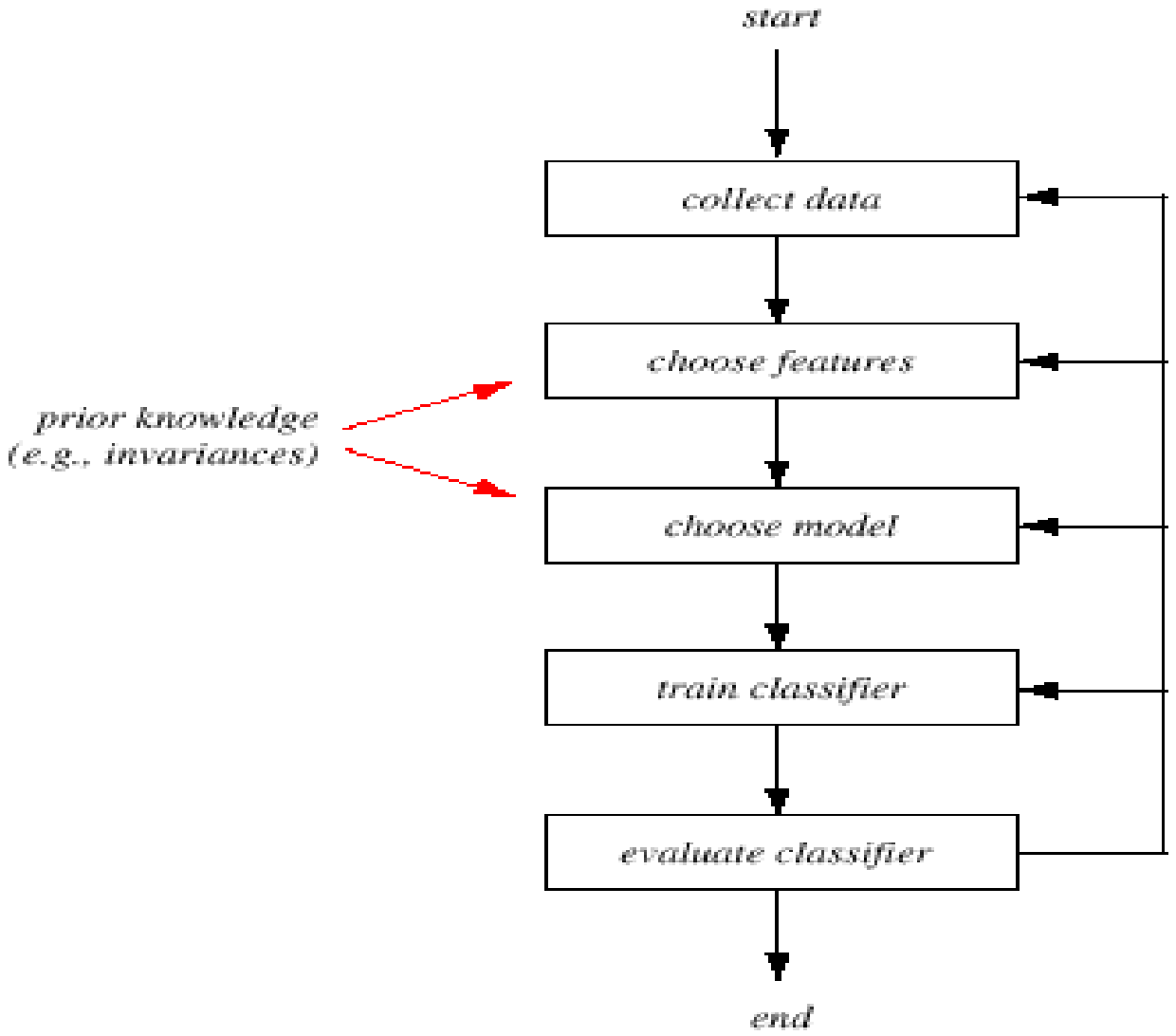
Problem Analysis

- 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!


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

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.


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

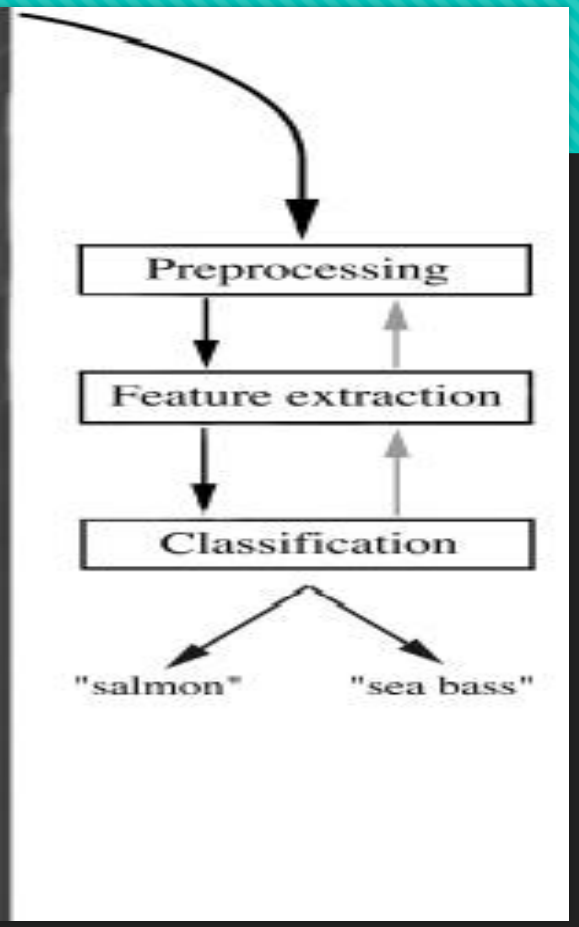
- Data Collection
  - How do we know when we have collected an adequately large and representative set of examples for training and testing the system?

- Feature Choice
  - Depends on the characteristics of the problem domain. Simple to extract, invariant to irrelevant transformation insensitive to noise.

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

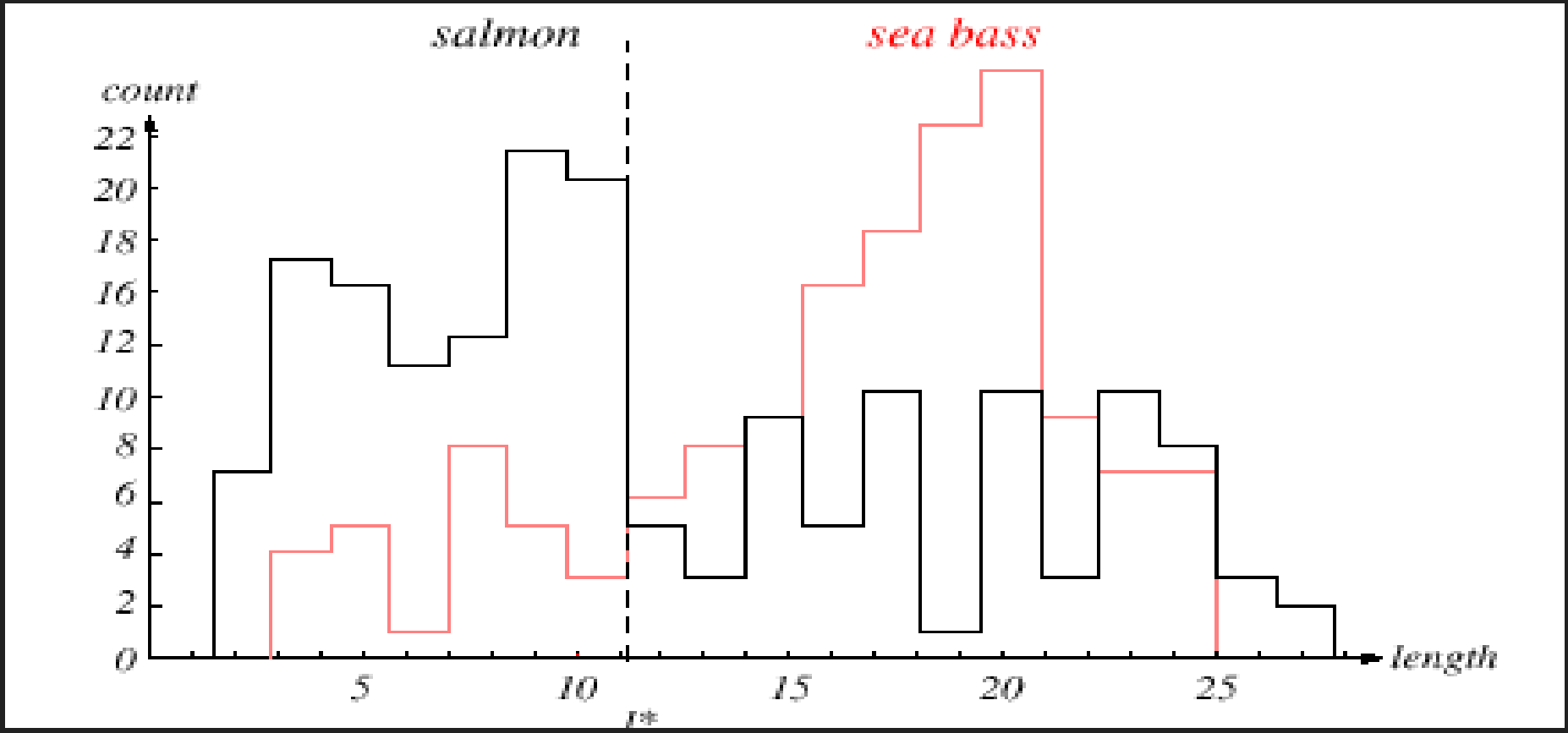
- Training
  - Use data to determine the classifier. Many different procedures for training classifiers and choosing models

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


Preprocessing
Feature extraction
Classification
"salmon"
"sea bass"

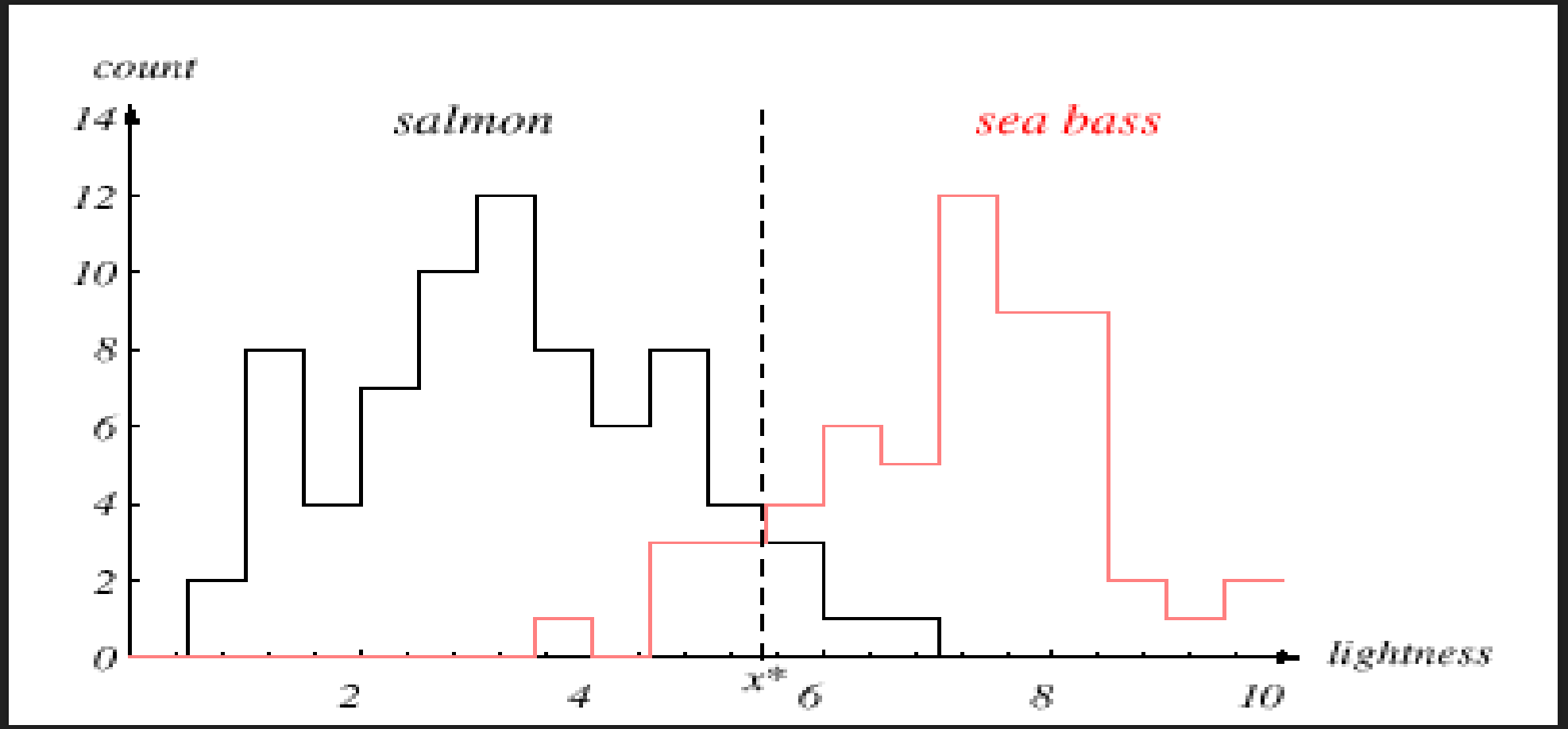
Classification

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



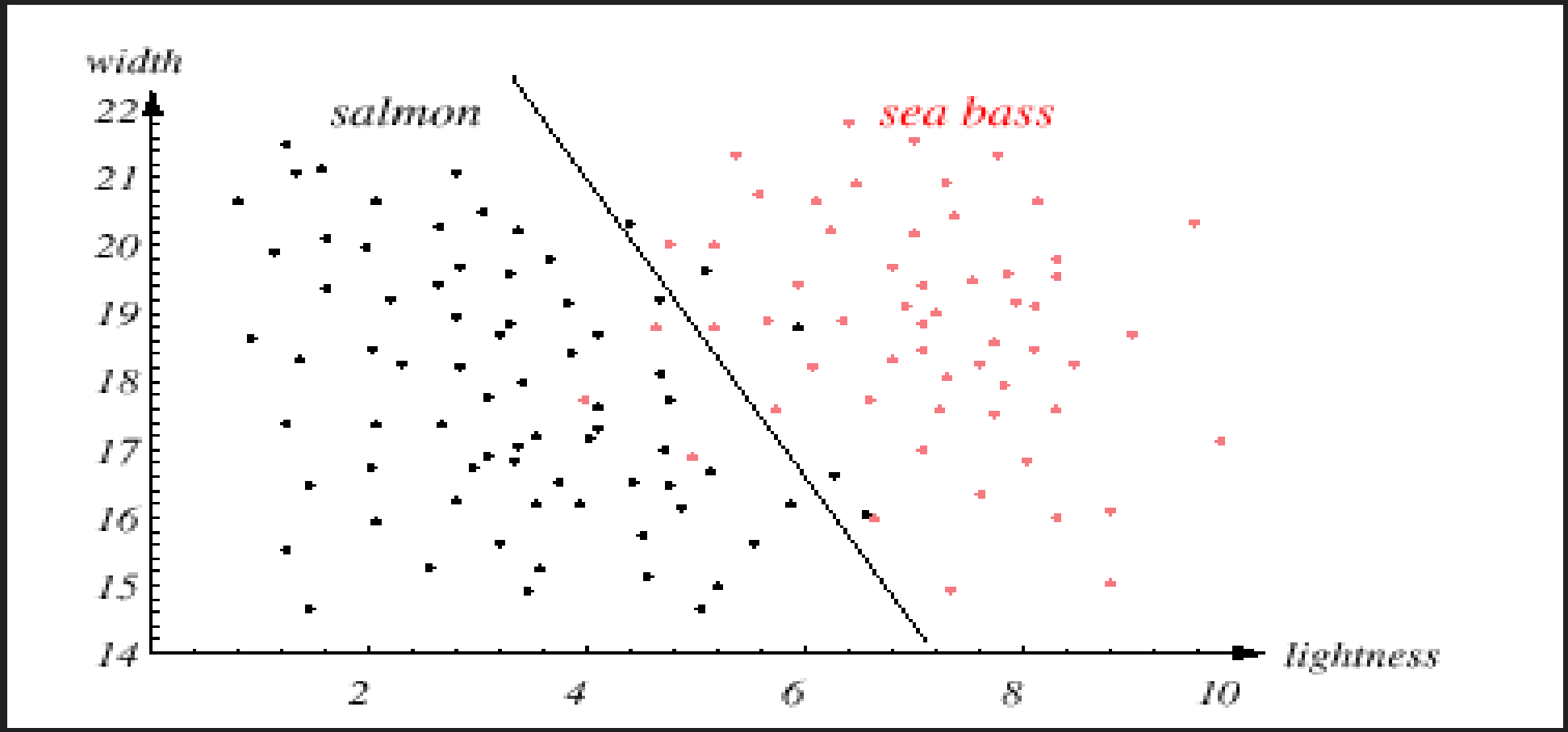
Classification

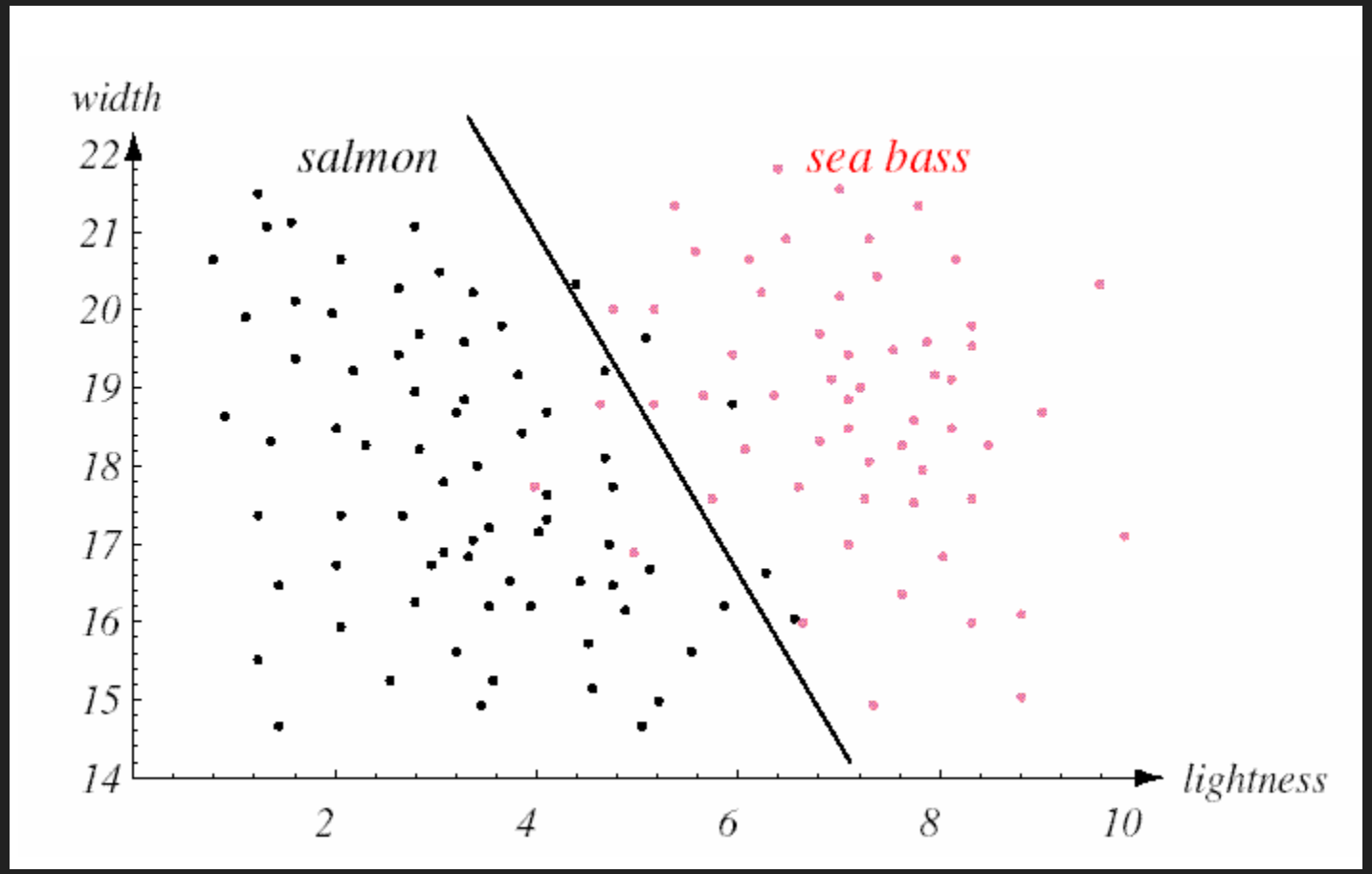
- Select the lightness of the fish as a possible feature for discrimination



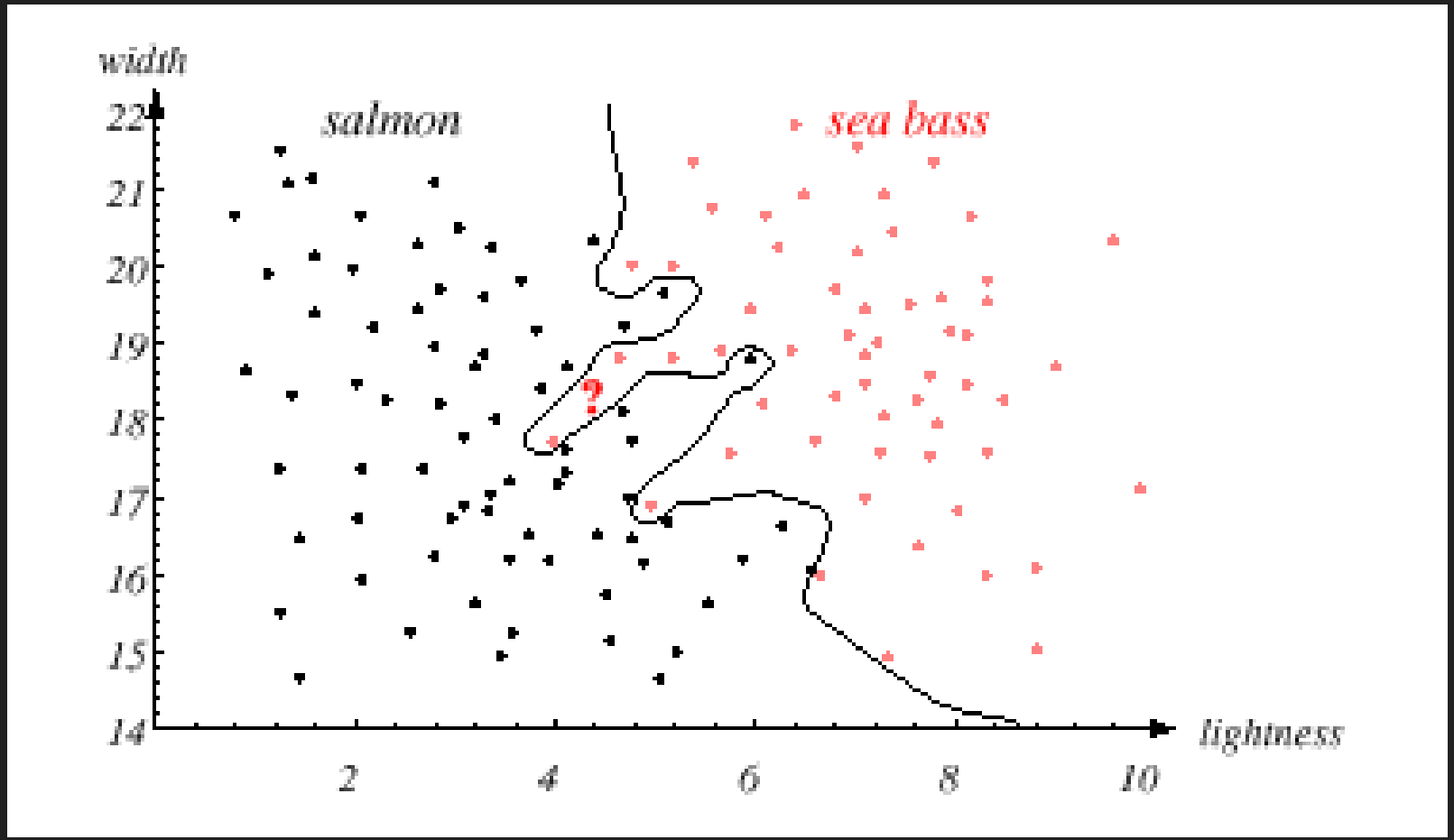
- Adopt the lightness and add the width of the fish

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

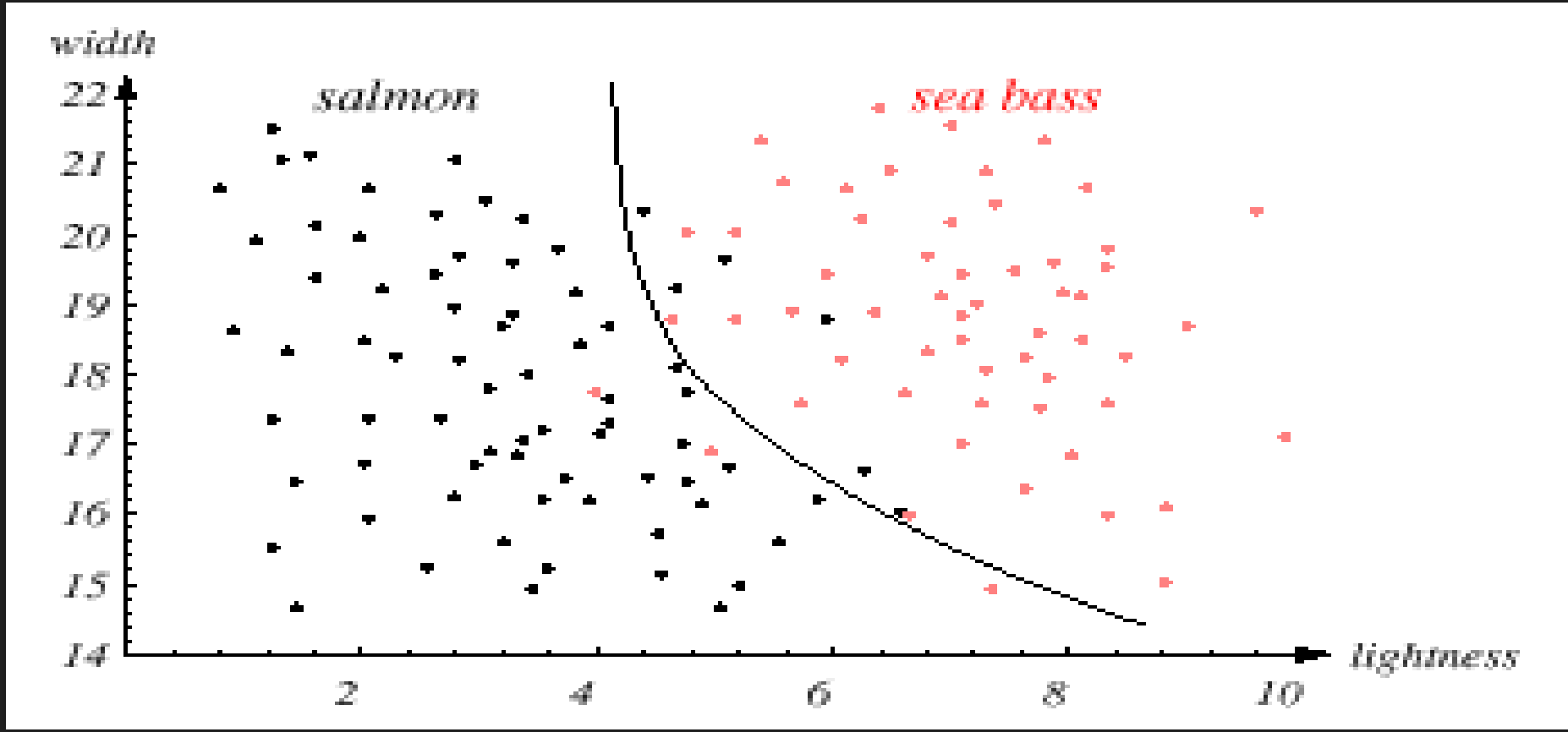

width
salmon
sea bass
22
21
20
19
18
17
16
15
14
2
4
6
8
10
lightness


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

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



Issue of generalization



No free lunch theorem

- هیچ کلاسیفایری آچار فرانسه نیست!
- با توجه به مسائل مشابهی که در گذشته کار شده، تصمیم گیری می کنیم.
- مسئله جدید ← آزمون و خطا با روش های مختلف

Classifier ensembles

Get the power of multiple models for a single decision.

Tripple trade-off

1. Complexity, C
2. Training set size, N
3. Generalization error, E