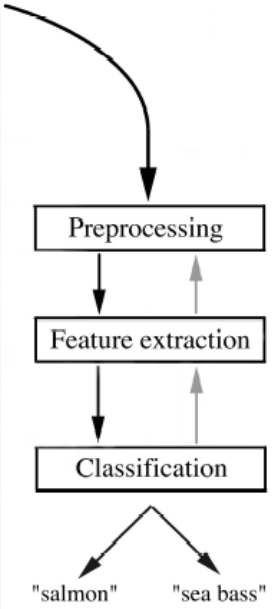
A Fishy Example I

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

- Salmon or Sea Bass?



Problem Analysis

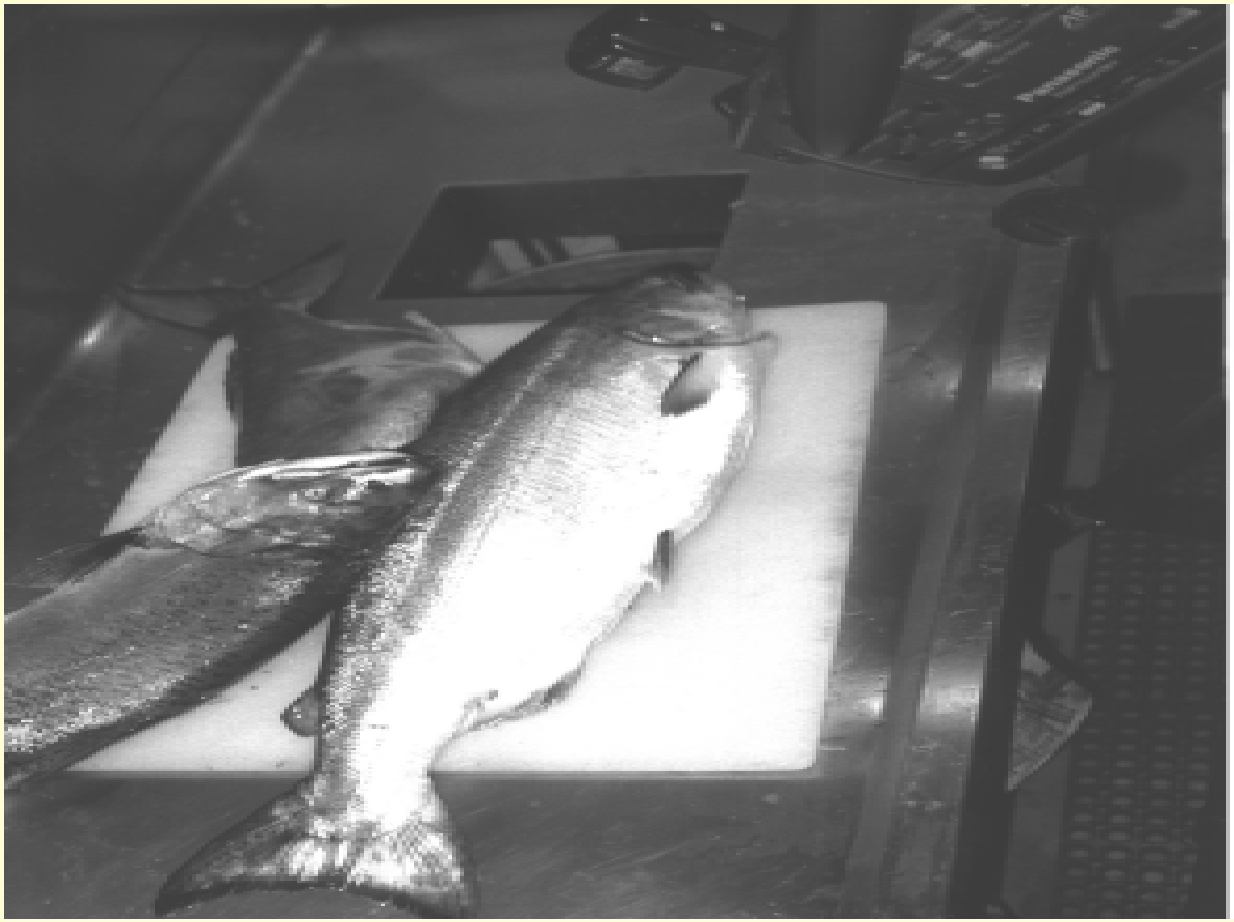
- Set up a camera and take some sample images to extract features
  - Length
  - Lightness
  - Width
  - Number and shape of fins(yüzgeç)
  - Position of the mouth, etc…

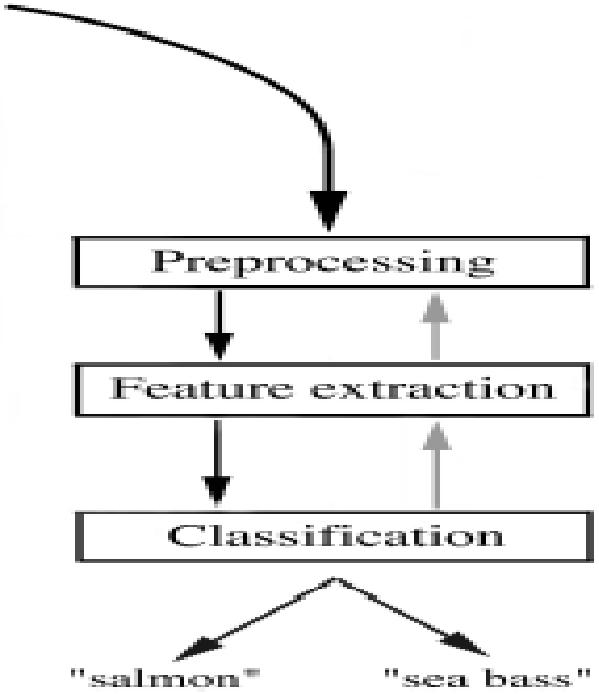
  This is the set of all suggested features to explore for use in our classifier!

Solution by Stages

- Preprocess raw data from camera
- Segment isolated fish
- Extract features from each fish (length,width, brightness, etc.)
- Classify each fish

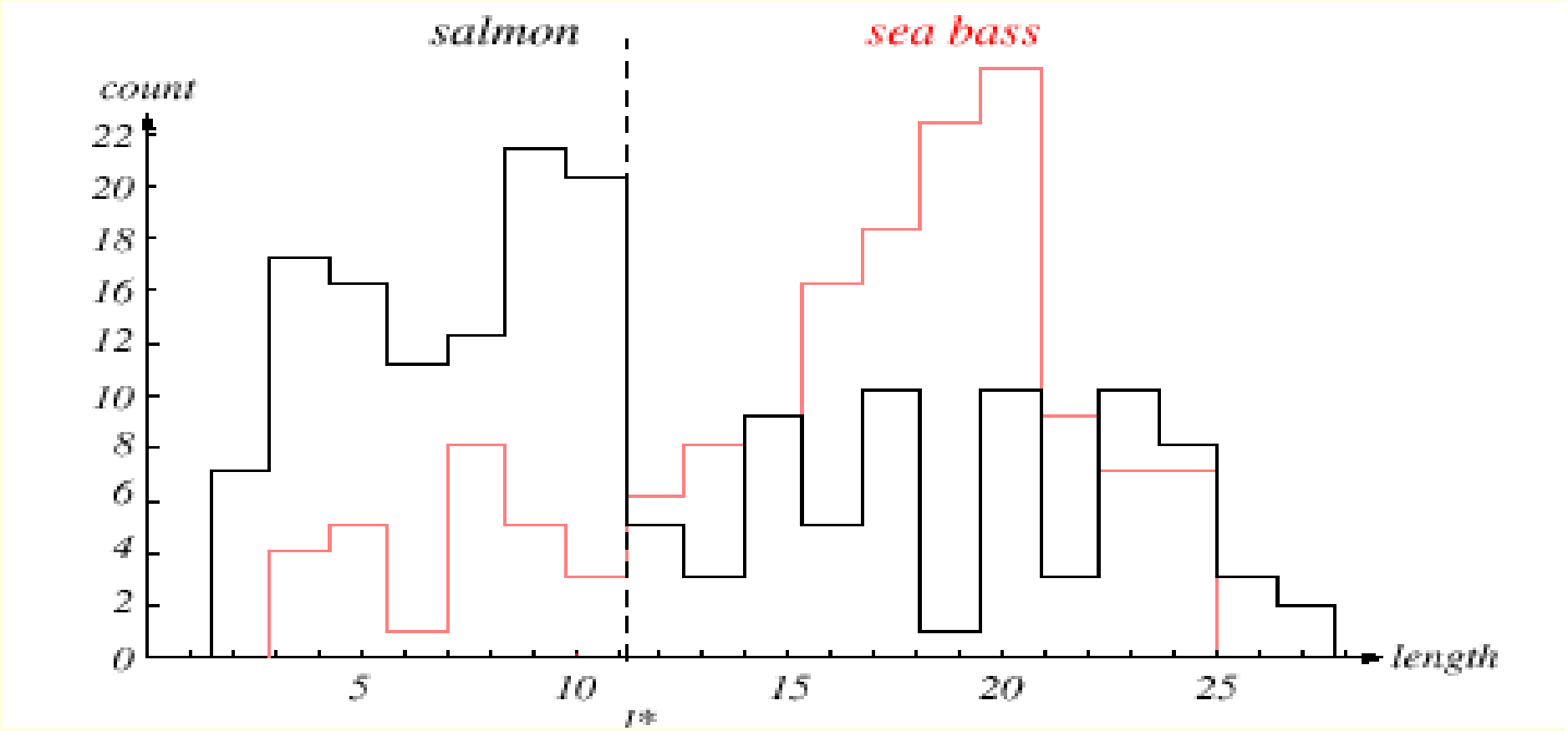
- Preprocessing
  - Use a segmentation operation to isolate fishes from one another and from the background
- Information from a single fish is sent to a feature extractor whose purpose is to reduce the data by measuring certain features
- The features are passed to a classifier




Preprocessing
Feature extraction
Classification
"salmon"
"sea bass"

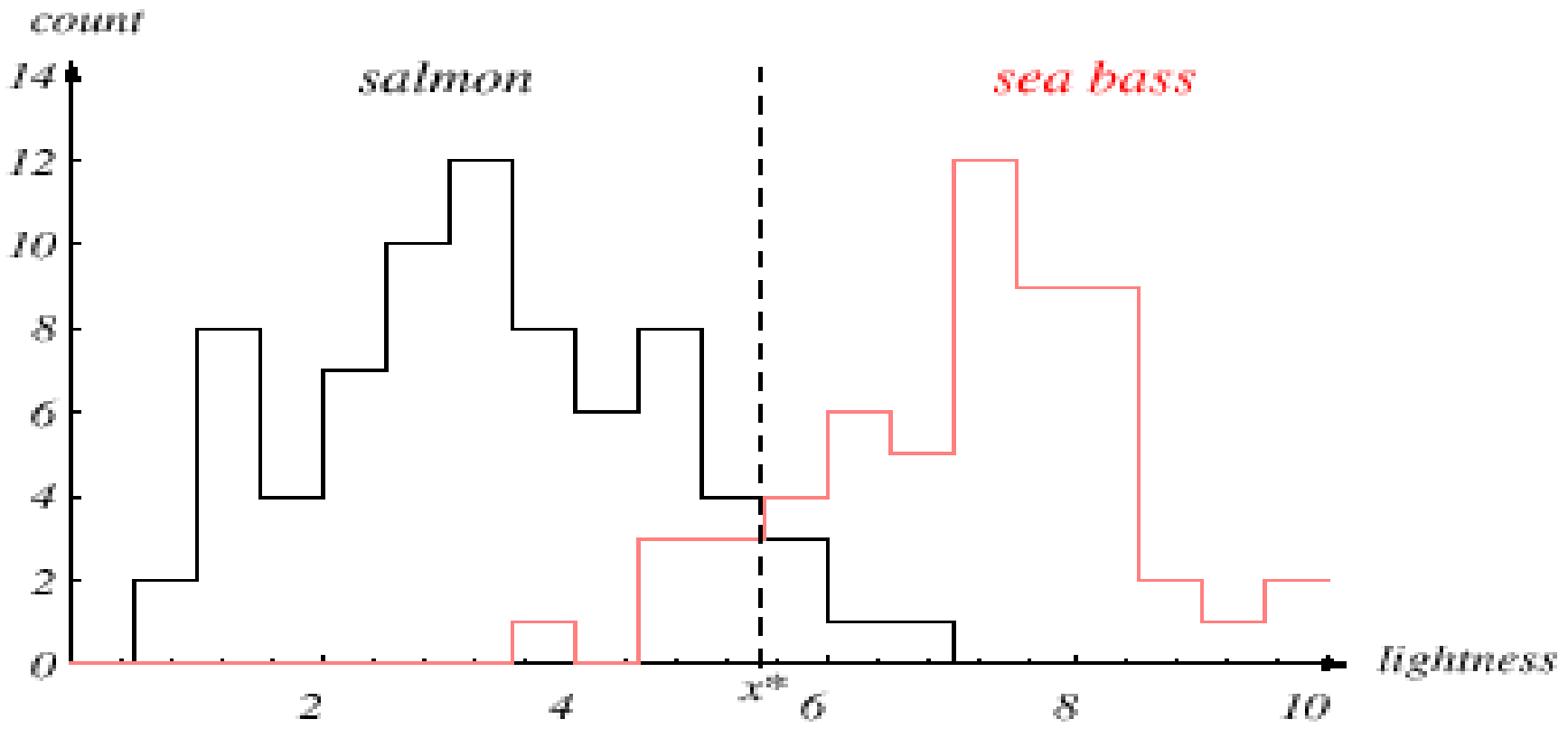
Classification

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*

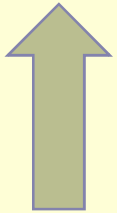
The length is a poor feature alone!

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

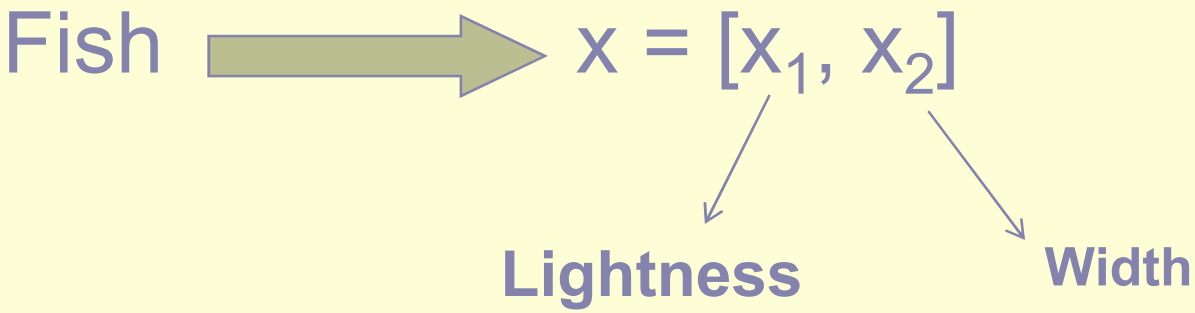
“Customers do not want sea bass in their cans of salmon”

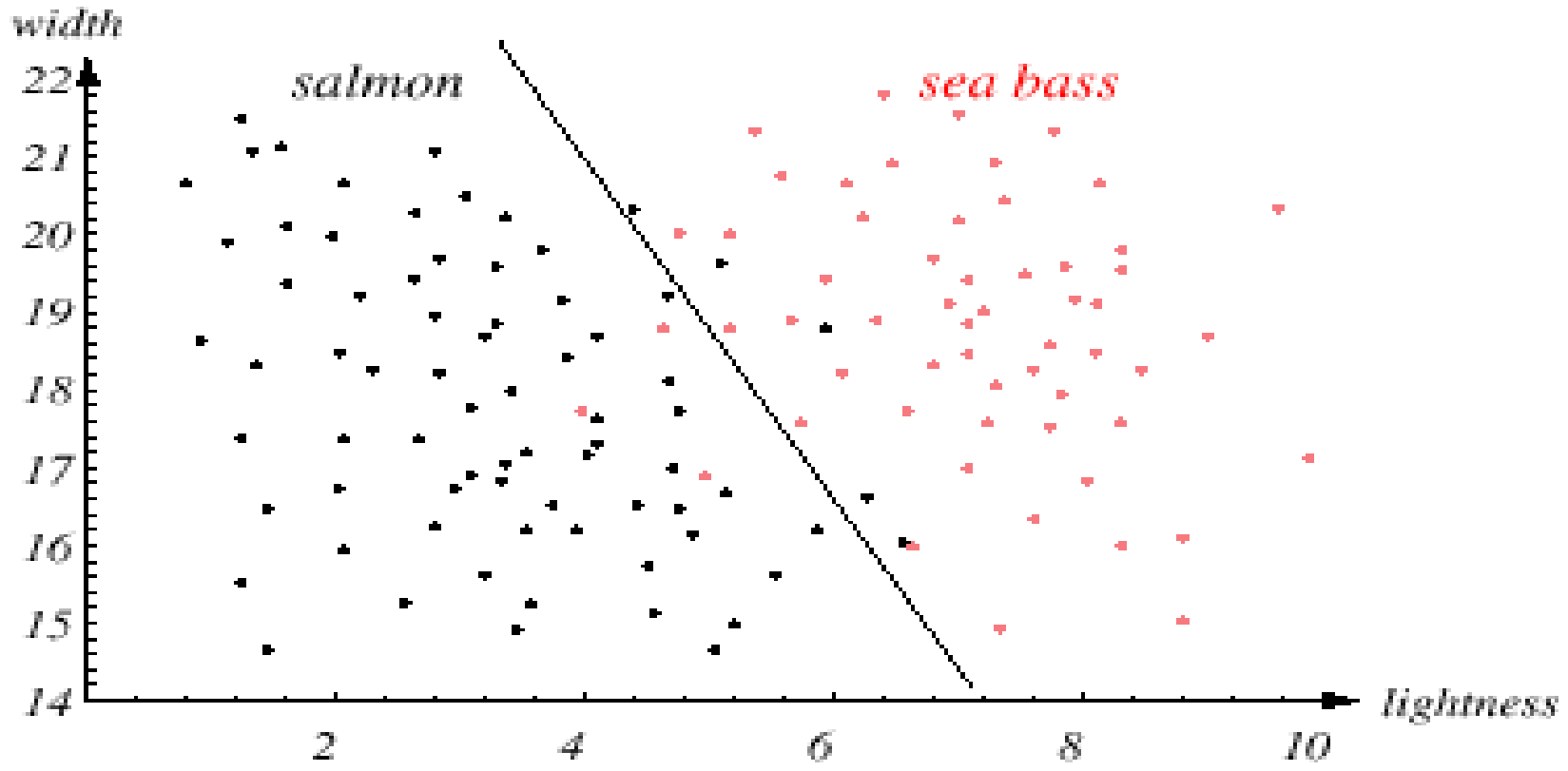
- Threshold decision boundary and cost relationship
- Move our decision boundary toward smaller values of lightness in order to minimize the cost (reduce the number of sea bass that are classified salmon!)



Task of decision theory

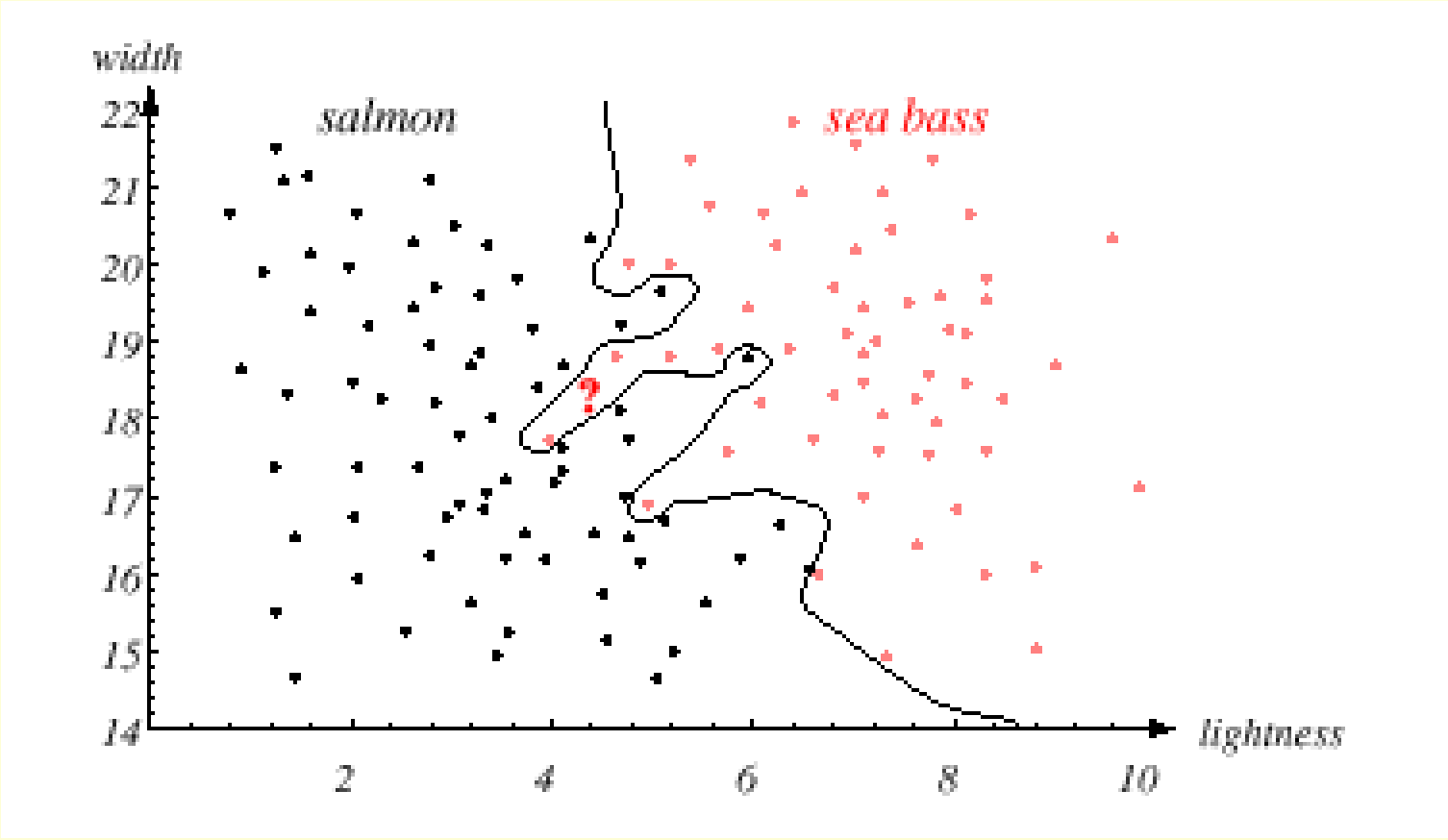
- Adopt the lightness and add the width of the fish



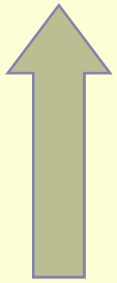

width
salmon
sea bass
22
21
20
19
18
17
16
15
14
2
4
6
8
10
lightness

- 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”

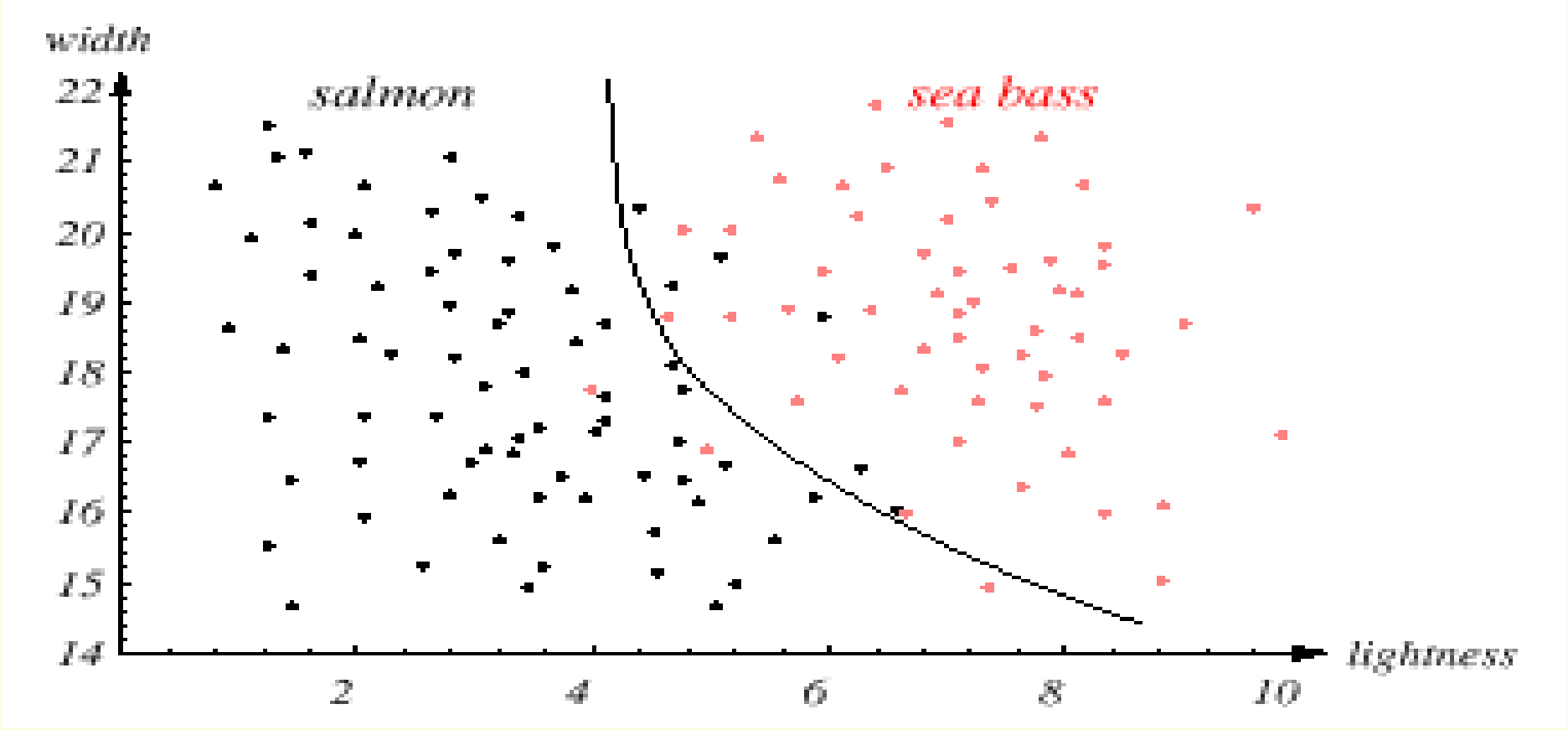
- Ideally, the best decision boundary should be the one which provides an optimal performance such as in the following figure:


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

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

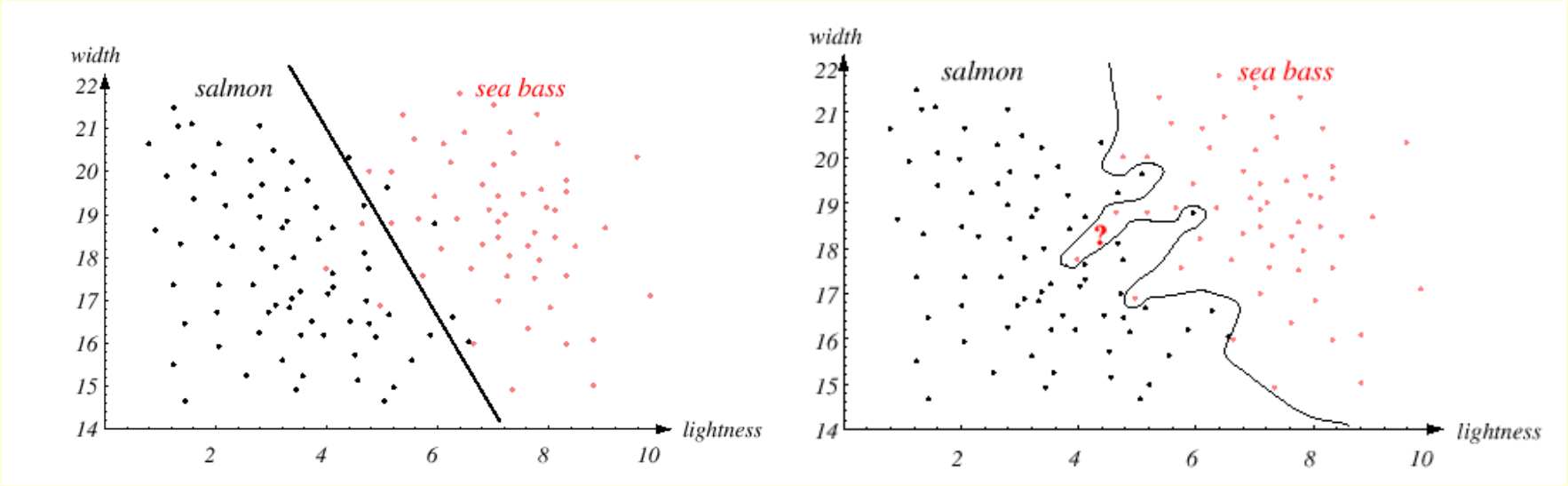


Issue of generalization!

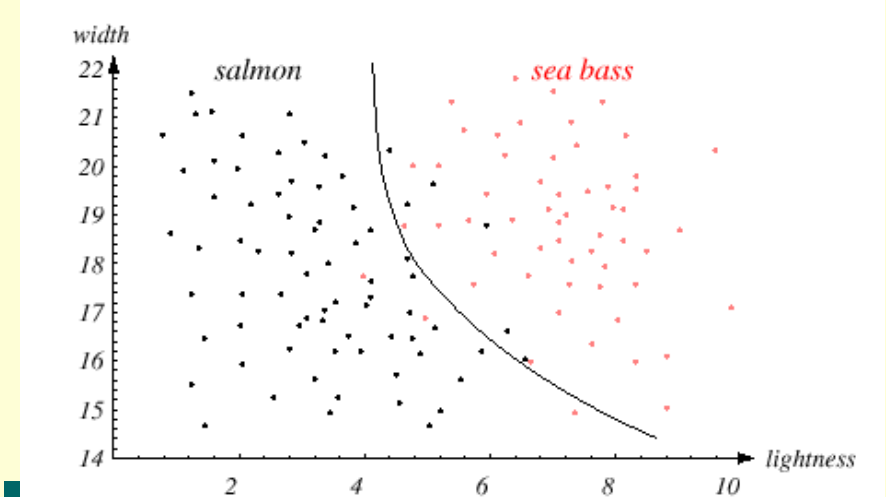
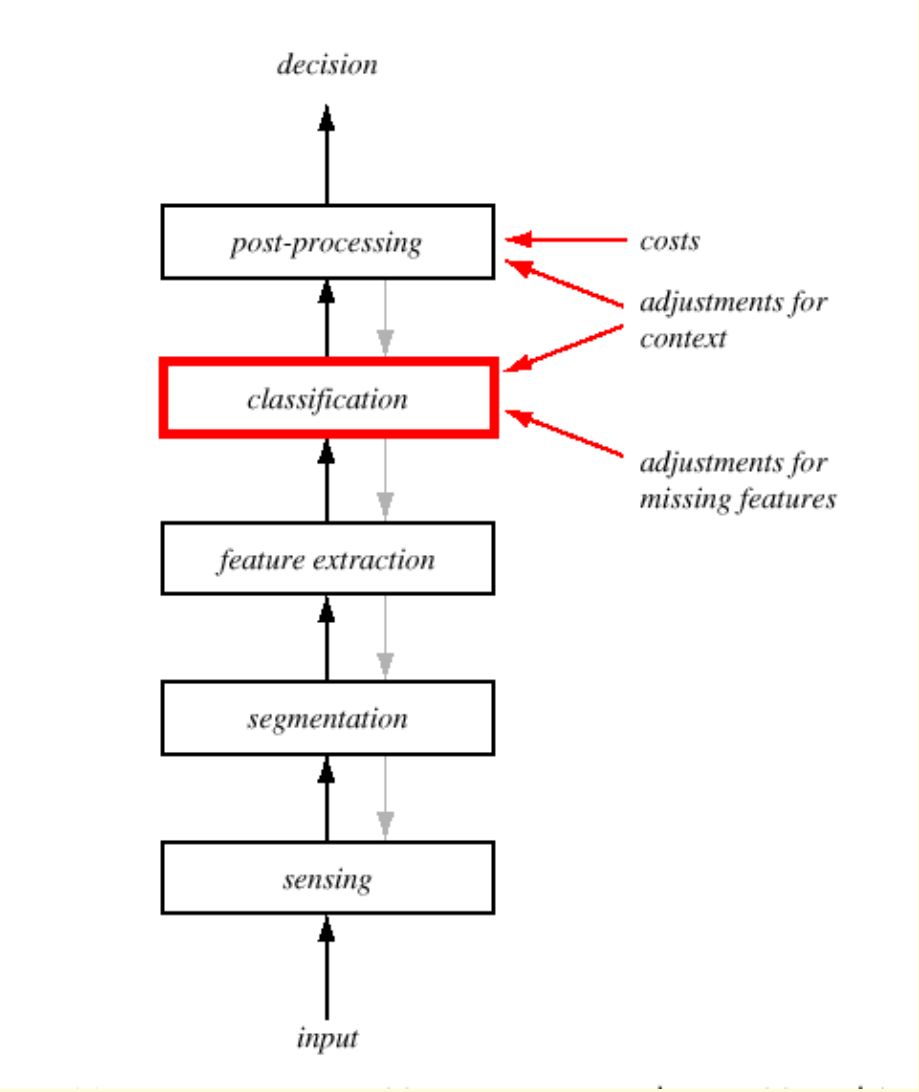

width
salmon
sea bass
22
21
20
19
18
17
16
15
14
2
4
6
8
10
lightness

Decision Boundaries

Observe: Can do much better with two features



Caveat: overfitting!



A Complete PR System