

Pattern Classification

All materials in these slides were taken from **Pattern Classification (2nd ed)** by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000 with the permission of the authors and the publisher

Capitulo 1: Introducción al reconocimiento de patrones (Secciones 1.1-1.6)

Machine Perception

Un Ejemplo

Sistemas de reconocimiento de
patrones

Ciclo de diseño

Adaptacion y Aprendizaje

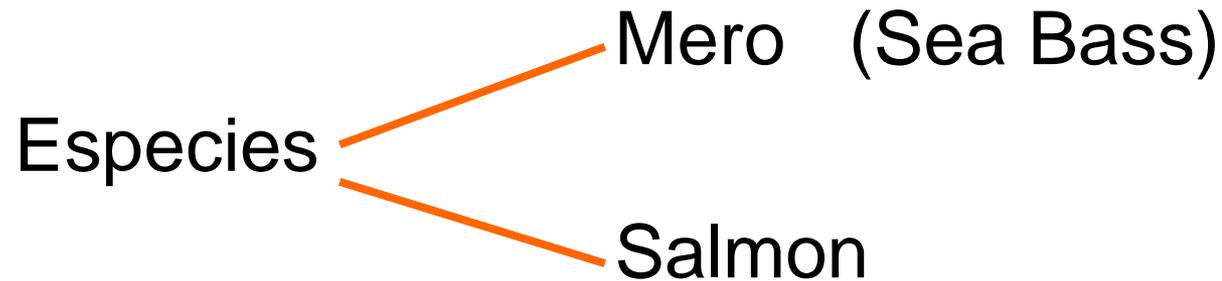
Conclusion

Machine Perception

- Construir una maquina capaz de reconocer patrones:
 - Reconocimiento de frases habladas
 - Identificacion de huellas digitales
 - Reconocimiento optico de escritura
 - Identificacion de secuencias de ADN

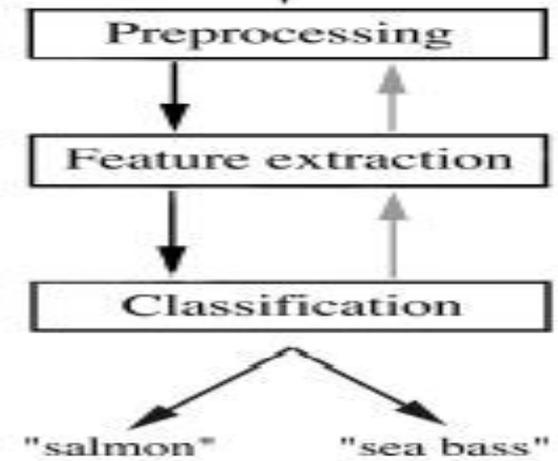
Ejemplo

- “Acomodar el pescado entrante en una cinta de transporte, de acuerdo con la especie, usando sensores ópticos”



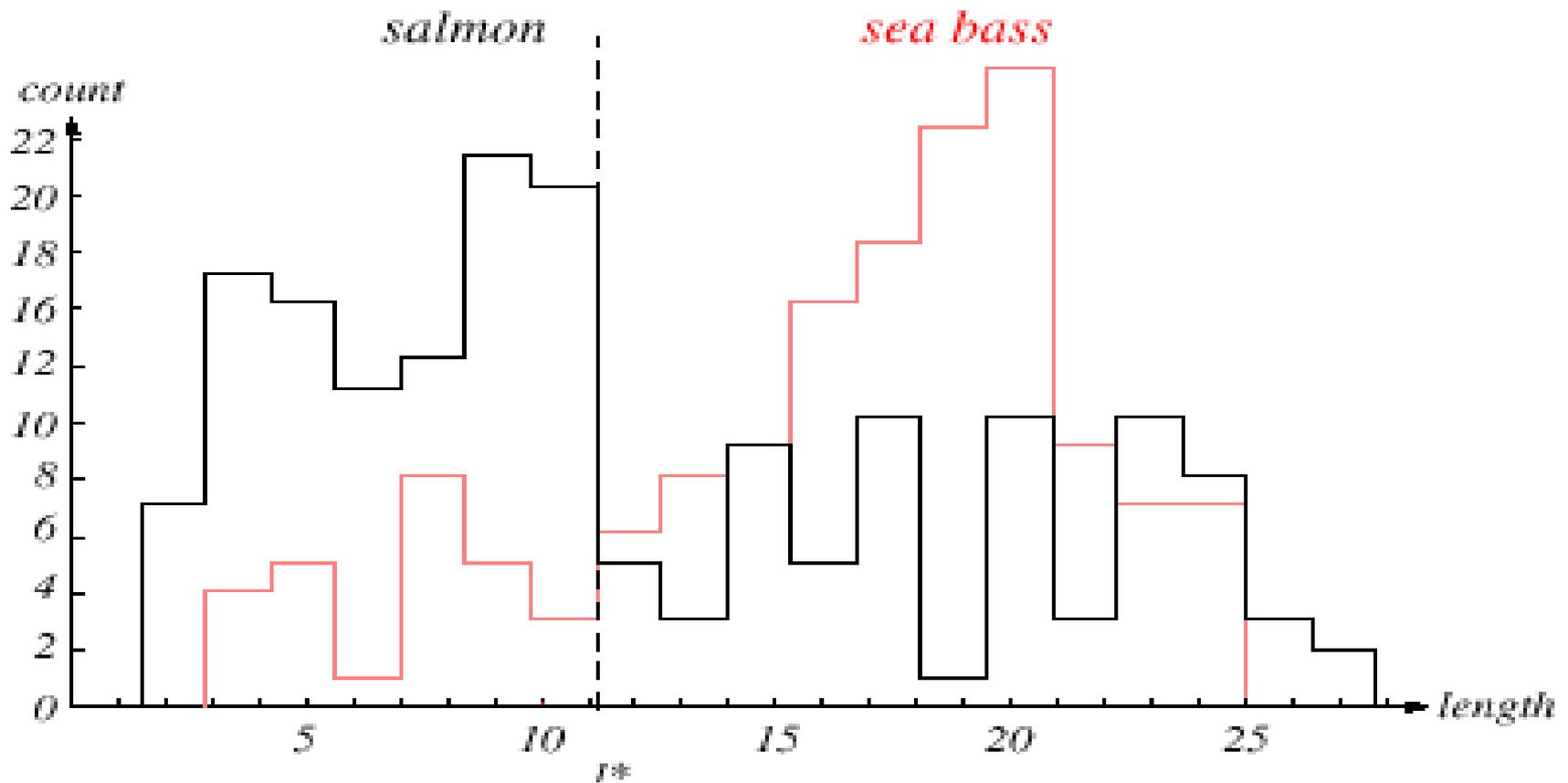
- **Análisis del Problema**
 - Montar una camara y tomar algunas imagenes de muestra para extraer características
 - Largo
 - Tono (color)
 - Tamaño
 - Numero y forma de las aletas
 - Posicion de la boca, etc...
 - Este es el conjunto de las características sugeridas para explorar en el clasificador

- Preprocesamiento
 - Usar segmentacion para aislar los peces uno de otro y con respecto al fondo
- Informacion de un solo pez se envia al extractor de caracteristica cuyo proposito es reducir la informacion midiendo ciertas características
- Las características se pasan al clasificador



- Clasificación
 - Se decide el conjunto de características a usar
 - Conjunto “suficiente”
 - Conjunto “mínimo”
- Seleccionar el largo del pez como una posible característica para discriminación
- Clasificación lineal,

Largo(x) < l*  x=salmon



En promedio son distintos, pero no es suficiente.....

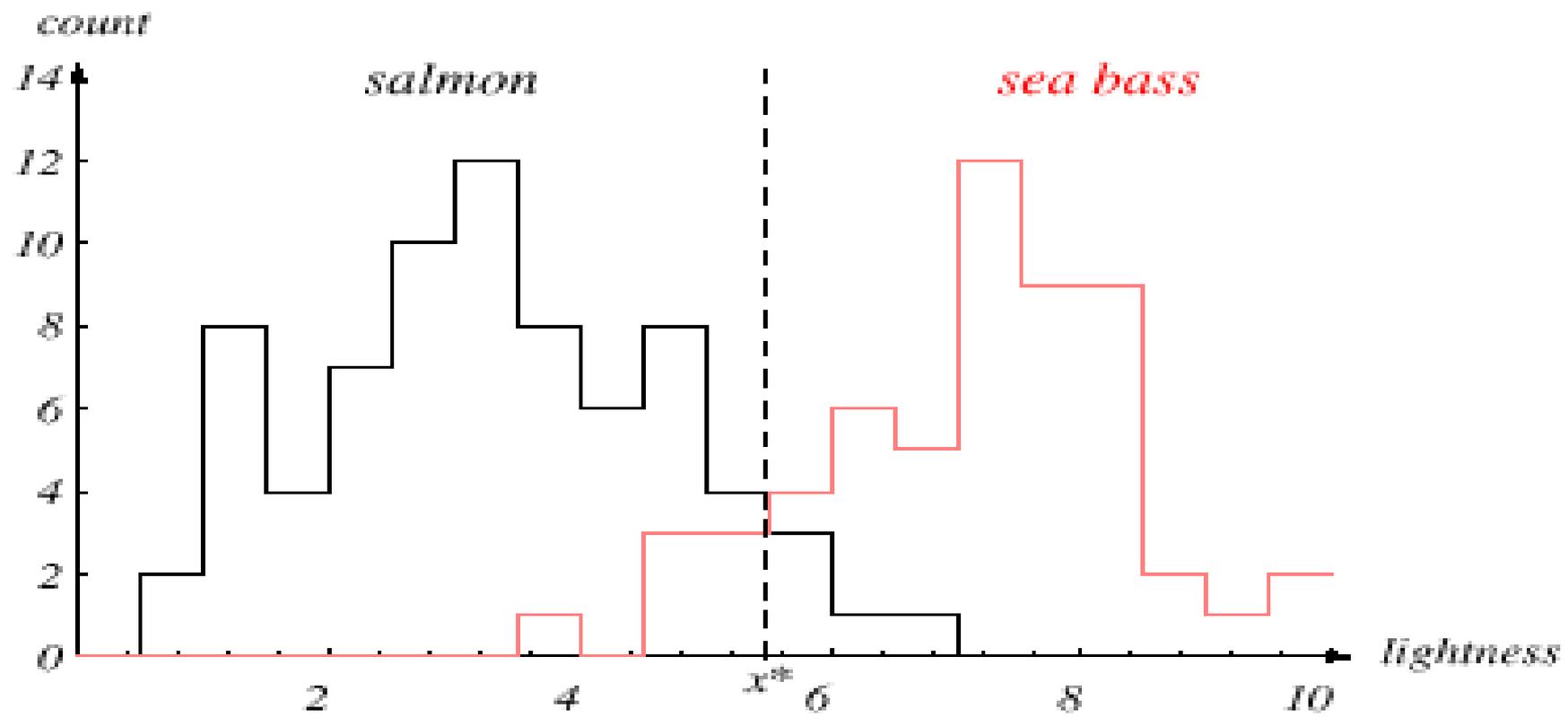
El **largo** es una característica muy pobre!

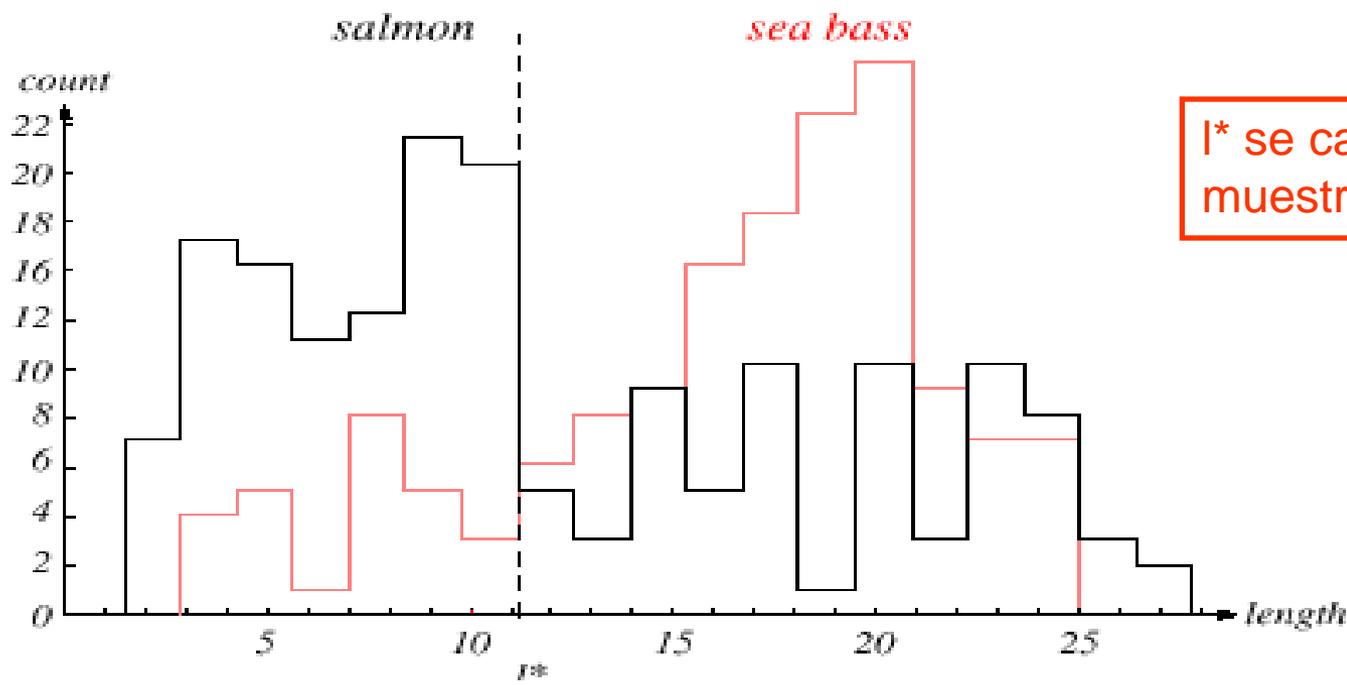
- Largo
- Tono (color)
- Tamaño
- Numero y forma de las aletas
- Posicion de la boca, etc...

Seleccionamos el **tono de las escamas** como una posible característica.

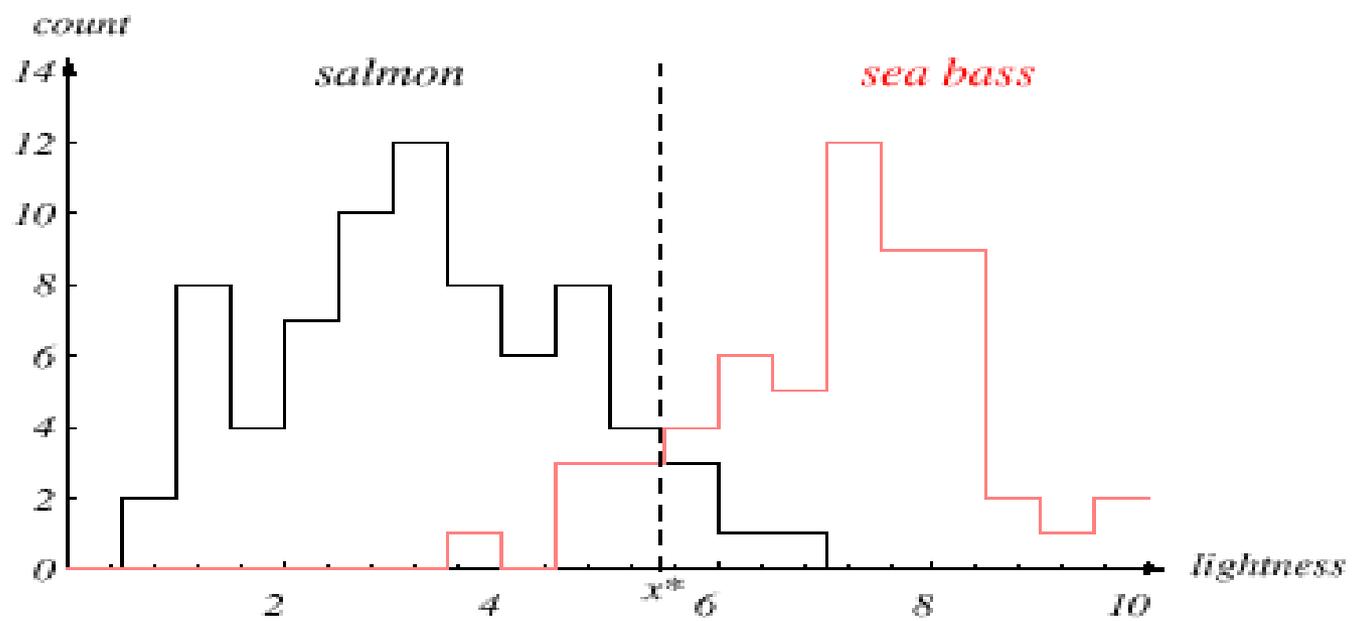
- Clasificación lineal,

Tono(x) $< x^*$  x=salmon

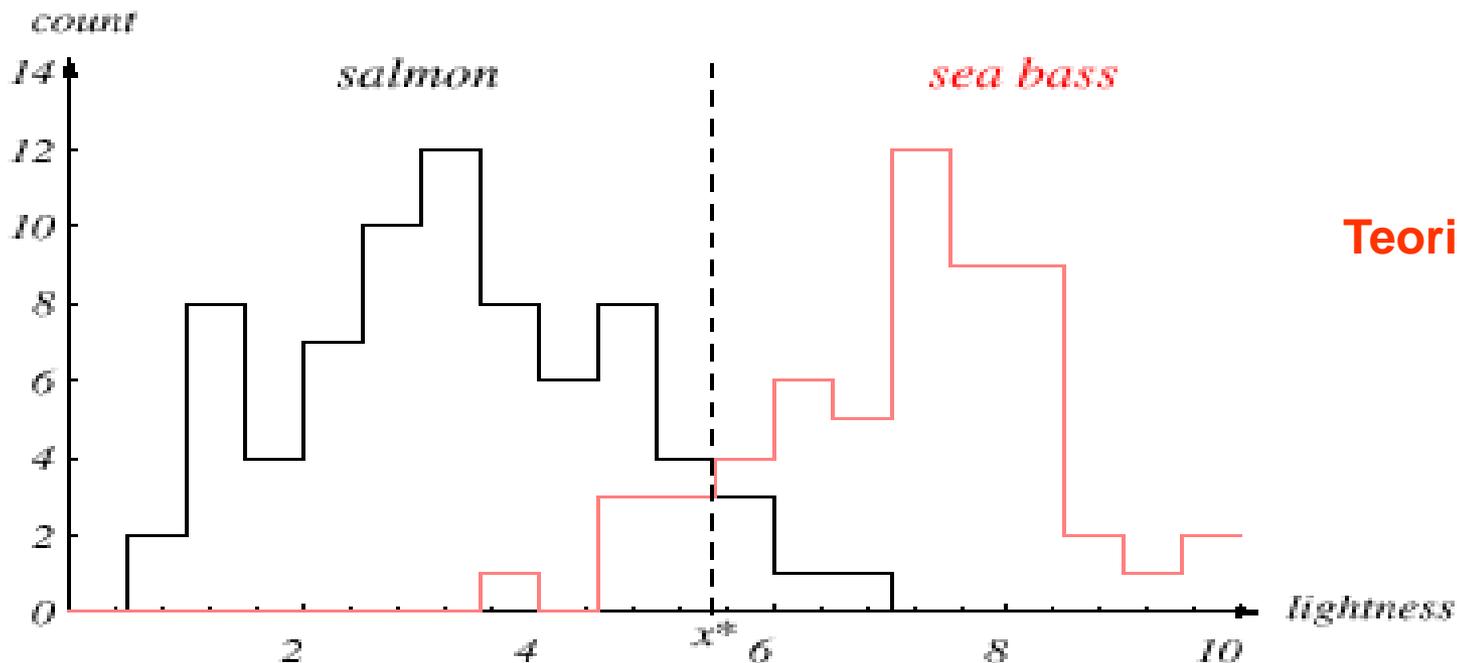




l^* se calcula con muestra de entrenamiento

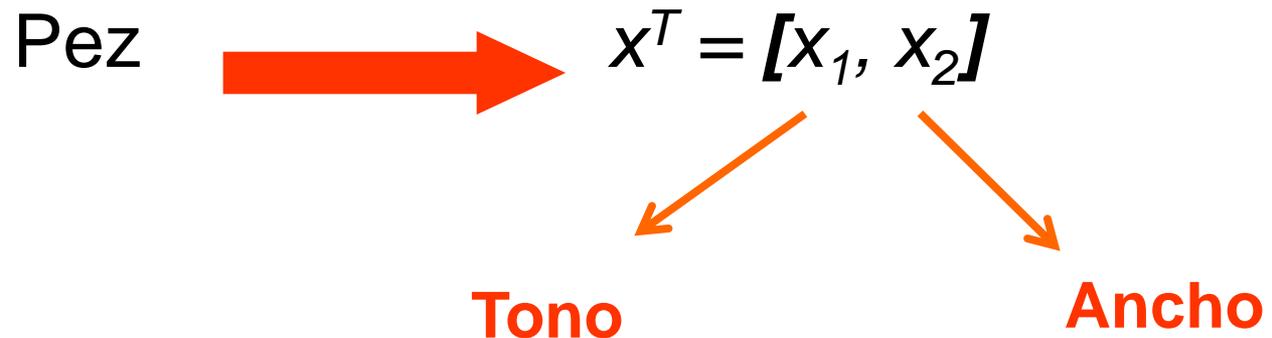


- El umbral de decision x^* y el costo de clasificacion
 - Mover nuestro borde de decision hacia valores menores del Tono para minimizar el costo (reducir el numero de Meros que son clasificados como Salmones !)



Teoria de la decision

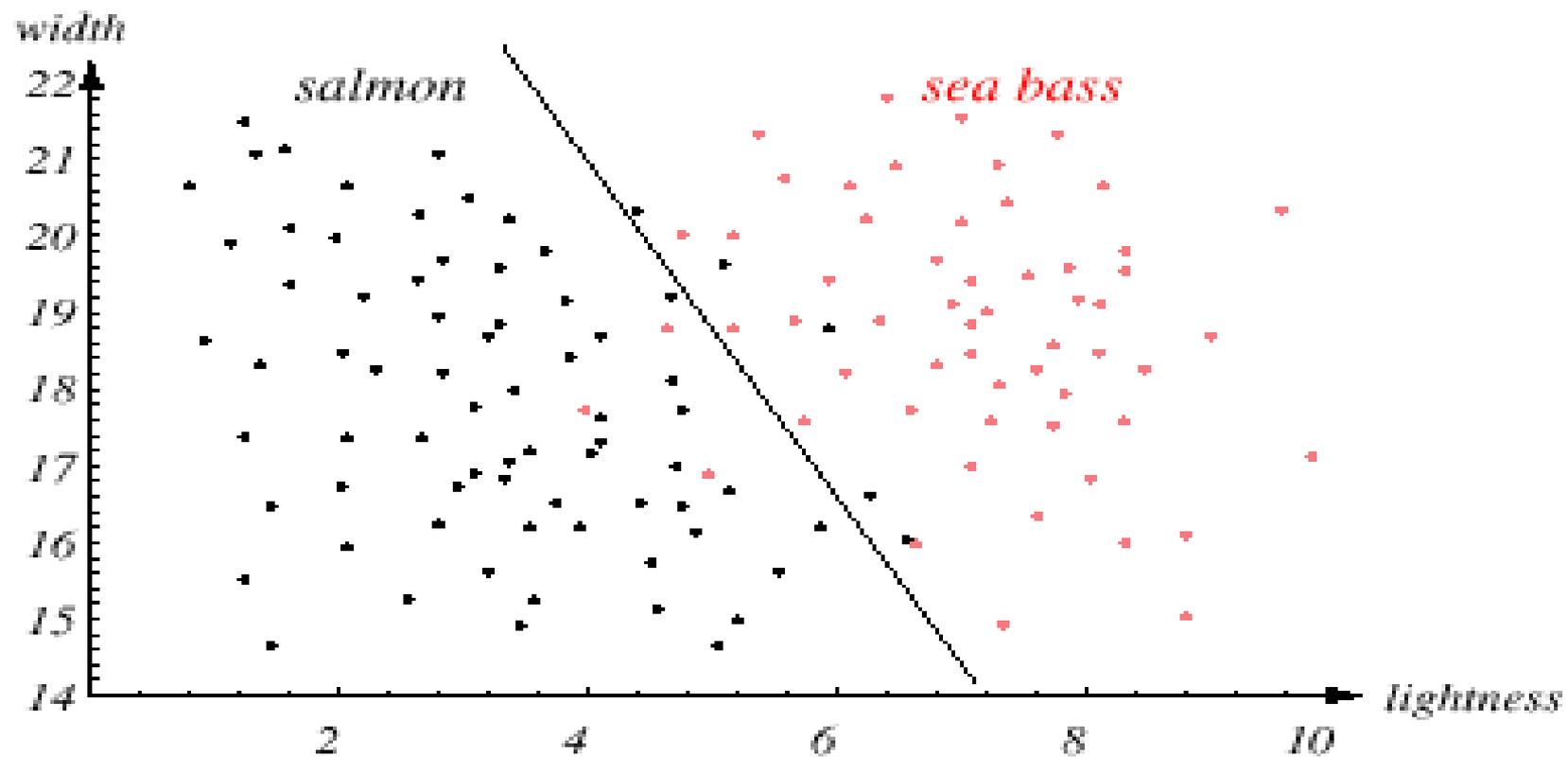
- Todavía hay algunos salmones mal clasificados.
- Adoptar el Tono y agregar el ancho de las escamas del pez



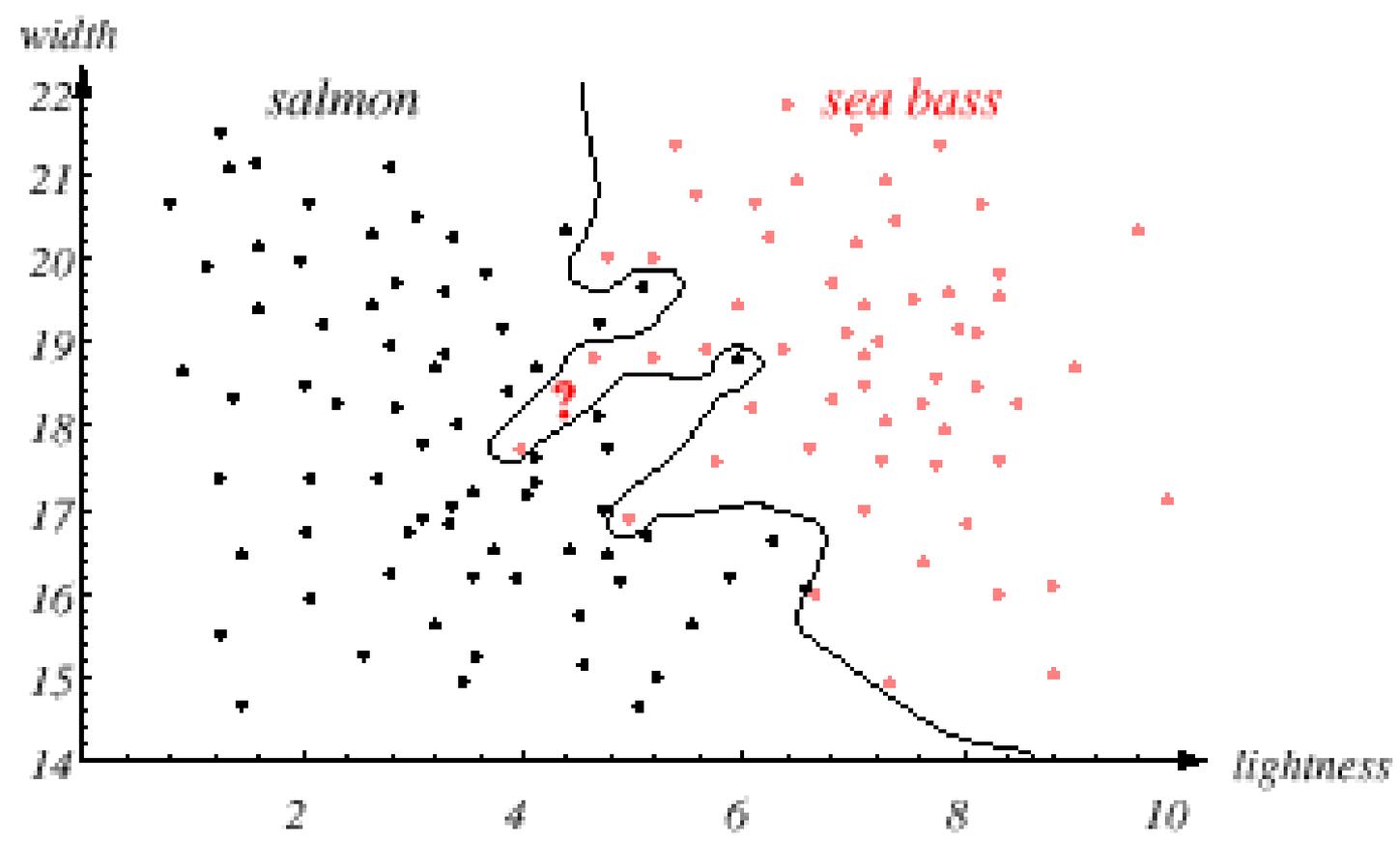
Clasificador: Decision Lineal,

$$x_2 < a x_1 + b,$$

clasifica (x_1, x_2) en salmon



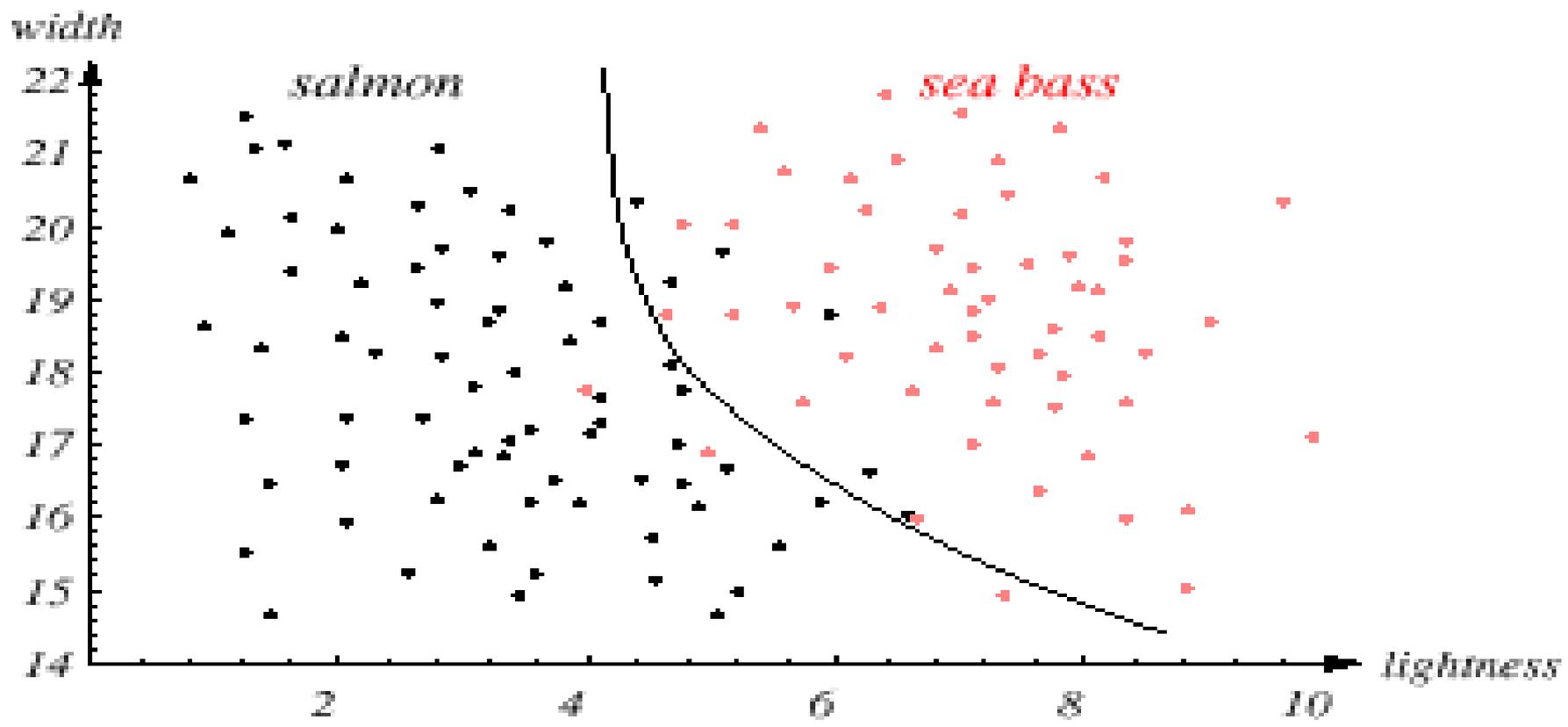
- Podríamos agregar otras características que sean no correlacionadas con las que ya tenemos.
- Hay que tomar precauciones para no reducir el desempeño agregando características “ruidosas”.
- Idealmente, el mejor borde de decisión debería ser el que provee óptimo desempeño como en la figura siguiente :



- Sin embargo, nuestra satisfaccion es prematura porque el objetivo central es designar un clasificador que clasifique correctamente un nuevo elemento

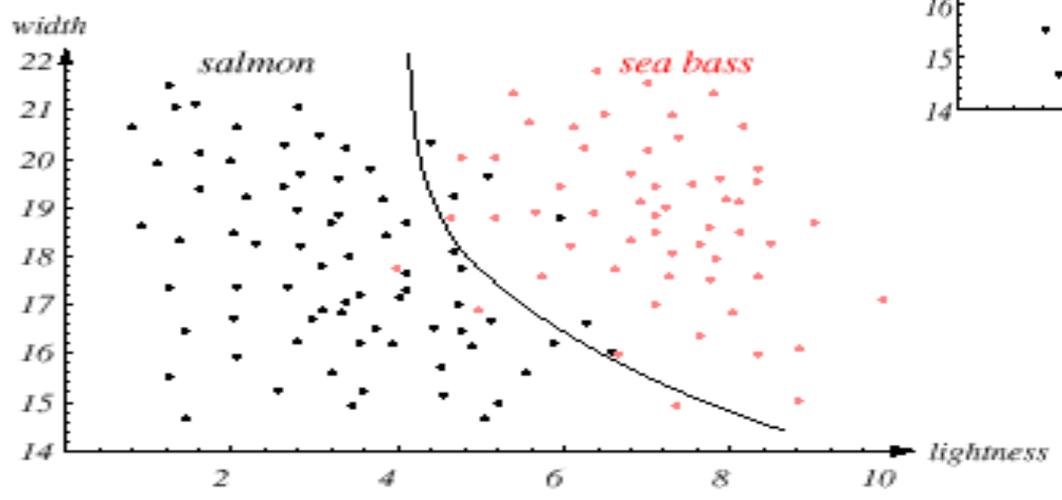
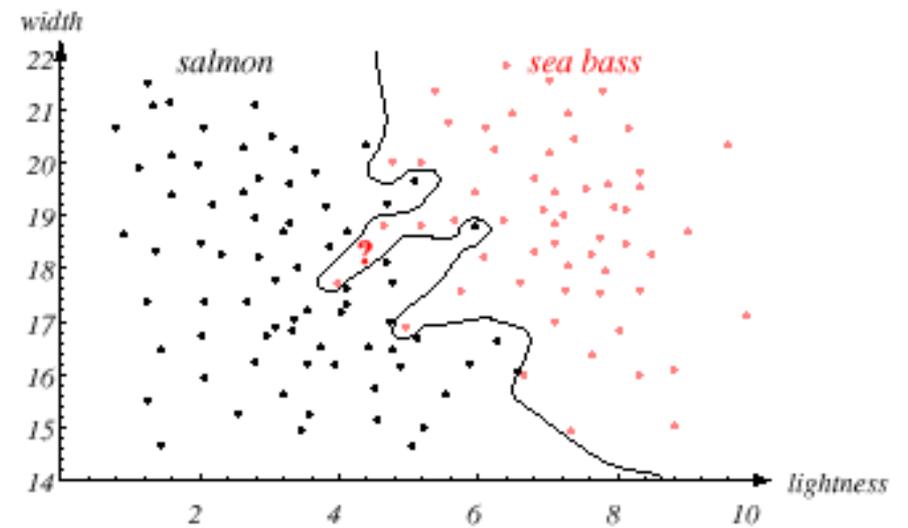
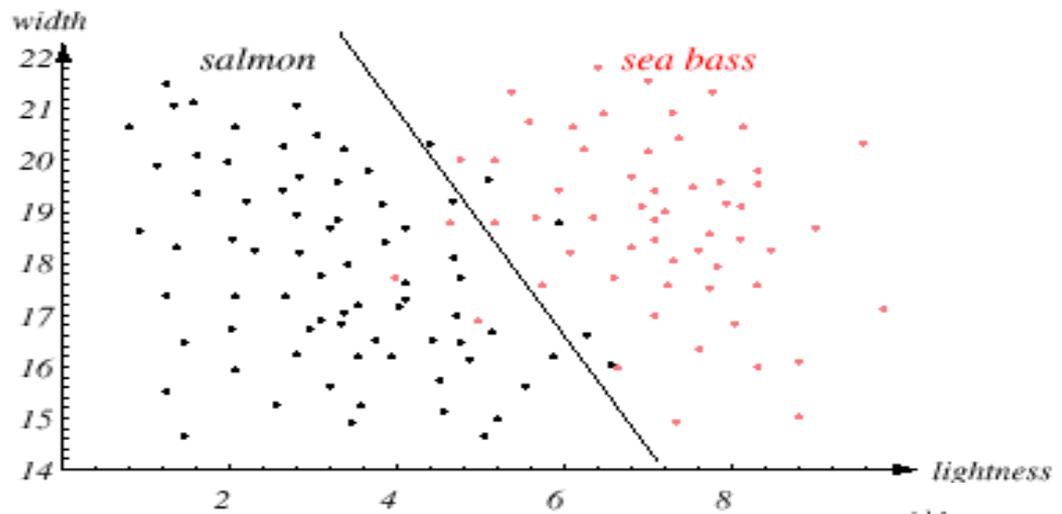


Se debe poder generalizar!



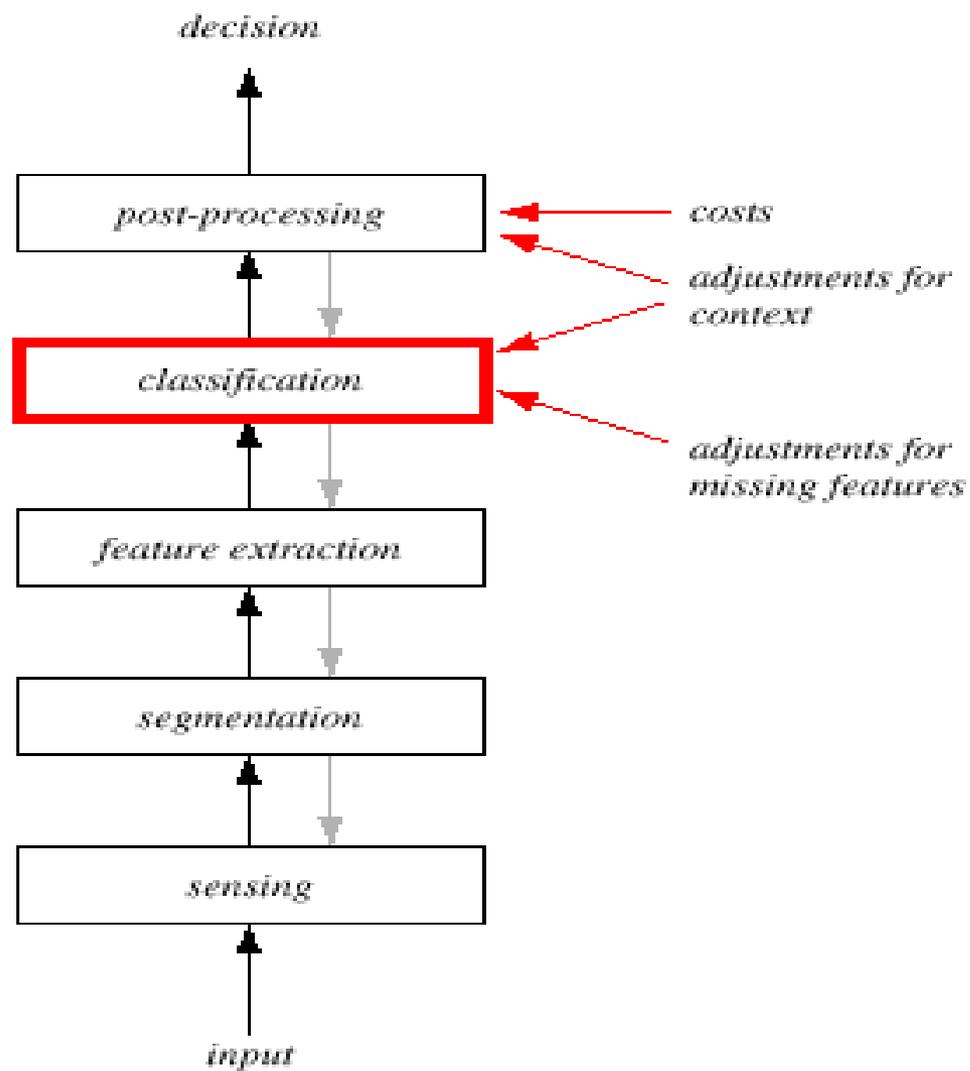
Objetivos del curso

- Lograr realizar en un grupo de datos dado distintos tipos de clasificaciones
 - decidiendo las características a utilizar
 - planteando las ecuaciones
 - implementando en Matlab dichas ecuaciones
 - generando mapas visuales de clasificacion
 - calculando tasas de error aparente y teoricas



Sistemas de reconocimiento de patrones

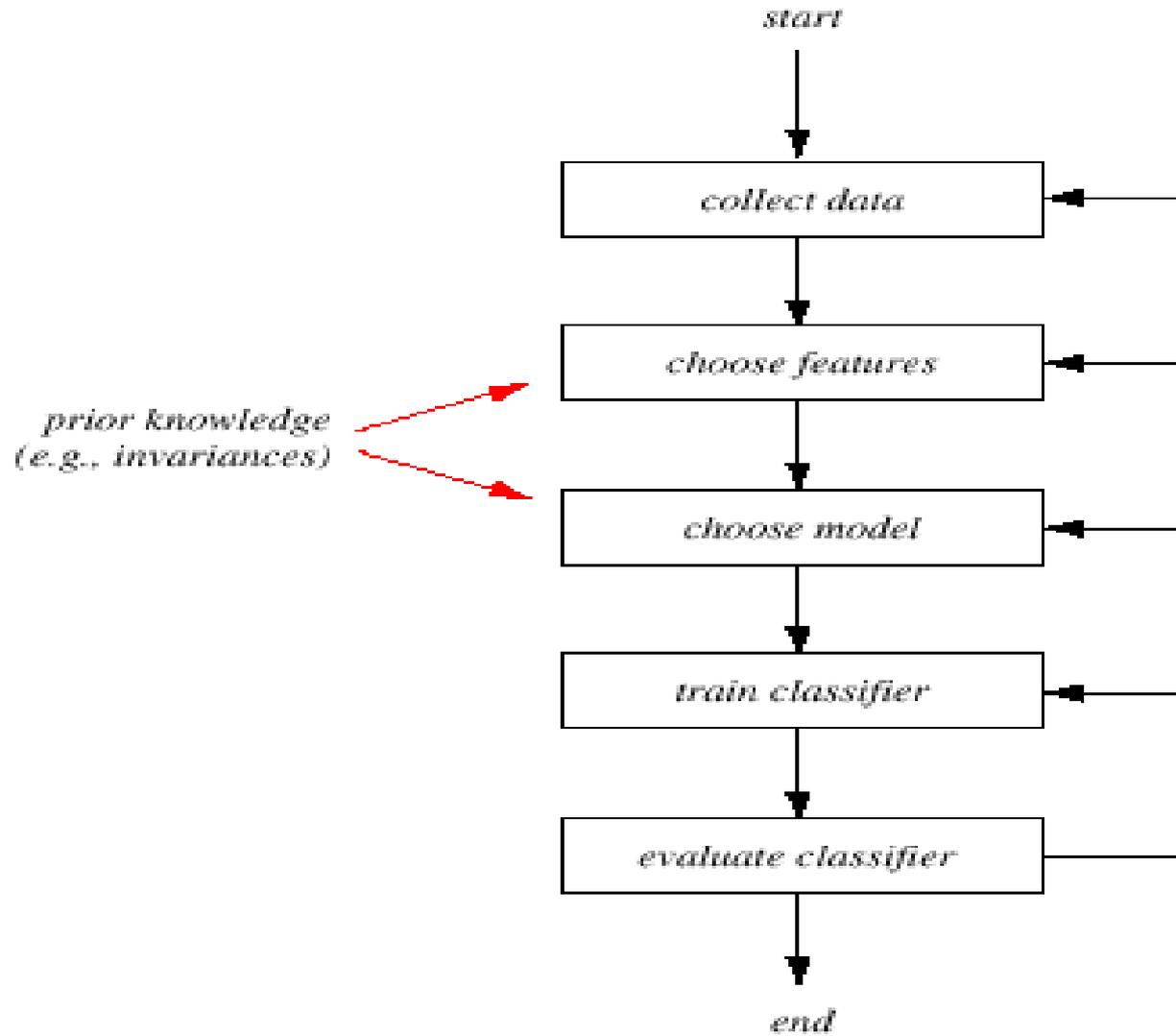
- Captacion
 - Transducer (camera or microfono) para captar los datos
 - Caracteristicas del transducer :
 - ancho de banda,
 - resolucion,
 - sensibilidad a la distorsion
- Segmentacion y agrupamiento de muestras
 - Patrones deberia estar separados y no solaparse
 - Seleccion muestras de entrenamiento



- Extraccion de características
 - Características discriminativas
 - Características invariantes con respecto a la traslacion, rotacion y escala.
- Clasificacion
 - Regla que usa el vector de características provisto para asignar el objeto a una categoria
- Post Procesamiento
 - Explotar la informacion de **contexto** del objeto para mejorar el desempeño

El Ciclo de Diseño

- Recoleccion de datos
- Eleccion de características
- Eleccion de Modelo
- Entrenamiento
- Evaluacion
- Reajuste
 - complejidad computacional.
 - errores aceptables e inaceptables



Recoleccion de Datos

- Como sabemos cuando hemos recolectado una coleccion adecuada, en tamaño y representatividad, de ejemplos para entrenar y testear el sistema?

Eleccion de características

- Dependen del problema de clasificacion, del sistema de vision que genera los datos, del costo de obtencion de los datos.
- Deben ser simples de extraer, invariantes bajo transformaciones de la imagen e inmunes o resistentes al ruido de captacion de la imagen.

Eleccion de Modelo Matematico

- Lineal
- Cuadratico
- Funcional
- Arbol
- Red neuronal
- Maquina de soporte vectorial
- etc

Entrenamiento

- Usamos datos para determinar los clasificadores.
- Muchos procedimientos diferentes para entrenar los clasificadores y elegir los modelos

Evaluacion

- Medidas de Desempeño bajo un modelo
 - Tasas de error
 - Costos
- Medidas de desempeño al cambiar de un grupo de características a otro

Complejidad computacional

- Cual es el balance entre computacion facil y desempeño?
- Como escala un algoritmo en funcion del numero de características, patrones y categorías?

Adaptacion y Aprendizaje

- Aprendizaje supervisado
 - Un oraculo provee las etiquetas de categoria y costo de cada patron en el conjunto de entrenamiento
- Aprendizaje no supervisado
 - El sistema forma grupos naturales de los patrones ingresados .

Clasificación en Imágenes

- Etiquetar toda una imagen como una clase
 - Imagen de interiores o de exteriores;
 - Imagen de Salmon o Imagen de Mero
- Dividir una imagen en regiones de diferente tipo.
 - Diferentes tipos de tejido en una imagen médica,
 - Diferentes ejemplares de peces en la cinta de transporte.

El segundo tipo de clasificación se denomina
segmentación

- Nombres distintos porque son problemas distintos
- Técnicas parecidas, podemos usar las mismas técnicas en ambos problemas.

Clasificación



Clase 1



Clase 2



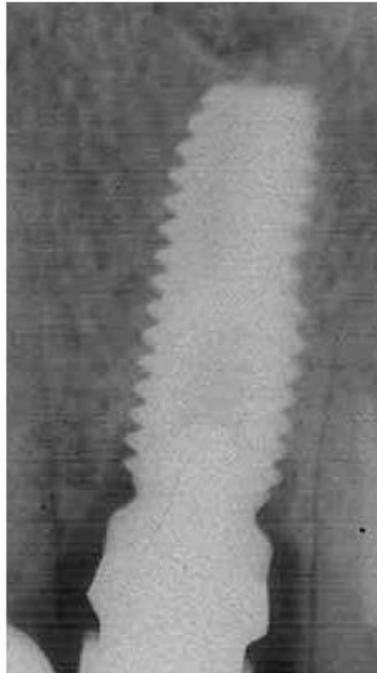
Clase 3



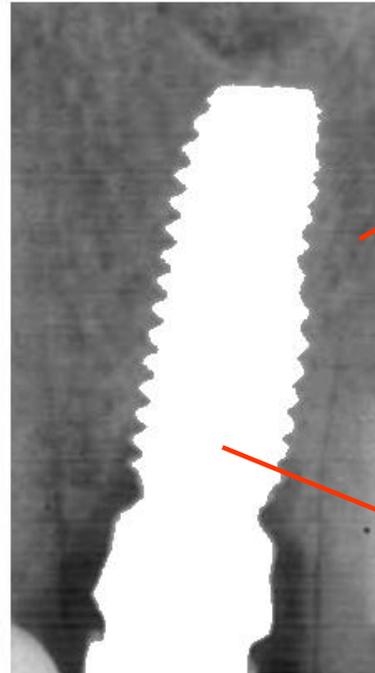
?

Segmentacion

Original



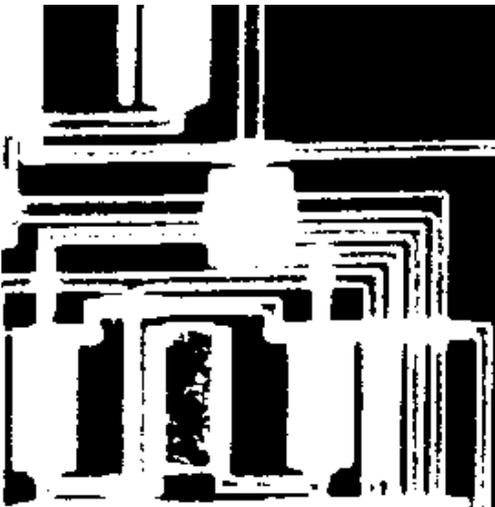
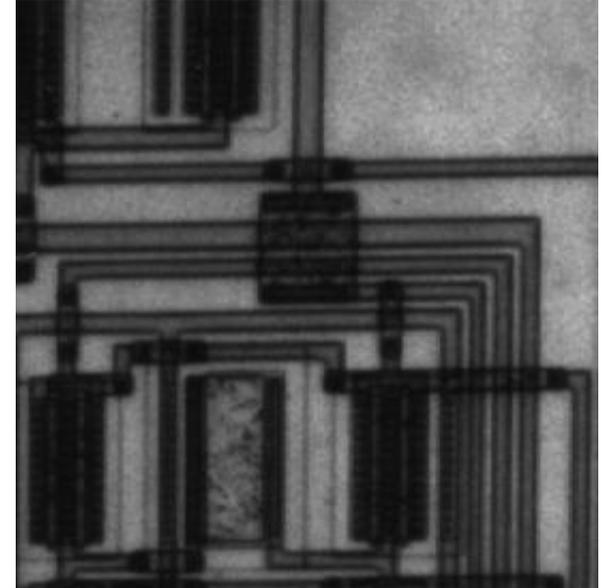
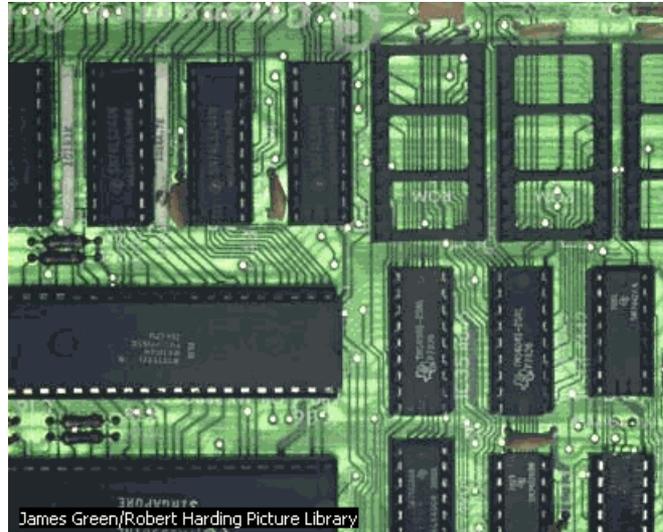
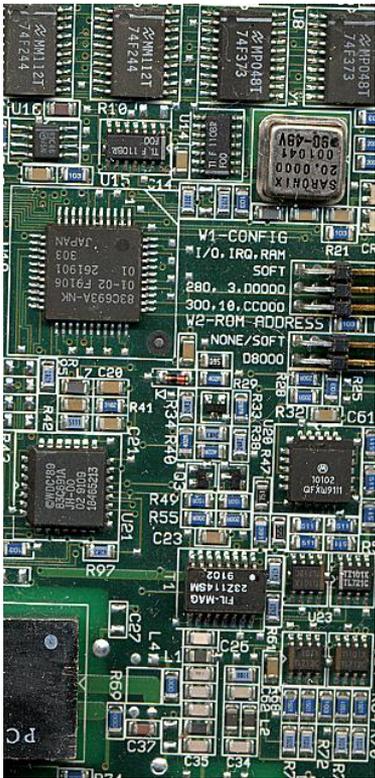
Segmentada



fondo

perno

Problema actual



Anotacion de características
 Palabras claves
 Clasifico una imagen como correspondiente a
 una bolsa de palabras

Ejercicios

- Estudiar el problema de clasificación de flores Iris de Matlab
 - Características decididas por el biólogo que recolecto los datos
 - numero de clases hipotéticas
 - cluster analysis
 - classification
- Estudiar cargado de datos y muestra de datos en pantalla según características elegidas.