

Aprendizaje automático

Introducción

Varios

- **Programa del curso**
- **Herramientas y lenguaje de programación**
 - } Python con Pytorch y otras bibliotecas
 - } Anaconda
 - } <https://colab.research.google.com/>
- **Administrativos:**
 - Contacto: maria.mora@itcr.ac.cr
 - Asistente: Ricardo Víquez (rychiold@gmail.com)
 - TecDigital
 - No se reciben tareas por otro medio
 - El asistente tiene 10 días para entregar resultados de una tarea.
 - Sitio alternativo para materiales (por si hay problemas con el TecDigital)
 - https://drive.google.com/drive/folders/1Fmi7Zin_0GjXpNIGnDLvkQGM5MWUDgMe?usp=sharing

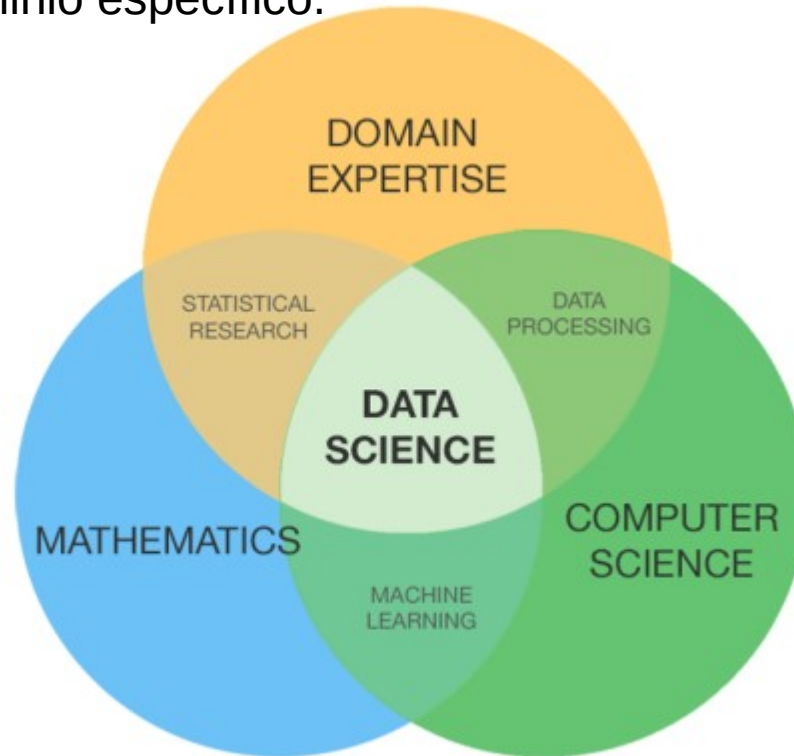
Contenidos

- Introducción
- Formas de aprendizaje
- Flujo de trabajo



Aprendizaje automático (ML)

- Parte fundamental de las Ciencia de Datos
- Mezcla de aplicación de conocimientos en matemática, estadística, y computación, usando grandes cantidades de datos, en un dominio específico.



Source: Palmer, Shelly. Data Science for the C-Suite.
New York: Digital Living Press, 2015. Print.



Aprendizaje automático (ML)

Definición

Aprendizaje automático es el estudio de algoritmos que:

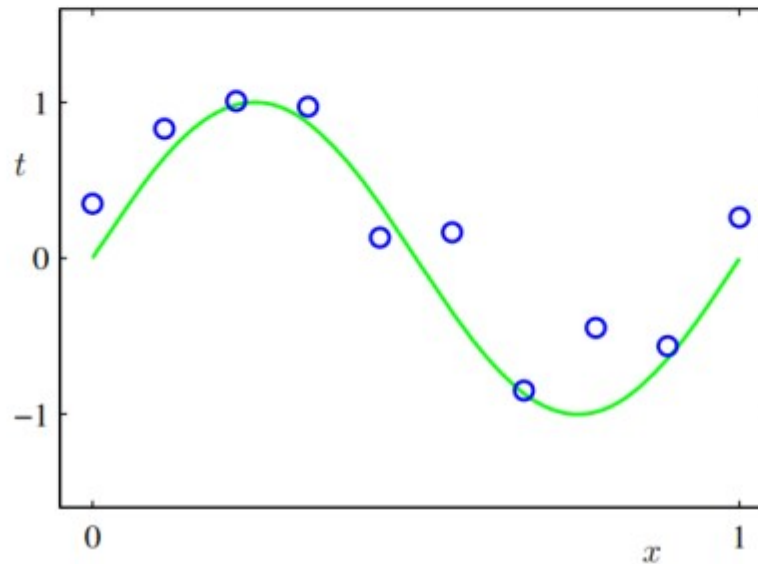
- Mejoran su desempeño P
- Al realizar una tarea T
- La mejora se basa en experiencia E

Una tarea de aprendizaje bien definida viene dada por $\langle P, T, E \rangle$

Tom Mitchell (1998)

Tipos de aprendizaje automático (supervisado)

- **Clasificación:** Proceso de asignar una muestra \vec{x}_i una etiqueta j , y en general para un conjunto de muestras \mathbf{X} , obtener un conjunto de etiquetas \mathbf{T}
- Si todas las etiquetas $j \in \mathbb{N}$ se realiza una **clasificación** y si $j \in \mathbb{R}$ se realiza una **regresión**



Experiencia: Conjunto de muestras

- **Conjunto de muestras X :** Conjunto de M muestras de un arreglo de características:

$$X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M\}$$

- Por ejemplo, ¿cual es el patrón para las muestras de la verdosidad ($\dim(\vec{x}_i) = 1 = N$)?:

$$X = \{\vec{x}_1 = 253, \vec{x}_2 = 254, \vec{x}_3 = 100, \vec{x}_4 = 255\}$$



Verdad conocida (*target*)

- **Conjunto de etiquetas T para un conjunto de muestras X :**

Para un conjunto de M muestras

$$X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M\}$$

existe un conjunto de etiquetas «correctas»:

$$T = \{t_1, t_2, \dots, t_M\}$$

Clasificación: las clases se codifican

- **Clase:** Para facilitar su representación una clase C_j tiene asociada una etiqueta $j = 1, 2, \dots, J$
- **Ejemplo 1:** la clase «banano» tiene etiqueta $j = 1$, la clase «sandía» la etiqueta $j = 2$
- **Ejemplo 2:** la clase «uva verde» $j = 1$, «uva morada» $j = 2$ y «uva roja» $j = 3$



Ejemplo: Un clasificador estadístico simple

Ejemplo

Clasificador de uvas verdes, moradas y rojas

- Se define el conjunto de **muestras de entrenamiento (ground truth)** X_e con sus correspondientes etiquetas T_e

$$X_e = \{\vec{x}_1 = 253, \vec{x}_2 = 254, \vec{x}_3 = 100, \vec{x}_4 = 255, \vec{x}_5 = 105\}$$

$$T_e = \{t_1 = 1, t_2 = 1, t_3 = 2, t_4 = 1, t_5 = 2\}$$

- Momentos estadísticos para cada clase:

$$\begin{aligned} \mu_1 &= 254 & \sigma_1 &= 1 \\ \mu_2 &= 102,5 & \sigma_2 &= 3,53 \end{aligned}$$

Media y desviación estándar

- Si se recibe una nueva muestra, p. ej. con valor $\vec{x}_6 = 252$, el clasificador hace para todas las clases C_j :

$$\begin{aligned} |\mu_1 - \vec{x}_6| &= 2 < 3\sigma_1 \\ |\mu_2 - \vec{x}_6| &= 149,5 > 3\sigma_2 \end{aligned} \Rightarrow t_6 = 1$$



Etapas básicas de un sistema de aprendizaje automático



1. Pre-procesamiento

- Modificación de datos de entrada para facilitar el aprendizaje (Transformaciones)
 - Estandarización y escalamiento
 - Procesamiento de datos faltantes
 - Eliminación de sesgos
 - Redundancia
 - Ruido
 - Discretización

1. Pre-procesamiento

El preprocesamiento consiste en aplicar a un conjunto de datos, representado matricialmente como $X \in \mathbb{R}^{n \times m}$, una transformación:

$$X_p = T(X)$$

- La cual modifique la información en cada observación o muestra $\vec{x}_i \in \mathbb{R}^n$
- y/o modifique la cantidad de muestras m con el fin de mejorar el desempeño de etapas posteriores.

1. Pre-procesamiento

Tipos básicos de atributos:

Atributos numéricos: Los atributos numéricos son atributos que toman una serie de valores discretos o continuos.

- Por ejemplo, atributos como temperatura, humedad o fecha pueden tomar infinidad de valores, por lo que son entonces **continuos**.
- Atributos numéricos discretos son aquellos que toman **valores naturales**, como por ejemplo cantidad_clientes.
- Las operaciones matemáticas están definidas en estos atributos, tal como la moda, media y mediana y las distancias.

1. Pre-procesamiento

Tipos básicos de atributos (continuación):

Atributos ordinales: atributos con un conjunto de valores que pueden ser ordenados, aunque las magnitudes de tales valores no sea conocida.

- Por ejemplo, **satisfaccion_usuario** el cual puede tomar los valores muy_satisfecho, poco_satisfecho, o insatisfecho, a los cuales se les pueden asignar los valores 3, 2 y 1 respectivamente.
- Pueden derivarse a partir de un atributo numérico, partiendo el dominio de tal atributo en intervalos, como por ejemplo el atributo **temperatura** puede tomar los valores 0-15, 16-25, 25-50.
- Para los atributos ordinales se puede calcular la moda (valor más frecuente), o mediana, con significado matemático, no así la media.

1. Pre-procesamiento

Tipos básicos de atributos (continuación):

Atributos binarios: Un atributo binario es un atributo cualitativo que puede tomar únicamente dos valores (Booleano), usualmente codificados como 0 o 1.

- Un ejemplo de atributo binario es **identificacion_nacional** el cual es 0 si la identificación es extranjera, y 1 de lo contrario.

1. Pre-procesamiento

Tipos básicos de atributos (continuación):

Atributos binarios. Existen dos tipos de atributos binarios:

- **Atributos binarios simétricos**: son atributos cuyos valores presentan usualmente igual probabilidad. Por ejemplo el atributo de **sexo** puede ser masculino o femenino con igual probabilidad usualmente.
- **Atributos binarios asimétricos**: se refiere así a los atributos cuyos valores pueden presentar probabilidades distintas.
- Ejemplo, pruebas de laboratorio
- Convención: asignar 1 al mas importante (p.e., VIH positivo)




1. Pre-procesamiento

Tipos básicos de atributos (continuación):

Atributos categóricos.

- Son atributos **cualitativos** que se refieren a símbolos o nombres de elementos.
- Cada valor representa algún tipo de categoría, código o estado, con valores sin ningún tipo de relación de precedencia.
- También se les conoce como enumeraciones.
- Por ejemplo:
 - `contextura_cuerpo`: delgado, grueso, muy delgado, etc.
 - `estado_civil`: soltero, casado o divorciado.

Definiciones: Clasificación

Muestra	Características			Clase
	Peso	Forma	Tamaño	
	liviano	redondo	pequeño	uva
	mediano	alargado	mediano	banano
	pesado	ovalada	grande	sandía

Preprocesamiento - Imágenes

- Leer la imagen
- Escalarla
- Eliminar ruido
- Segmentar (de ser necesario)

<https://vanceai.com/denoise-ai/>

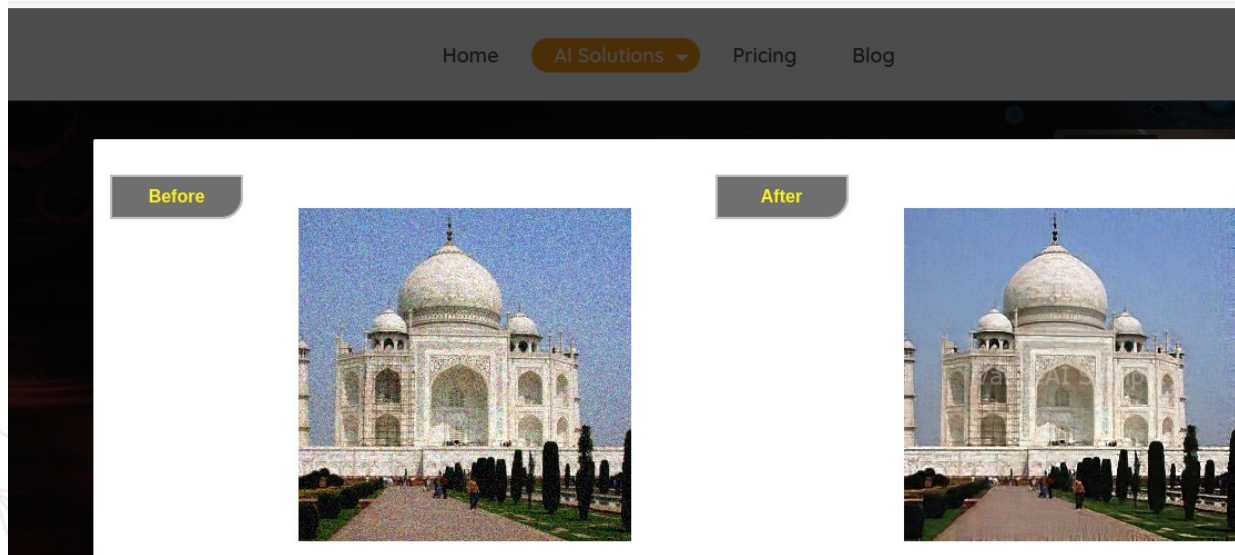


Imagen: <https://docs.gimp.org/2.8/es/plugin-rgb-noise.html>

¿Hasta donde preprocesar los datos?

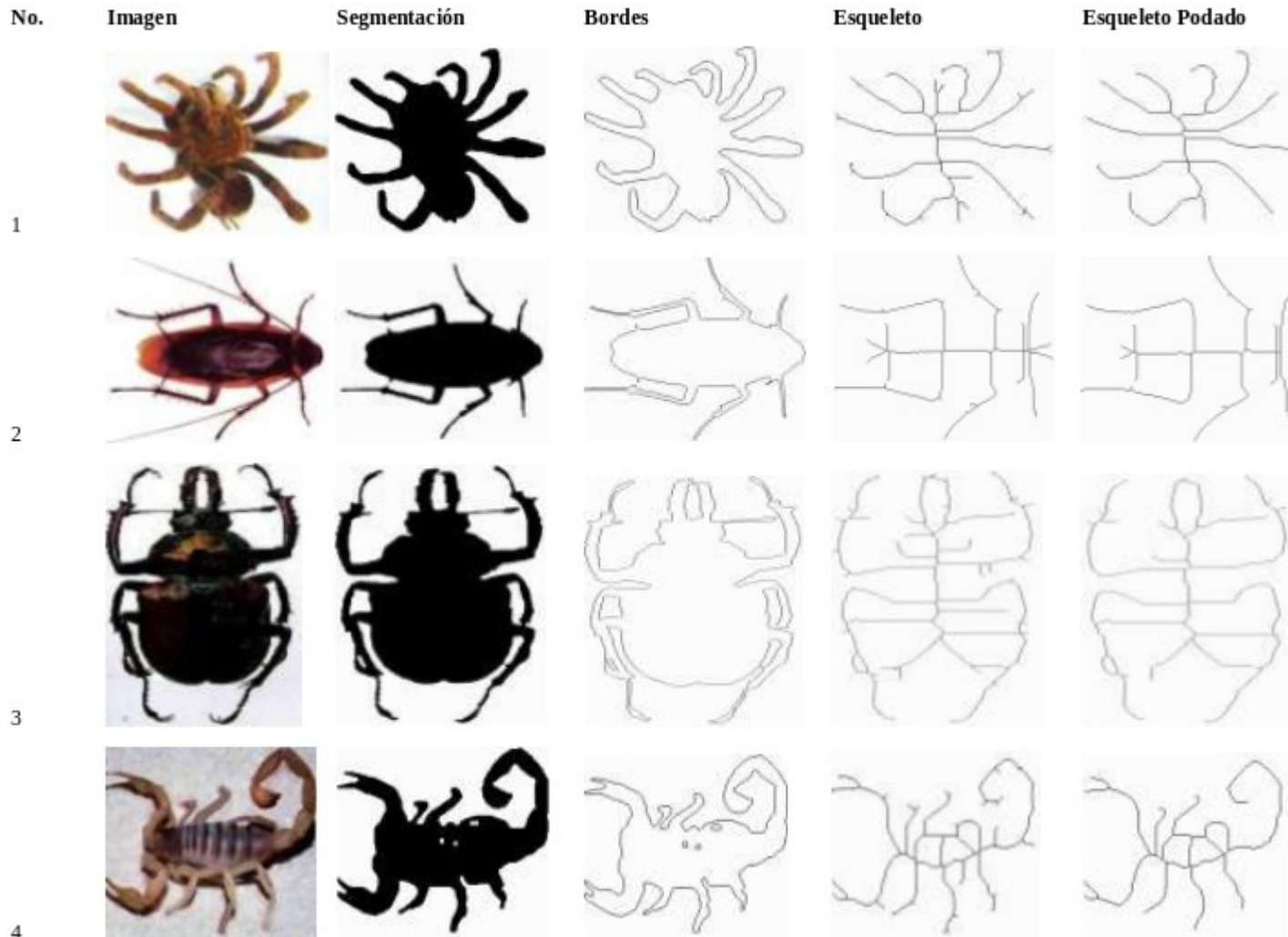


Imagen: Buchelly, F;Pencue, L. (2009). Clasificación e identificación de artrópodos mediante procesamiento digital de imágenes. Recuperado de <https://www.redalyc.org/pdf/903/90312171012.pdf>

Normalización / Estandarización

- Existen algoritmos susceptibles a escala
 - Estas técnicas permiten tener todas las características en el mismo orden de magnitud (o cercano)
- Puede invisibilizar valores muy extremos
- Difieren en bandas de acotación de los datos
- Ambas deben "traducirse" de vuelta para interpretación humana

2. Extracción características

- No todos los atributos en un conjunto de datos son "tipos básicos" (numéricos o categóricos)
- Describir tipos de datos complejos en **categorías o valores numéricos**
- Las características se convierten en columnas del conjunto de datos

Etapas básicas: Extracción de características

Ejemplo de los róbalos (sea-bass) y salmones empacados.

- **Entrada:** conjunto de datos preprocesados
 $V = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_M\}$, con $\dim(\vec{u}_i) = 307200$
- **Salida:** El funcional $G(\vec{v}_i) = \vec{x}_i = \langle x_1, x_2 \rangle$ extrae características de **ancho** del pescado, y a la «**claridad**» del **color** del pescado

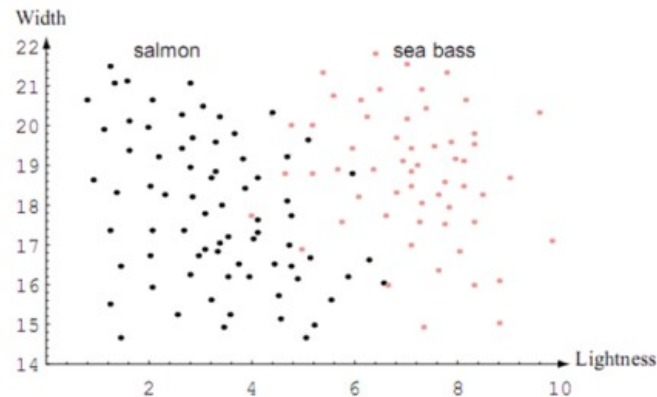


Figura: Diagrama de dispersión del espacio de muestras con dimensión $N = 2$

3. Predicción

- **Problema:** Partiendo de un conjunto de datos de ejemplo, donde se tienen todas las variables de entrada X y todas las variables de salida Y , **determinar una función h que aproxime la verdadera función $Y = f(X)$**
- **Hipótesis:** La función aproximada que estamos aprendiendo
- **Aprendizaje:** **Explorar el espacio de funciones h de manera que encontremos una con buen desempeño**
 - No sólo en datos de entrenamiento sino en ejemplos nuevos (datos de prueba)
 - Si tiene buen rendimiento fuera de los datos de entrenamiento, **"generaliza bien"**

Tipos de predicción

- **Clasificación**
 - Cuando la Y proviene de un número finito de alternativas
- **Regresión**
 - Cuando la Y es un número (posiblemente) real
- Cuando f es estocástica buscamos aprender una distribución condicional de probabilidad $P(Y | X)$. La probabilidad de Y dado X.

4. Validación / Evaluación

- **Premisa:** Existe una distribución desde la que se pueden muestrear todos los atributos y etiquetas/resultados
- Los ejemplos de "hoy" nos pueden permitir inferir una **hipótesis** que se acerque lo suficiente para predecir ejemplos "futuros"
- Podemos conectar el pasado al futuro si los ejemplos son independientes y distribuidos idénticamente (i.i.d. assumption)

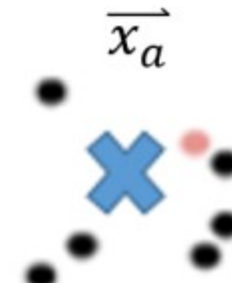
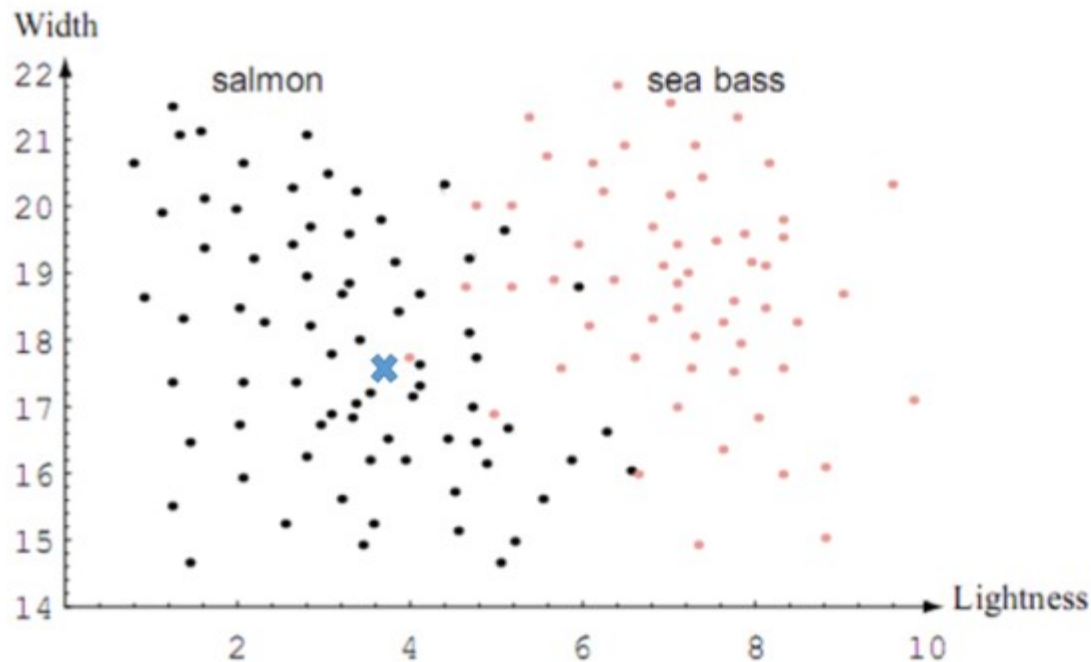
Evaluación de hipótesis

- Debemos determinar **qué tan cerca está nuestra hipótesis de la realidad**
- Tasa de errores
 - Proporción de ocasiones donde $h(x) \neq y$ para un (x, y) particular
- Se evalúa en subdivisiones de los datos para asegurarse que se consideran datos no incorporados durante el entrenamiento

Algunos ejemplos de algoritmos

Clasificación, no paramétrica

- **K vecinos más cercanos:** Para un punto nuevo \vec{x}_a , se calculan los $K = 6$ vecinos más cercanos, usando la distancia Euclidiana,
- En este caso $t_a = 1$ (C_1 corresponde a la clase salmón)



Clasificación, paramétrica lineal

- Dado un conjunto de muestras de entrenamiento X_e , construye un hiperplano o función que minimice el error de clasificación
- Error cuadrático medio, discriminante lineal de Fisher, RN de perceptrón, entre otros.

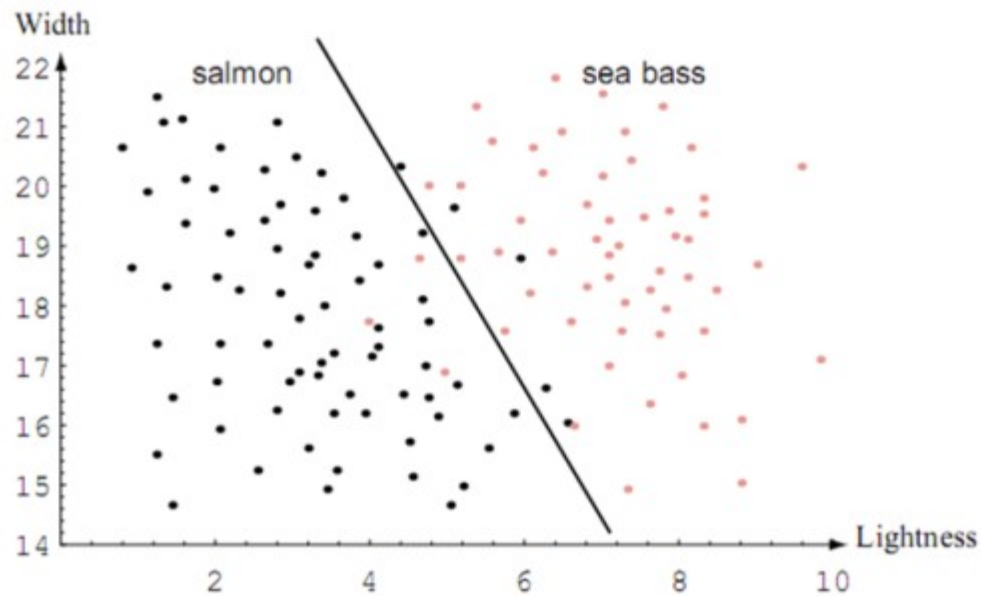


Figura: Espacio de muestras de entrenamiento X_e , con $N = 2$

Clasificación con superficie no lineal

- Clasificadores más sofisticados **generan superficies de decisión no lineales** (máquinas de soporte vectorial, redes neuronales, convolucionales, Bayesianas, etc)

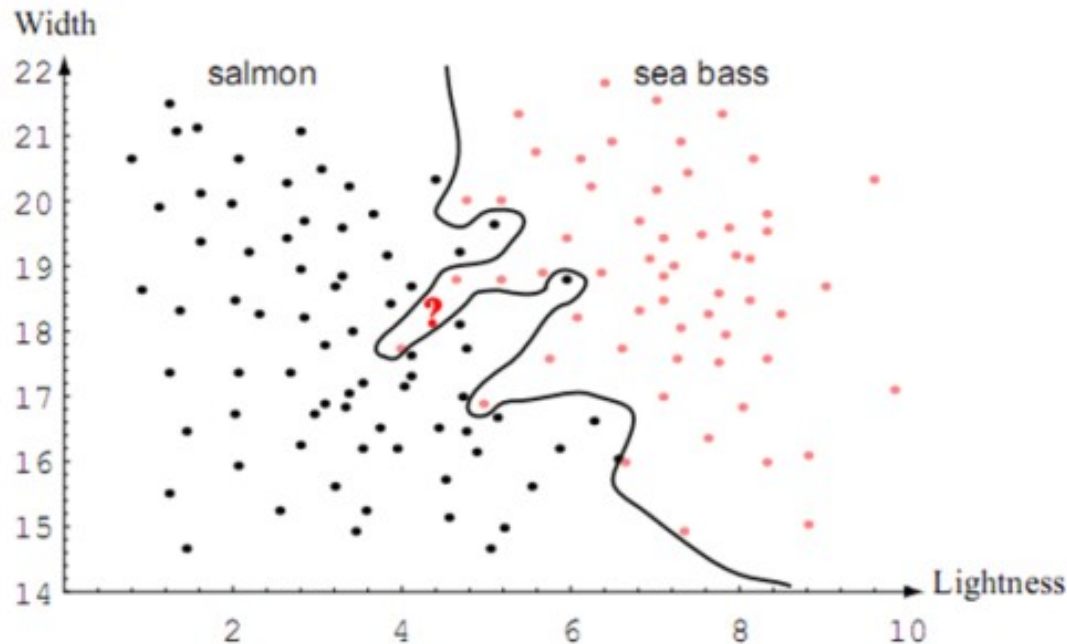


Figura: Sobre ajuste: Una superficie que se sobreajusta, confía al 100 % en X_e lo cual no es aconsejable...

Clasificación con superficie no lineal (regularizada)

- Clasificadores más sofisticados generan superficies de decisión no lineales (máquinas de soporte vectorial, redes neuronales, convolucionales, Bayesianas, etc)

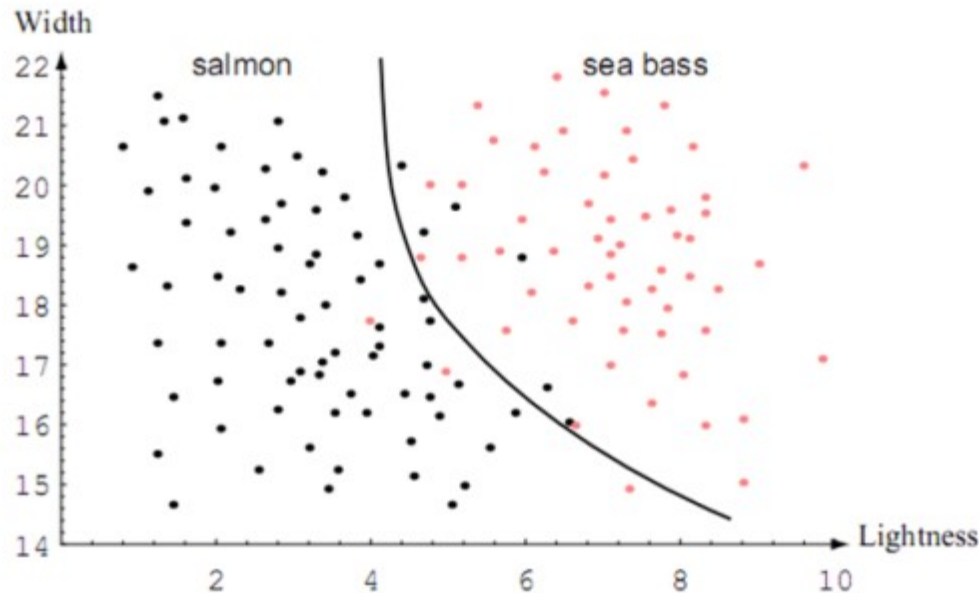
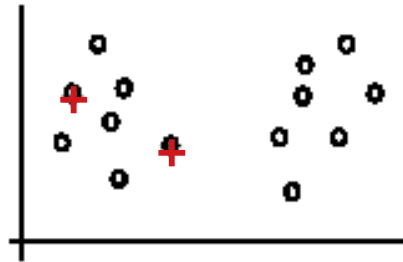


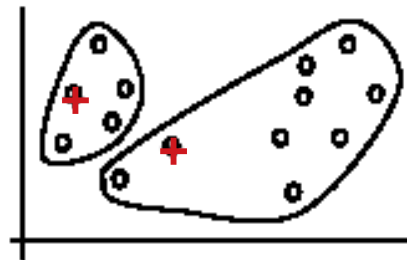
Figura: **Mejores resultados con superficies regularizadas**, para evitar sobre ajuste

Clasificación no supervisada: clustering

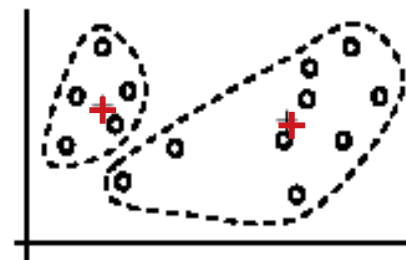
- Los **clasificadores anteriores** necesitan un conjunto de muestras etiquetadas X_e , por lo que se les dice supervisados.
- **No supervisados:** agrupan los datos para encontrar por sí solos las muestras de cada clase.



(A). Random selection of k centers



Iteration 1: (B). Cluster assignment



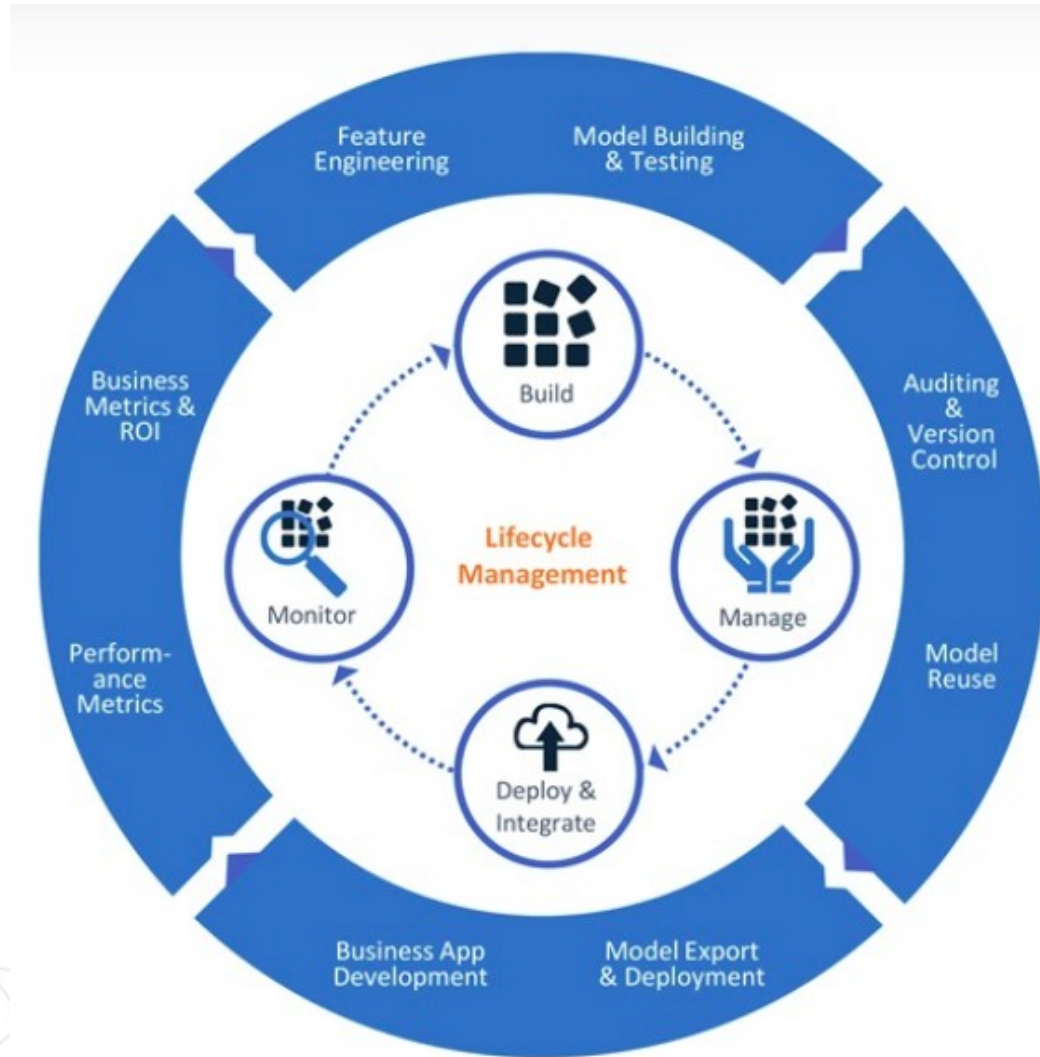
(C). Re-compute centroids

Funcionamiento del **algoritmo K-medias**, (Bing Liu, 2019)

Ciclo de vida de un Sistema de aprendizaje automático



De forma más amplia Operaciones de Machine Learning (MLOps)





Referencias

TEC | Tecnológico
de Costa Rica

