
Classification Problems

Sub2

Iván G Torre

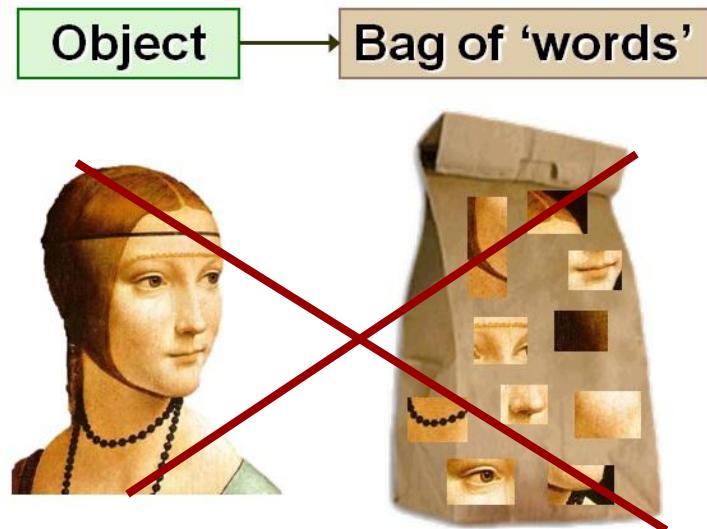
Limitaciones del modelo anterior

- Supone independencia de las variables, lo cual no siempre es cierto



Limitaciones del modelo anterior

- NLP es mucho más complicado que Bag of Words



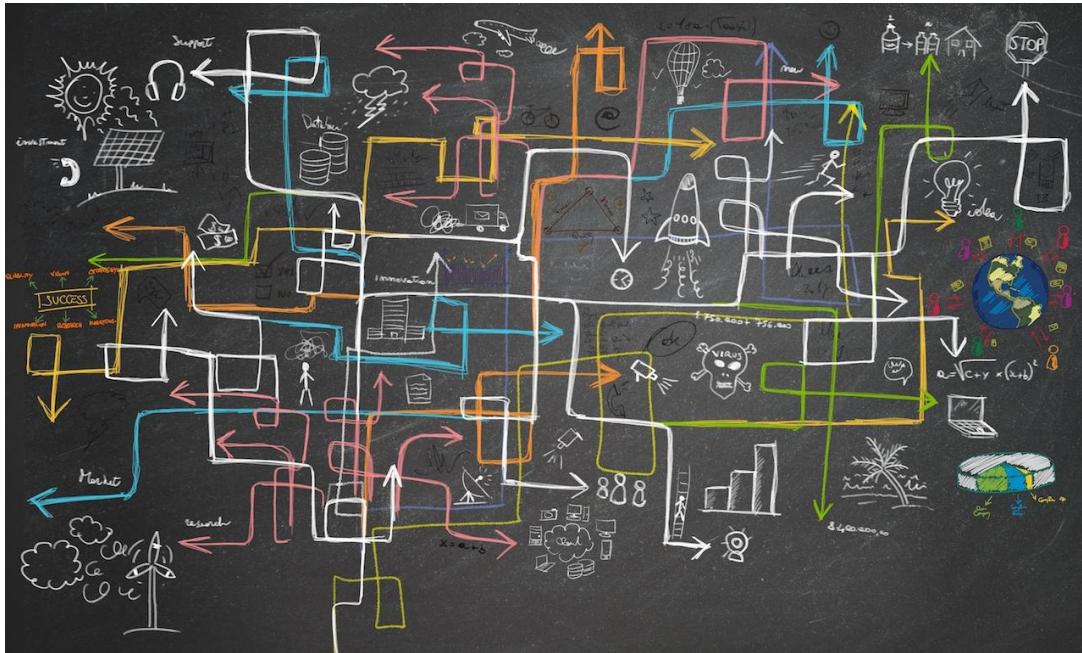
Limitaciones del modelo anterior

- En problemas más complejos con muchas features no son óptimos



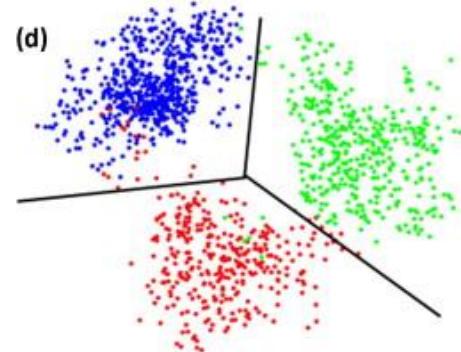
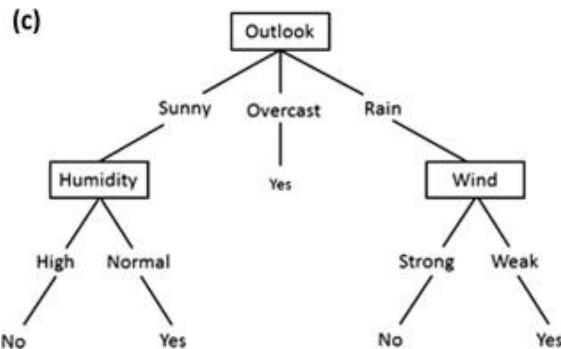
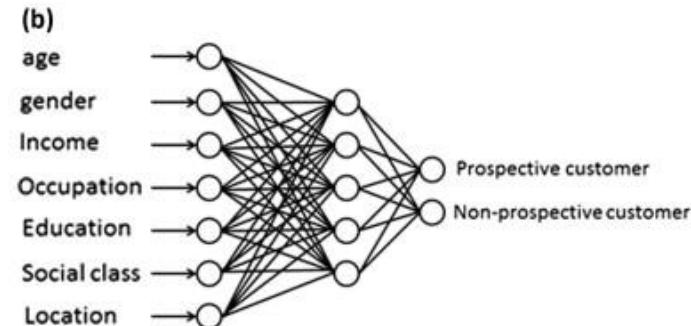
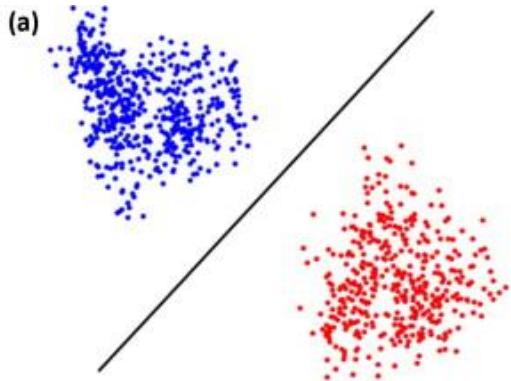
Limitaciones del modelo anterior

- Pueden aumentar el modelo de complejidad y tiempo de cálculo rápidamente

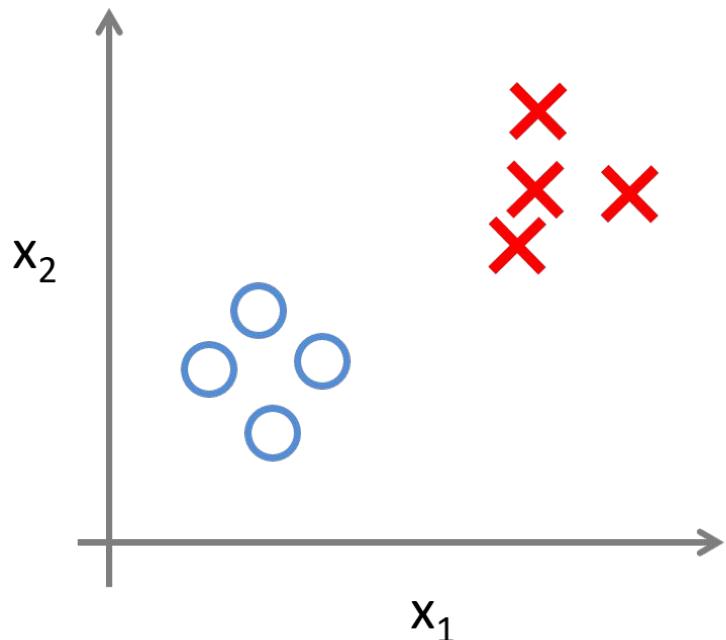


Otros algoritmos de clasificación populares

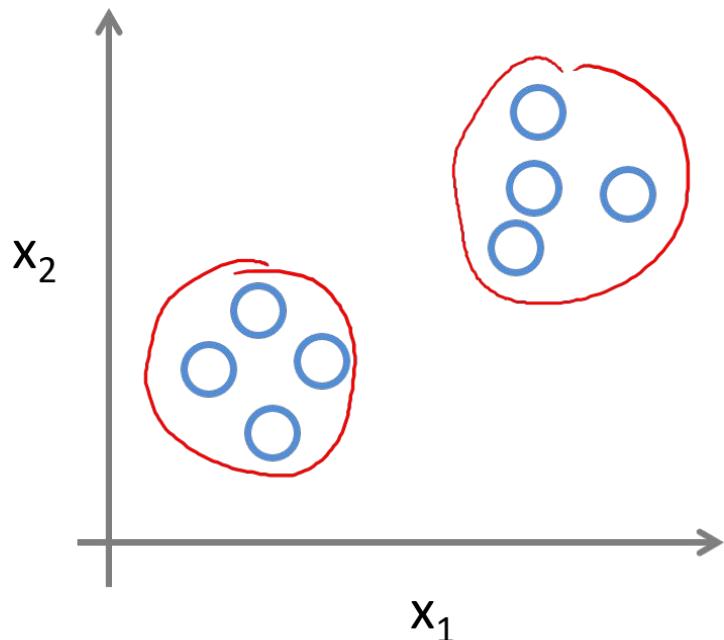
- SVM
- KNN
- Random Forest
- XGBoost



Supervised Learning



Unsupervised Learning



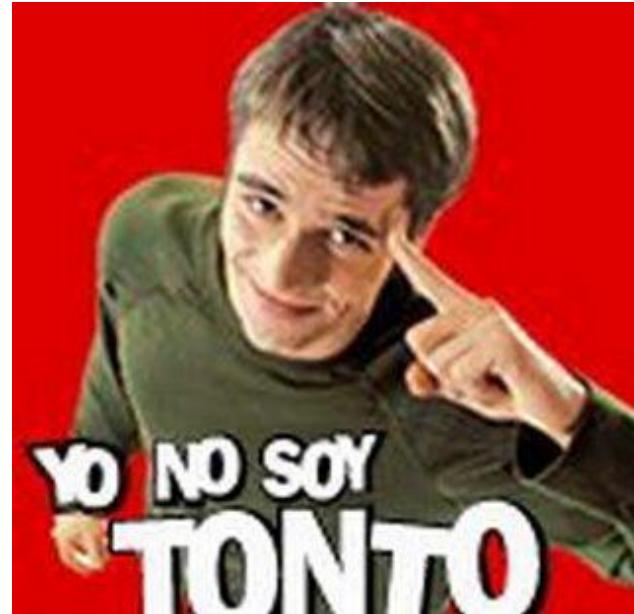
Aprendizaje Supervisado:Datos

- Data con etiquetas:
 - Complicado
 - Caro
 - ¿Calidad?
 - Serie histórica



Aprendizaje Supervisado

- ¿Solución?:
 - Tomadlos gratis



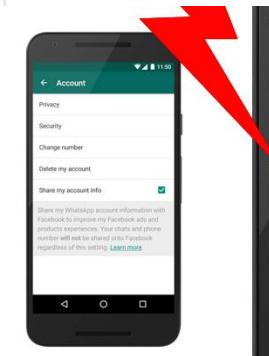
Aprendizaje Supervisado: Labels

This screenshot shows the original Facebook profile page for Mark Zuckerberg. At the top, it says "facebook" and "Mark Zuckerberg's Profile (This is you)". Below that is a large profile picture of Mark Zuckerberg. To the right of the picture, his name is displayed along with his education at "Harvard" and his location "Facebook San Francisco, CA". Underneath this, there is a section for "Sex: Male", "Birthday: May 14, 1984", "Hometown: Dobbs Ferry, NY", and "Relationship Status: In a Relationship". Below this is a "Mini-Feed" section showing 10 stories, including comments from Ezra Callahan, edits to his profile picture, and posts from the group "LOCKDOWN". Further down, there are sections for "Status" (with 2 updates this week), "Friends" (Harvard Friends, 147 friends at Harvard), and "Information" (Contact info, Email, AIM Screenname).



Tweets: 5 | Tweets & retweets: 1

Not Mark Zuckerberg
@NotMarkZuckerberg
Facebook stock is soaring because I blew away Wall St's estimates for earnings per share, spend per advertiser, per user.



Change number

Delete my account

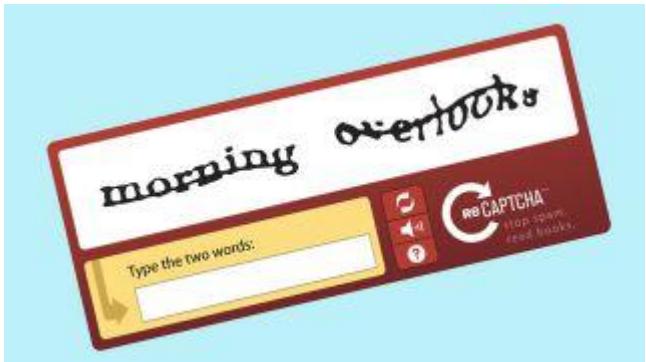
Share my account info

Share my WhatsApp account information with Facebook to improve my Facebook ads and products experiences. Your chats and phone number will not be shared onto Facebook regardless of this setting. [Learn more](#).

This screenshot shows the "Create New Account" form. It includes fields for "Phone Number" (with a dropdown for "Country: US (+1)"), "Create a PIN" (with a note "Must be 6 digits" and "Numbers Only"), "Confirm PIN", "Email Address", "Password" (with a note "Must be at least 6 characters"), and "Confirm Password". Handwritten annotations with orange arrows point to the "Phone Number" field and the "Email Address" field. The "Continue" button is at the bottom right.

Aprendizaje Supervisado: labels

- Data con Labels: Esto es lo más complicado!



Jane
Last Name: Smith
Email: stopall1
Pick your color:
 Red
 Green

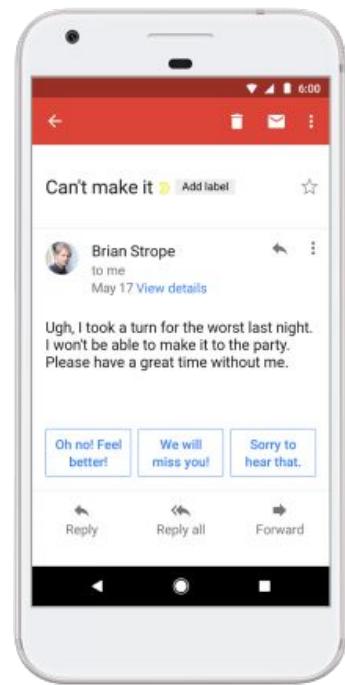
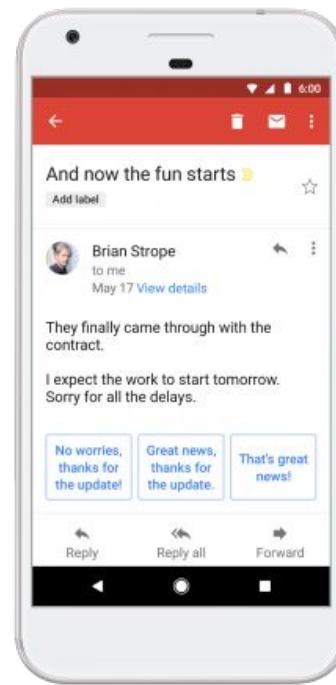
Select all squares with street signs.

Selecciona todos los cuadrados de carteles de calles.

CAPTCHAS
Cada día más complicados

más en cuantaranazon.com

Aprendizaje Supervisado: labels



Aprendizaje Supervisado: labels



G Cronología HOY viernes 7 sept. 2018 1 lugar para corregir

+ 5 MÁS

En bicicleta - 3,7 km 14 minutos

toom Baumarkt 11:03-11:31 CONFIRMAR

Am Sundheimer Fort 1, 77694 Kehl, Alemania

En bicicleta - 400 m 2 minutos

Lidl 11:33-12:12 CONFIRMAR

Boschstraße 3, 77694 Kehl, Alemania

En bicicleta - 3,3 km 14 minutos

This block shows a screenshot of Google Maps' Timeline feature. It displays a series of cycling routes and their details. The first route is to "toom Baumarkt" at "Am Sundheimer Fort 1, 77694 Kehl, Alemania". The second route is to "Lidl" at "Boschstraße 3, 77694 Kehl, Alemania". Both routes are marked as confirmed.



Aprendizaje Supervisado: gamification



Internet Interventions
Volume 6, November 2016, Pages 89-106
open access



Gamification for health and wellbeing: A systematic review of the literature

Daniel Johnson ^a , Sebastian Deterding ^b, Kerri-Ann Kuhn ^a, Aleksandra Staneva ^a, Stoyan Stoyanov ^a, Leanne Hides ^a

Show more

Scientists launch virtual reality game to detect Alzheimer's

<https://doi.org/10.1016/j.invent.2017.08.001> August 29, 2017

Under a Creative Commons



QQ

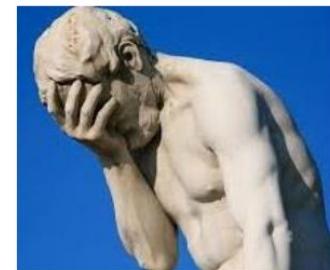
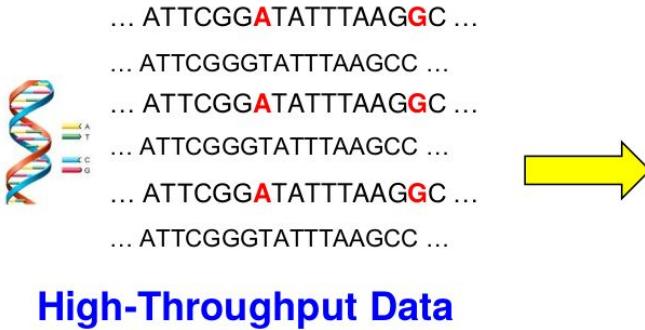
PLAY EDUSPOT GET INVOLVED! ABOUT TUBERSPOT

Play 1 minute against Malaria!



<https://game.malaria-spot.org/>

Aprendizaje Supervisado: diagnosis



Discovery

Bottleneck #1: Knowledge

Bottleneck #2: Reasoning

Aprendizaje Supervisado: diagnosis

The screenshot shows the PubMed homepage (<https://www.ncbi.nlm.nih.gov/pubmed>) with a yellow overlay containing text about the service.

PubMed
PubMed comprises more than 27 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

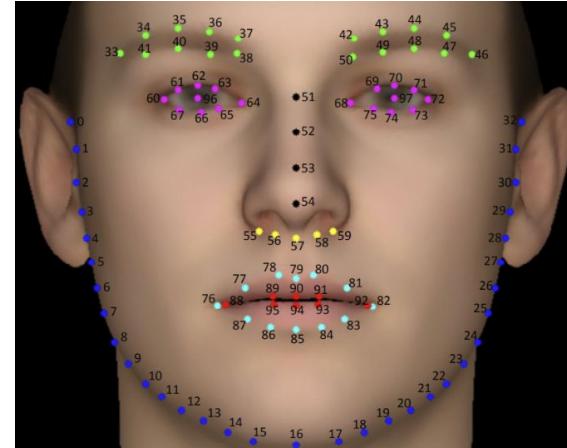
Everyday
PubMed: 4000 new papers
Expert can curate <10

Using PubMed

- [PubMed Quick Start Guide](#)
- [Full Text Articles](#)
- [PubMed FAQs](#)
- [PubMed Tutorials](#)
- [New and Noteworthy](#)

Aprendizaje Supervisado: data annotation

Data Annotation: The Billion Dollar Business Behind AI Breakthroughs



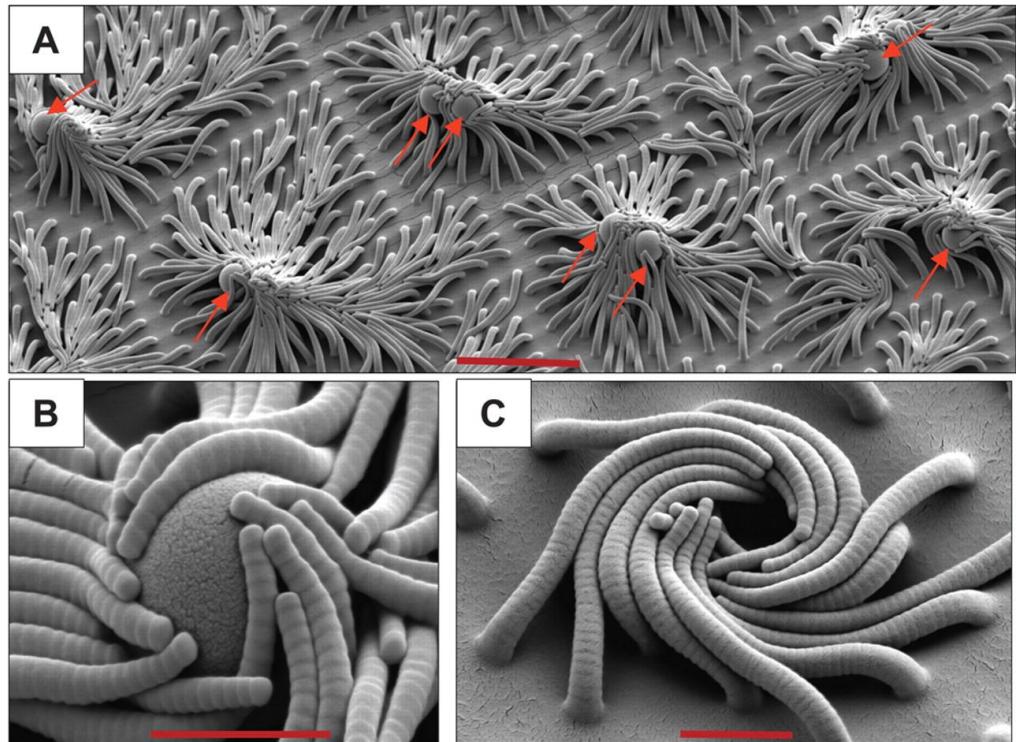
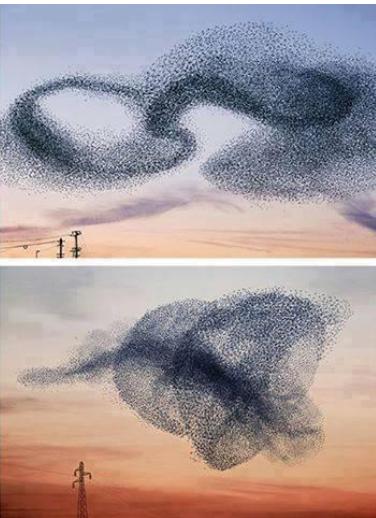
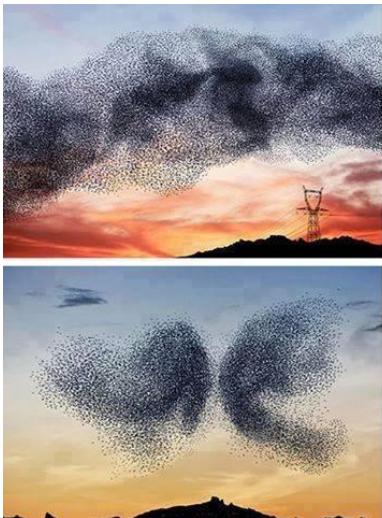
Aprendizaje No supervisado

- No train data



Aprendizaje No supervisado

- Auto-organización

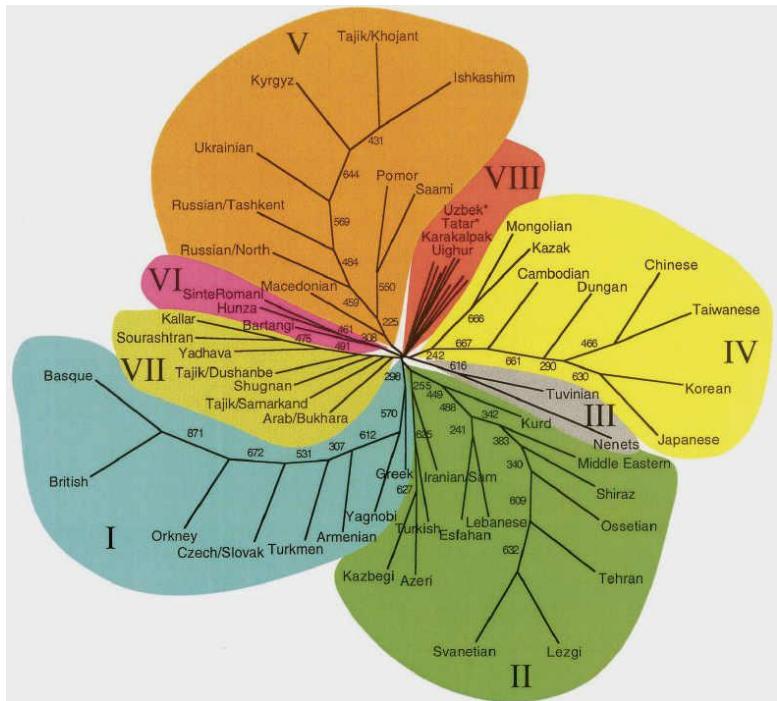
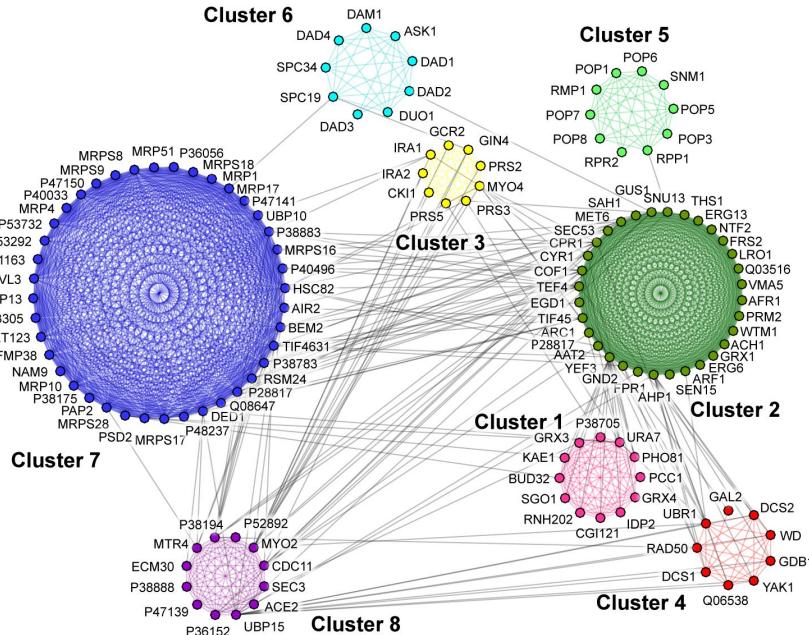


Aprendizaje No supervisado

- Existen correlaciones (lineales o no lineales)

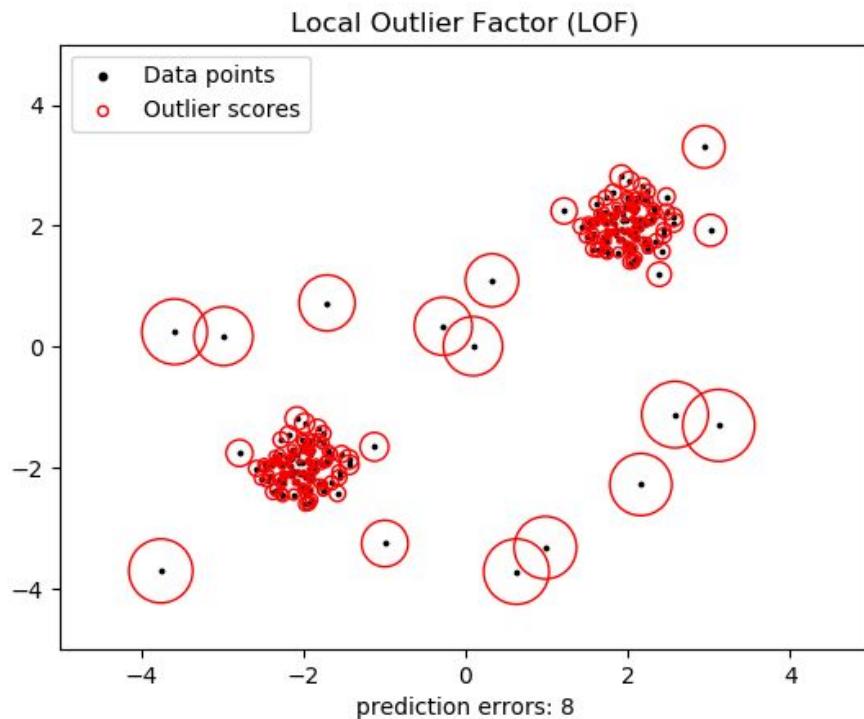


Aprendizaje No supervisado: Clustering

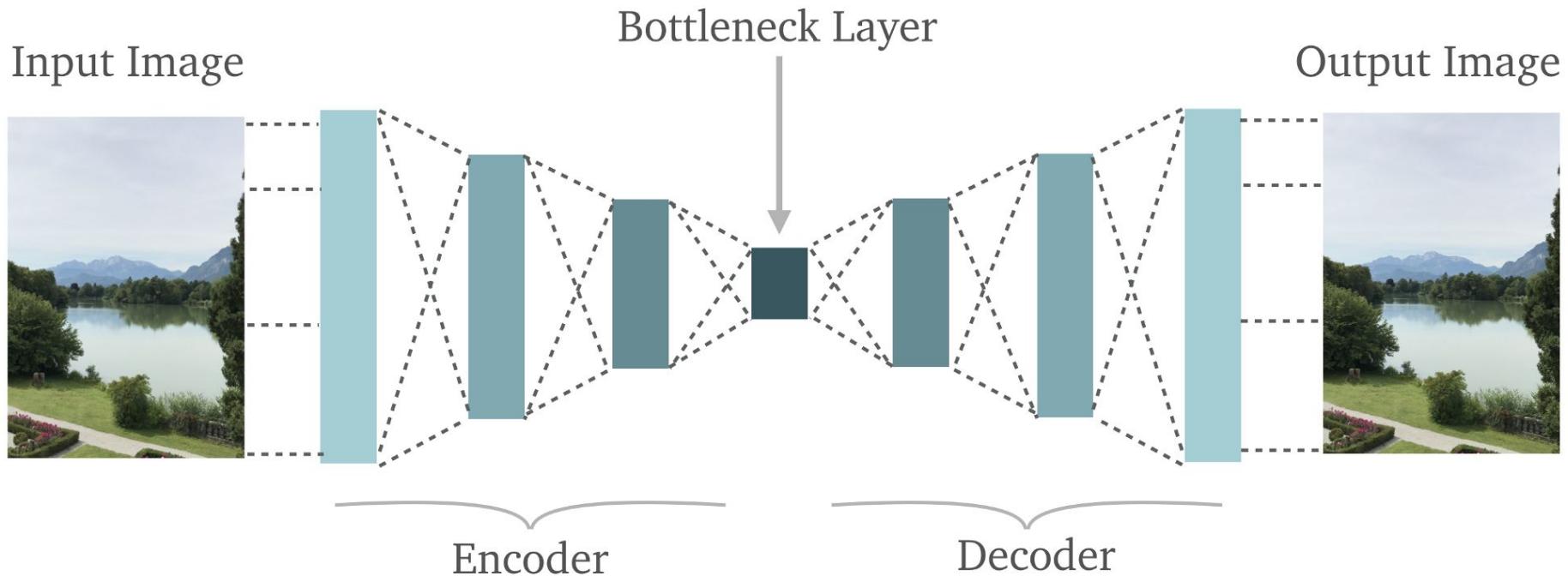


Aprendizaje No supervisado: Detección anomalías

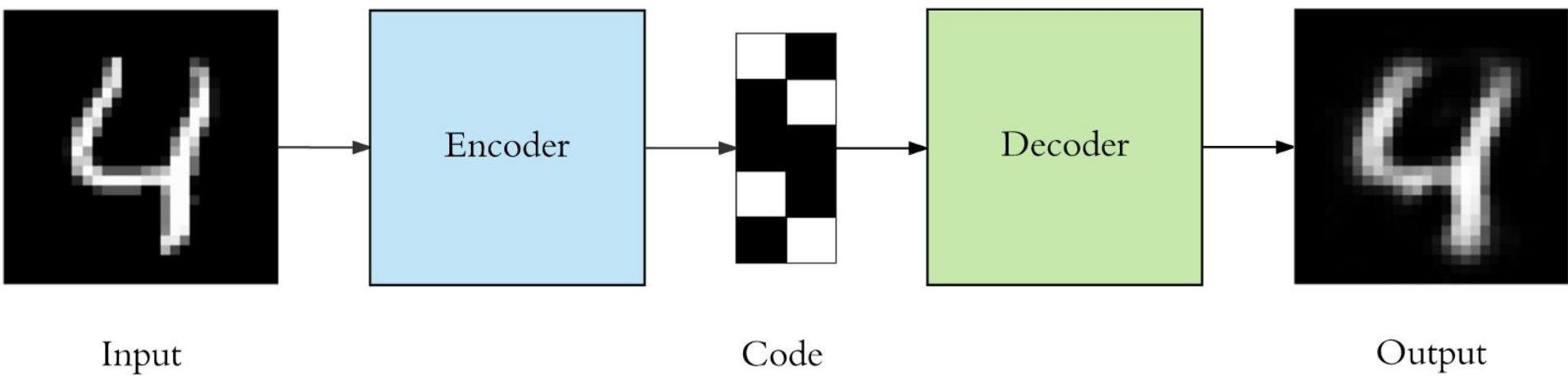
- Fraude
- Ataque
- Detección de bots
- Fallo no catastrófico



Aprendizaje No supervisado: Compresión



Aprendizaje No supervisado: Compresión

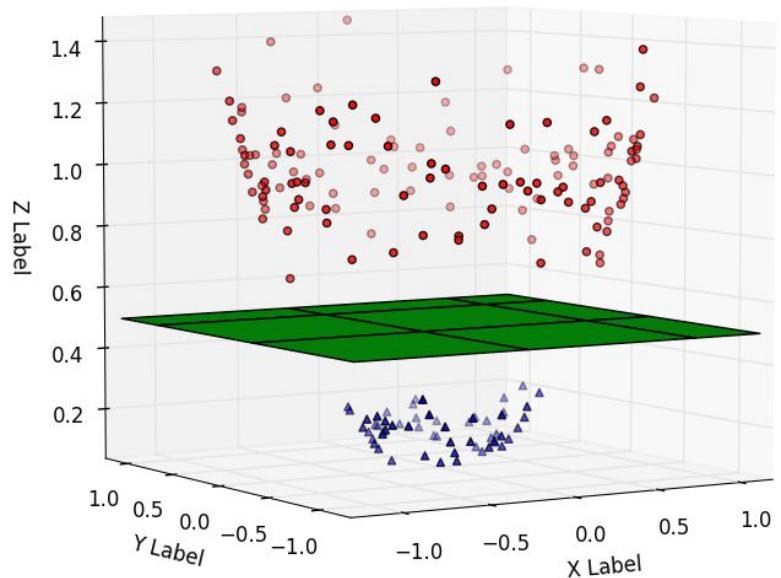


Algoritmos de clasificación

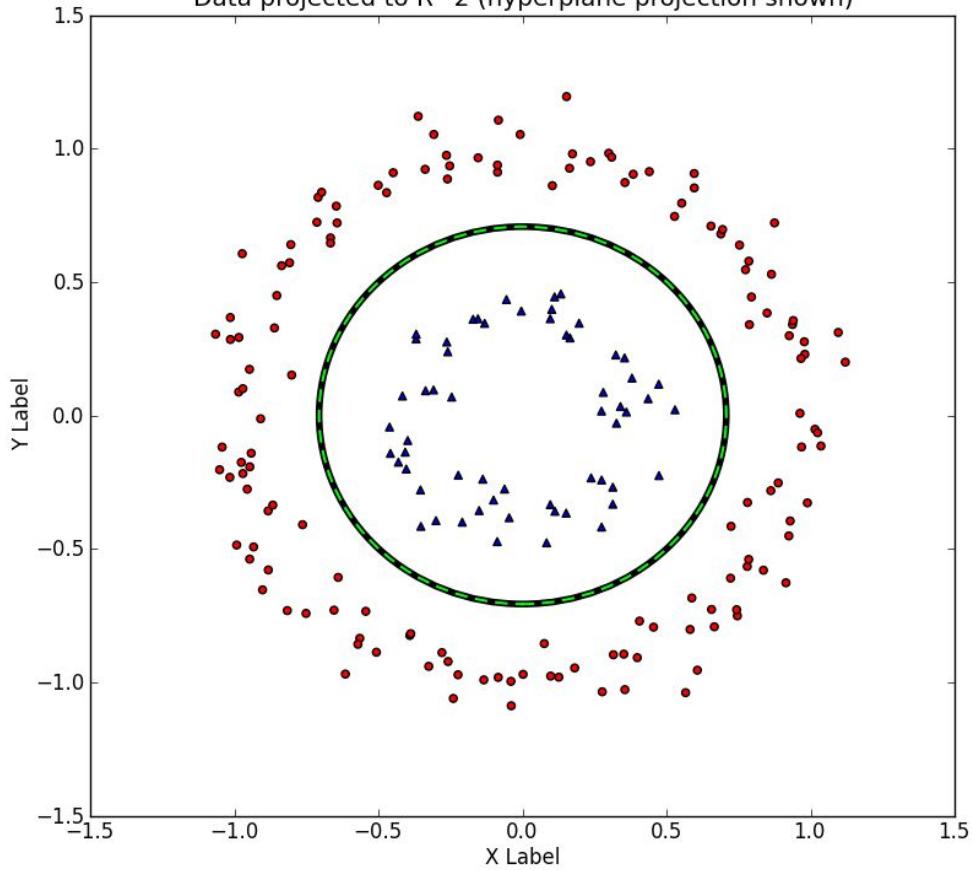
Algoritmos de clasificación: SVM

- Busca el mejor hiperplano posible para separar las clases (o conjunto de hiperplanos)

Data in \mathbb{R}^3 (separable w/ hyperplane)

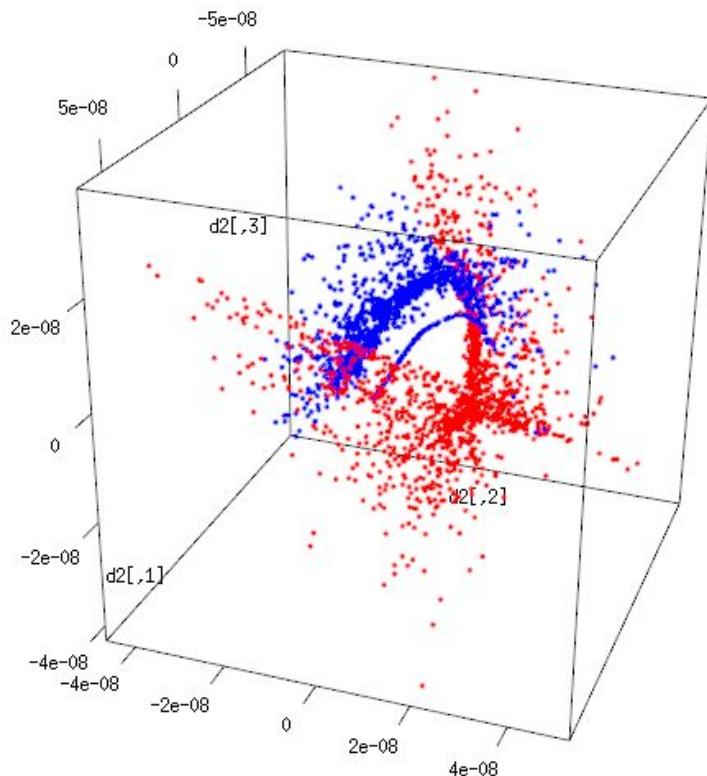
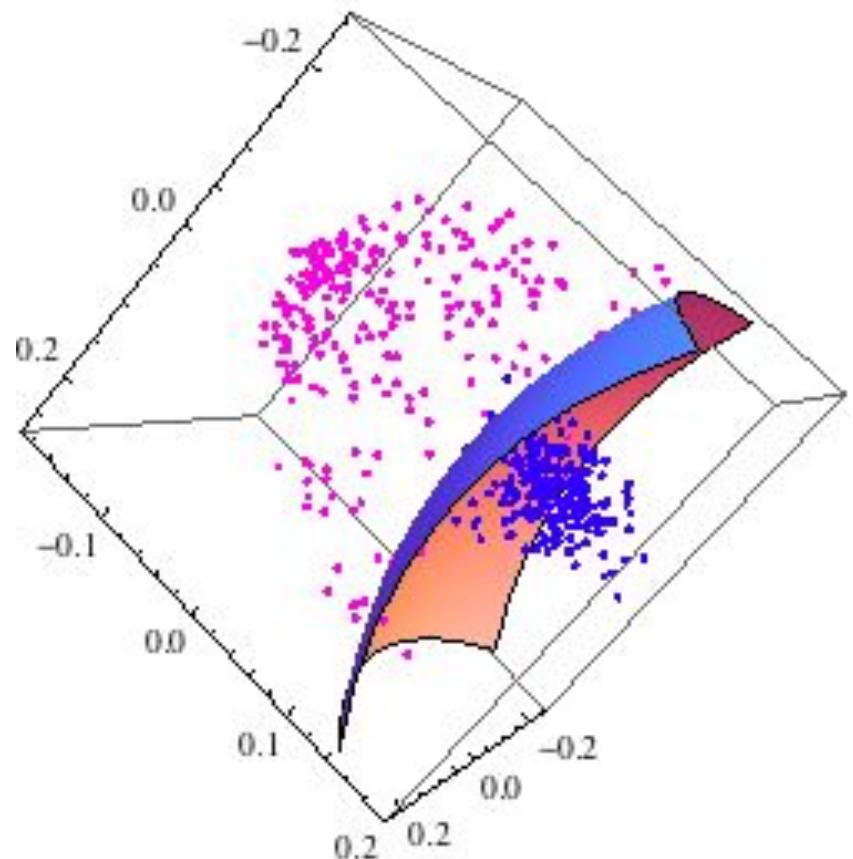


Data projected to \mathbb{R}^2 (hyperplane projection shown)



Algoritmos de clasificación: SVM

- Clasificador binario no probabilístico



Algoritmos de clasificación: SVM Multiclas: Uno Contra todos

- Se entrena un modelo por cada clase: 1 contra todos.



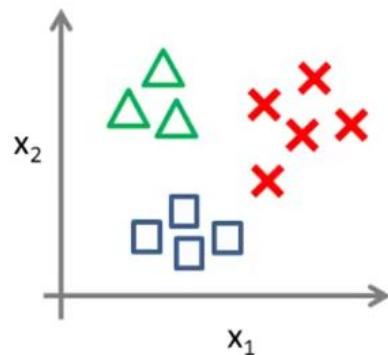
Algoritmos de clasificación: SVM Multiclas: Uno Contra todos

- Necesitamos una métrica para predecir a cuál pertenece en caso de conflicto

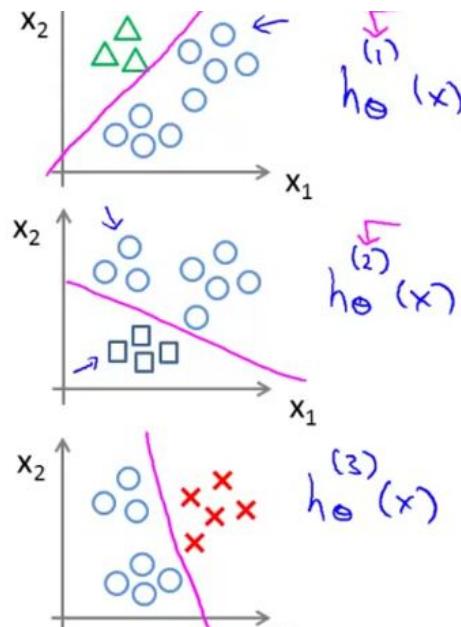


Algoritmos de clasificación: SVM Multiclas: Uno Contra todos

One-vs-all (one-vs-rest):



- Class 1: ←
- Class 2: ←
- Class 3: ←



Algoritmos de clasificación: SVM Multiclasificación: Uno contra uno

- Se entrena $k*(k-1)/2$ clasificadores para un problema de k clases.

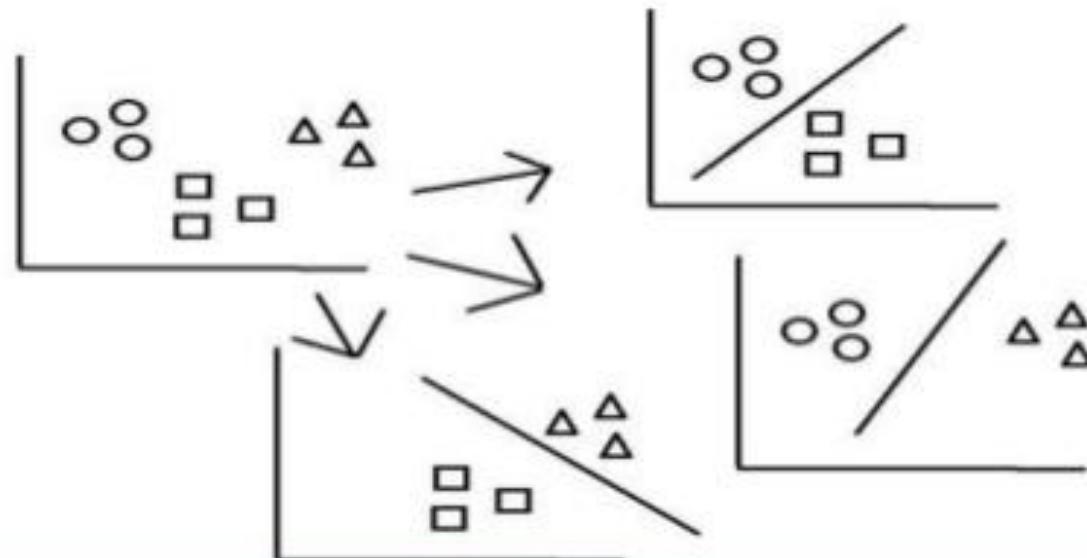


Algoritmos de clasificación: SVM Multiclas: Uno contra uno

- También se necesita una métrica para elegir

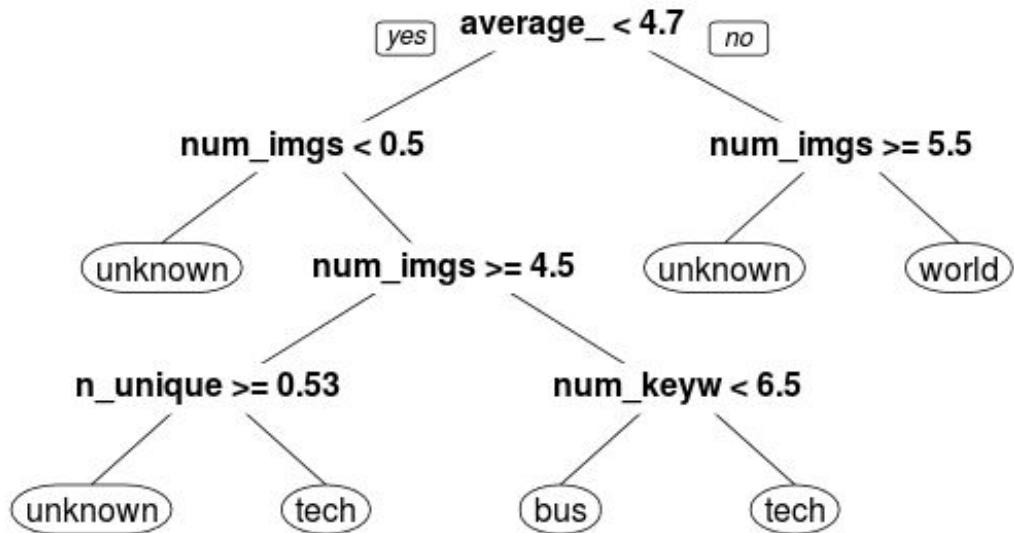


Algoritmos de clasificación: SVM Multiclas: Uno contra uno



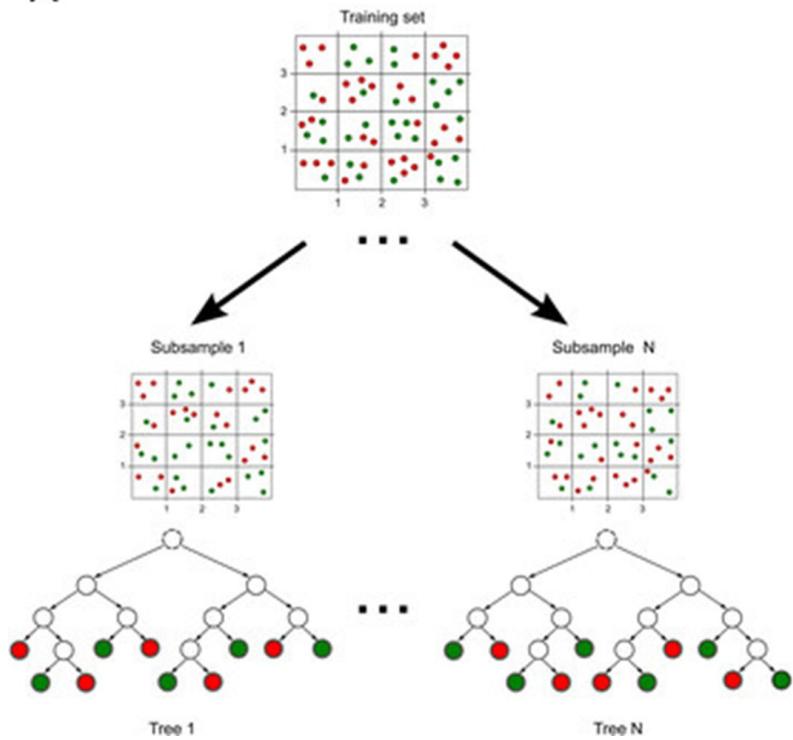
Algoritmos de clasificación: Árbol de decisión

Gini Split Criteria

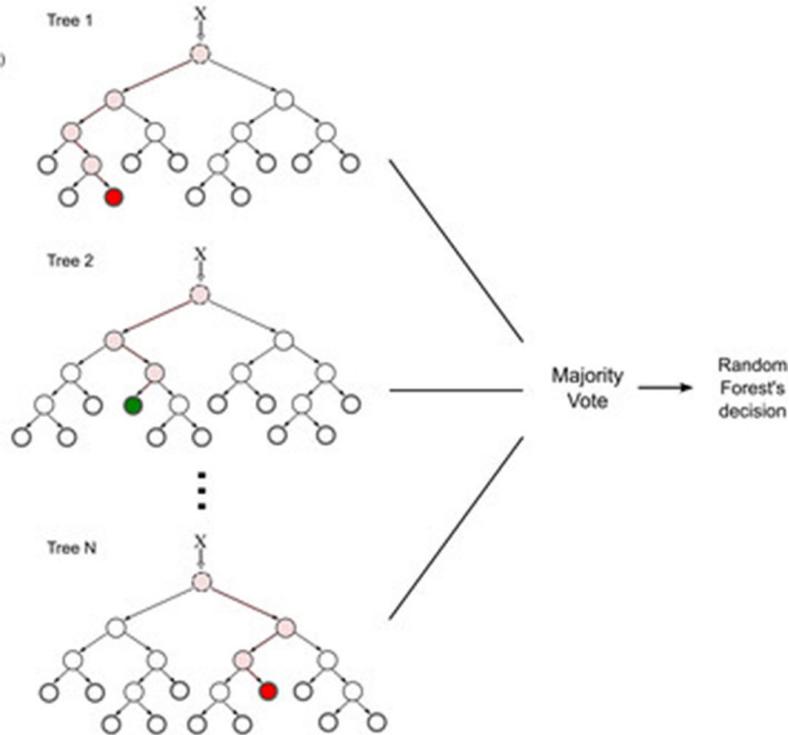


Algoritmos de clasificación: Random Forest

A

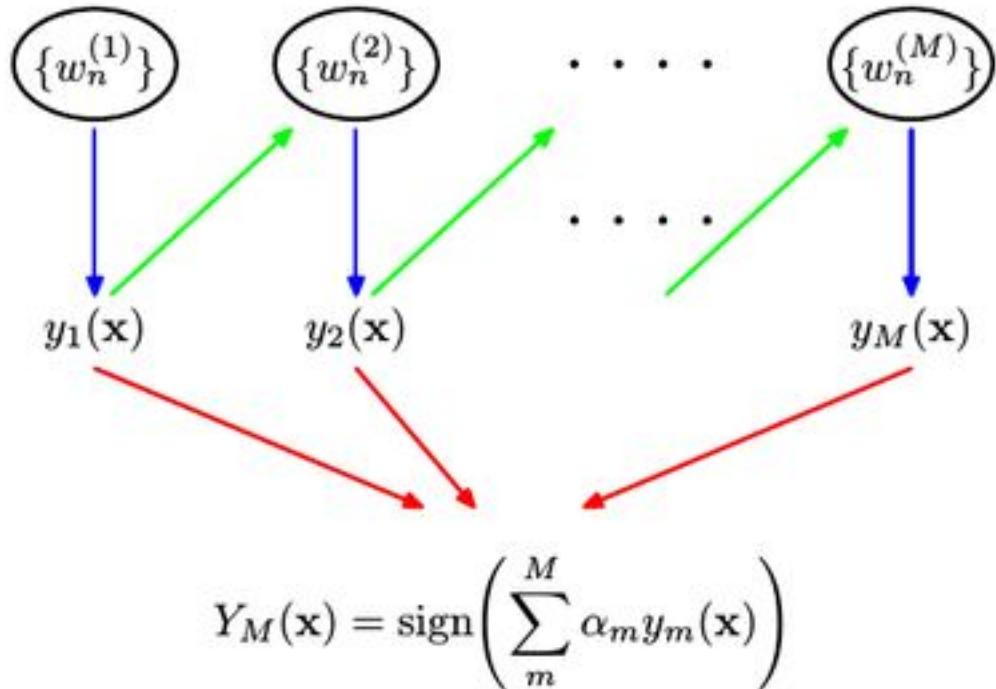


B



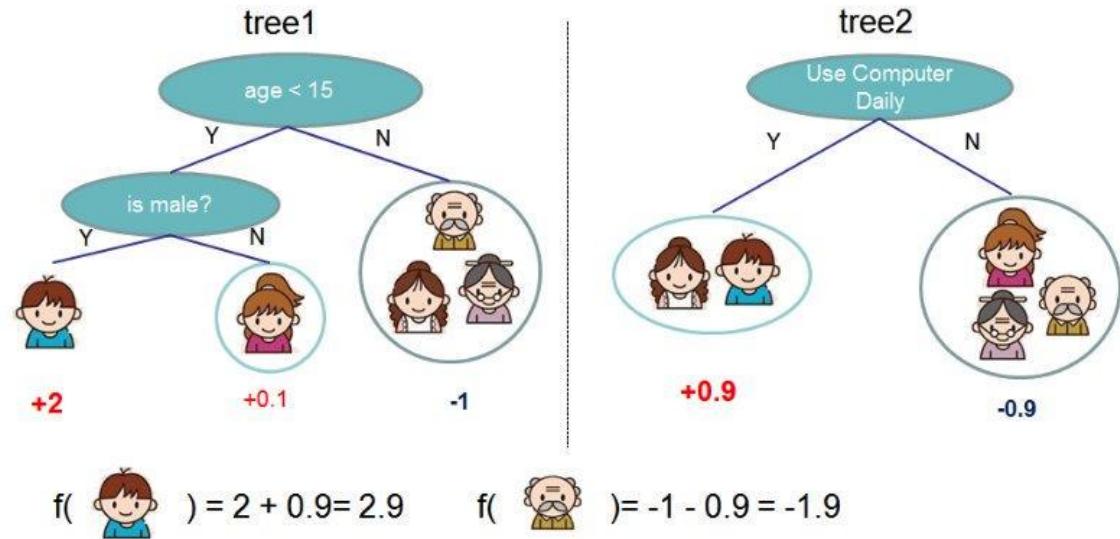
Algoritmos de clasificación: XGboost

Cada clasificador es entrenado en función de un peso del anterior. El peso que se le da al anterior clasificador vendrá dado por su rendimiento.



Algoritmos de clasificación: XGboost

Se construyen diversos árboles de decisión que se complementan



Multilabel y Multiclasificación

multi-class

- crossing?
- waiting?
- queuing?
- walking?
- talking?

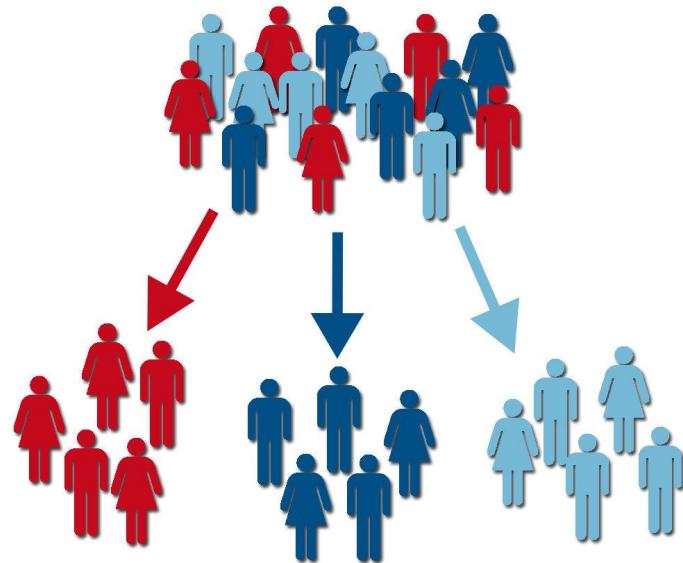


multi-label

- crossing?
- waiting?
- queuing?
- walking?
- talking?

correlation

Práctica 2: Supervised client segmentation y rendimiento de algoritmos



TYPES OF CUSTOMER SEGMENTS

	NPV PER CUSTOMER
	<ul style="list-style-type: none">• VALUE CONVENIENCE IN DELIVERY, ORDERING• HIGH INCOME• LONG RELATIONSHIP, LARGE REFERRALS
	<ul style="list-style-type: none">• BRAND BUYERS, NOT PRICE SENSITIVE• HIGHEST INCOME, MORE OFTEN MALE• EXPENSIVE TO ACQUIRE, BUT BUY MOST INITIALLY AND REFER MORE
	<ul style="list-style-type: none">• NOT CONCERNED WITH PERISHABLES OR DELIVERY TIME WINDOWS• SMALL SPENDING GROWTH
	<ul style="list-style-type: none">• INFLUENCED BY RETAILER BRAND, SUGGESTIONS, AND PROMOTIONS• LOW INCOME• SMALL SPENDING GROWTH/REFERRAL
	<ul style="list-style-type: none">• PRICE IS PRIMARY AND PERISHABLES ARE NOT IMPORTANT• LOW INCOME• SMALL PURCHASES

SOURCE: BAIN/MAINSPRING ONLINE RETAILING SURVEY