

## Tipos de aprendizaje

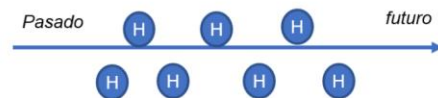
1. **Conceptos.**
2. **Regresión.**
3. **Segmentación.**
4. **ANN.**



**Variable:** son atributos o características susceptibles de cambios que adoptan diferentes valores dentro de un mismo entorno.

TIPOS			
Numéricas		Categorías	
Discretas	Continuas	Nominales	Ordinales
Numero de carros Edad	Temperatura	Genero	NSE

Nota: las variables numéricas son susceptibles de cálculos matemáticos es decir se pueden sumar, restar, dividir etc.



## MODELAMIENTO DE PROBLEMAS

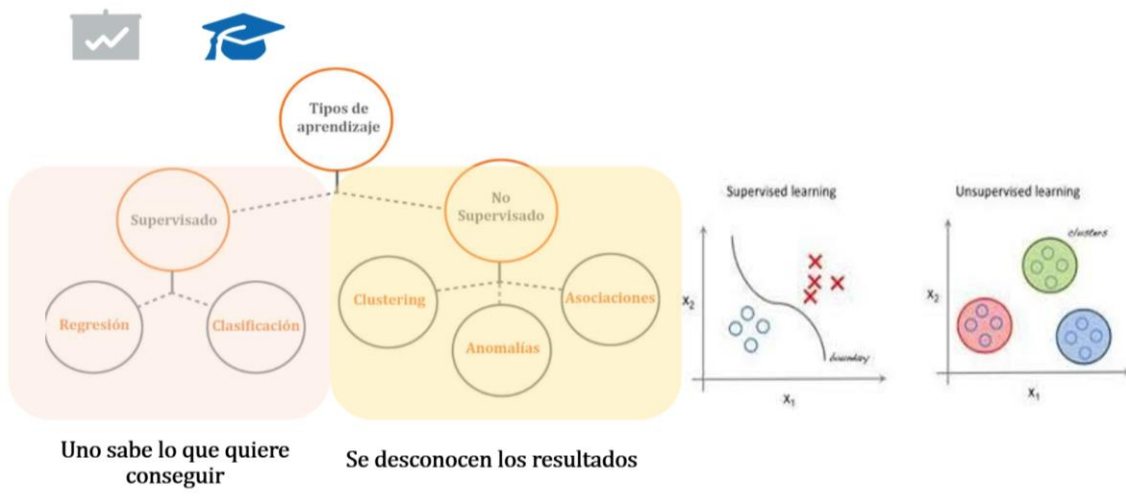
**Columna:** Es una variable.

**Registro:** Característica, cosa, objeto, elemento

**Numéricas:** Cualquier función matemática, operaciones. Enteros (Discretas), decimales (Continuas, temperatura)

**Categorías:** Codificación de variables, asignaciones (0, 1). Ordinales (Nivel socioeconómico), Nominal (Genero).

No todo sujeto temporalidad



## Aprendizaje:

**Supervisado:** (Guía-variable objetivo- Target) Ej. Carros que se venden en periodos (Meses), variable objetivo año anterior incremento %. Satisfacción del cliente. (Que tanto recomendaría, satisfacción del servicio). Variable objetivo-pregunta (Nivel de satisfacción- estimación)

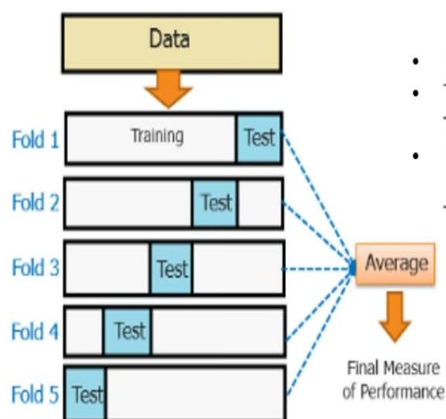
**Regresión:** Proceso variable continua

**Asociación:** Variable categórica, escala Glasgow

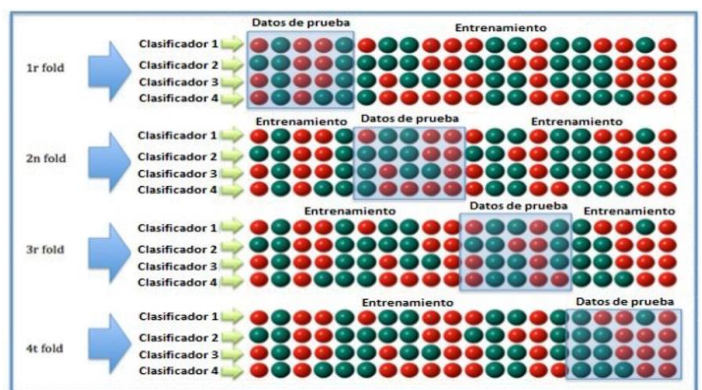
Fraude (1 ó 0)

**No Supervisado:** (No está la variable Objetivo). No hay asociación. Ej. Proceso e Segmentación (Clustering). Delimitación de generaciones gustos, preferencias (Centenian, milenium). Determinar segmentos, hábitos en común, caracterización. Homogéneos, heterogéneos.

Reglas de asociación, reglas de negocio, asociación



- Técnica para validar modelos
- Método para estimar que tan preciso es el modelo
- K-fold CV
  - Más popular
  - K usualmente es de 10
  - Cada Muestra es usada en las dos sub muestras de aprendizaje y validación



Subconjunto de datos de entrenamiento.

Ej. Riesgo de fraude el 70% del conjunto de datos. Testeo

Diferenciar, corregir proceso de datos, tema cíclico.

Testeo capacidades del modelo, entrenamiento.

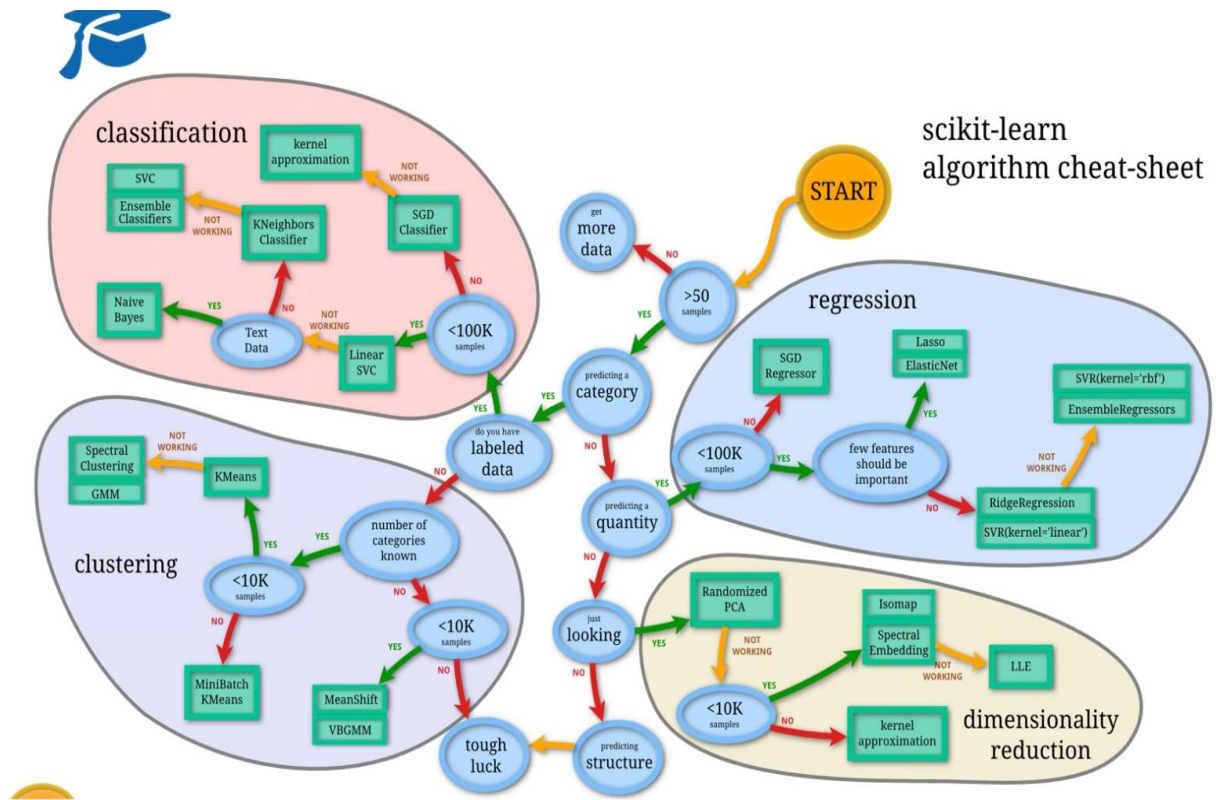
80/20

70/30

**Metodología:** Variación cruzada, partición subconjunto de datos. Para modelo testing

Ej. 100.000.000 de datos (partición)

Gestión de los procesos: Servidor validación 1 a 1 100.000.000.000 consumo nube, gestión de tecnología, consumo de energía. Responsable de las herramientas.



Librería de PHYTON. Análítica avanzada.

Mas de 50 muestras (registros)

Etiqueta guía. Que camino. Clasificación o camino clúster

Variable Guía.

Productos de Belleza. 4 subgrupos

Cliente nuevo: Correr modelo de clasificación, intersección

Predecir categoría.

Modelo de intersección. Componentes. Reducción de dimensionalidad.

Conjunto de información gestionar 100 variables, muy complejo hay que dimensionar, componentes principales, asociar. Ej. Hábitos de consumo

Gestionar conjunto de datos

Tratamiento de los datos.

Capturar la mayor cantidad de información es lo que busca el modelo.

Estadística se traduce en **VARIANZA**

Descriptiva:

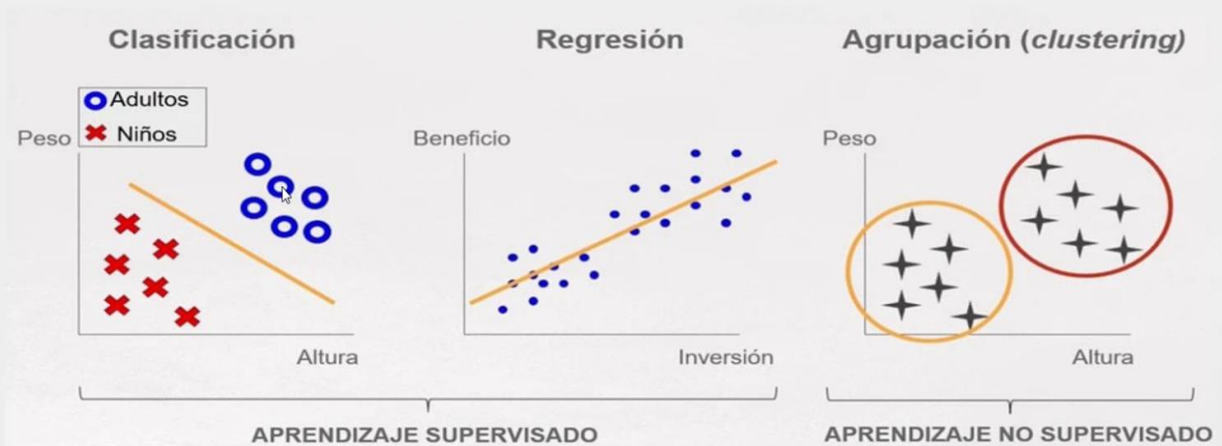
Datos faltantes, se deben gestionar.

Ej. Variable EDAD, imputación por promedio (30-40 años)

Tasa de datos faltantes

Muchas metodologías.

## Técnicas de Machine Learning



# MODELOS DE APRENDIZAJE AUTOMATICO



## SUPERPERFILES.

Estadística Bayesiana y clásica.





YouTube search results for "statquest".

**Left Sidebar:**

- Inicio
- Shorts
- Suscripciones
- Biblioteca
- Historial
- Mis videos
- Ver más tarde
- Videos que me gust...
- Mostrar más
- Suscripciones
- NoticiasUnoColo...
- Colectivo Millenn...
- Daniel Coronell
- Shark Tank Colo...
- DW Documental
- Armagedda
- Javier Traba Pistolas
- Mostrar 68 más
- Explorar
- Tendencias
- Música

**Main Content:**

**Filters:**

**All Courses & Tutorials Online**  
Anuncio - Codecademy  
Join The Millions Learning To Code With Codecademy For Free  
[Learn more](#)

**Data Analytics**  
Use data to answer questions

**StatQuest with Josh Starmer**  
832,500 suscriptores • 227 videos  
Statistics, Machine Learning and Data Science can sometimes seem like very scary topics, but since each technique is really just ...  
[Suscribirse](#)

**Lo último de StatQuest with Josh Starmer**

**Long Short-Term Memory (LSTM)...**  
...Clearly Explained!!!  
25,631 visualizaciones • hace 9 días  
StatQuest with Josh Starmer  
Basic recurrent neural networks are great, because they can handle different amounts of sequential data, but even relatively small ...  
[Substituir](#)

**StatQuest Live Stream!!!**  
Cosine Similarity for NLP and CatBoost!!!  
5,728 visualizaciones • Emitido hace 2 semanas  
StatQuest with Josh Starmer  
On the last live stream I said I'd keep talk about out CatBoost, and we will by talking about the Cosine Similarity. This metric is also ...

**The Cosine Similarity for NLP and CatBoost**  
5,728 visualizaciones • Emitido hace 2 semanas  
StatQuest with Josh Starmer  
On the last live stream I said I'd keep talk about out CatBoost, and we will by talking about the Cosine Similarity. This metric is also ...

statquest

**StatQuest with Josh Starmer**  
 @statquest  
 832.000 suscriptores

Unirme Suscrito

INICIO VIDEOS SHORTS EN DIRECTO LISTAS COMUNIDAD CANALES INFORMACIÓN

Nuestros miembros  
 ¡Gracias a todos los miembros del canal!

Ed S L K

Unirme

StatQuest ▶ Reproducir todo

**PCA...**  
**Step-by-Step!!!**  
  
 StatQuest: Principal Component Analysis (PCA)...  
 StatQuest with Josh Starmer  
 2 M de visualizaciones  
 • hace 4 años  
 Subtítulos

**Logistic Regression...**  
**...Clearly Explain**  
  
 StatQuest: Regresión Logística  
 StatQuest with Josh Starmer  
 1,5 M de visualizaciones  
 • hace 4 años  
 Subtítulos

**In Statistics...**  
**Probability is not Likelihood**  
  
 StatQuest with Josh Starmer  
 1 M de visualizaciones  
 • hace 4 años  
 Subtítulos

**Maximum Likelihood...**  
**...Clearly explained!!!!**  
  
 StatQuest with Josh Starmer  
 1 M de visualizaciones  
 • hace 5 años  
 Subtítulos

**p-values...**  
**...Clearly Explain**  
  
 StatQuest: P Values, clearly explained  
 StatQuest with Josh Starmer  
 891.563 visualizaciones  
 • hace 5 años  
 Subtítulos

**PCA Main Ideas...**  
**...in only 5 min**  
  
 StatQuest: ¡Ideas principales de PCA en solo 5 minutos!  
 StatQuest with Josh Starmer  
 877.663 visualizaciones  
 • hace 4 años  
 Subtítulos

statquest

INICIO VIDEOS SHORTS EN DIRECTO LISTAS COMUNIDAD CANALES INFORMACIÓN

**A Gentle Introduction to...**  
**...Machine Learning**  
  
 A Gentle Introduction to Machine Learning  
 StatQuest with Josh Starmer  
 601.989 visualizaciones  
 • hace 3 años  
 Subtítulos

**Cross Validation...**  
**...it's no big deal!!!**  
  
 Machine Learning Fundamentals: Cross...  
 StatQuest with Josh Starmer  
 739.070 visualizaciones  
 • hace 4 años  
 Subtítulos

**Machine Learning Fundamentals...**  
**The Confusion Mat**  
  
 Machine Learning Fundamentals: The...  
 StatQuest with Josh Starmer  
 432.737 visualizaciones  
 • hace 4 años  
 Subtítulos

**Machine Learning Fundamentals...**  
**...Sensitivity and Specificity**  
  
 Machine Learning Fundamentals: Sensitivity...  
 StatQuest with Josh Starmer  
 208.357 visualizaciones  
 • hace 2 años  
 Subtítulos

**Machine Learning Fundamentals...**  
**...Bias and Variance**  
  
 Fundamentos del aprendizaje automático: sesgo y varianza  
 StatQuest with Josh Starmer  
 901.501 visualizaciones  
 • hace 4 años  
 Subtítulos

**Entropy (for data science)**  
**...Clearly Explain**  
  
 Entropy (for data science) Clearly Explained!!!  
 StatQuest with Josh Starmer  
 336.281 visualizaciones  
 • hace 1 año  
 Subtítulos

Neural Networks / Deep Learning ▶ Reproducir todo

This playlist has everything you need to know about Neural Networks, from the basics, all the way to image classification with Convolutional Neural Networks.

**Neural Networks Clearly Explained!!!**  
**Look inside the black box**  
  
 Neural Networks Pt. 1: Inside the Black Box  
 StatQuest with Josh Starmer  
 415.435 visualizaciones  
 • hace 2 años  
 Subtítulos

**The Chain Rule...**  
**...Clearly Explain**  
  
 The Chain Rule  
 StatQuest with Josh Starmer  
 102.952 visualizaciones  
 • hace 2 años  
 Subtítulos

**Gradient Descent...**  
**...Step-by-Step**  
  
 Gradient Descent, Step-by-Step  
 StatQuest with Josh Starmer  
 833.072 visualizaciones  
 • hace 3 años  
 Subtítulos

**Backpropagation for Neural Networks...**  
**...Clearly Explain**  
  
 Neural Networks Pt. 2: Backpropagation Main Ideas  
 StatQuest with Josh Starmer  
 218.616 visualizaciones  
 • hace 2 años  
 Subtítulos

**Backpropagation Details...**  
**...Part 1**  
  
 Backpropagation Details Pt. 1: Optimizing 3 parameters...  
 StatQuest with Josh Starmer  
 83.320 visualizaciones  
 • hace 2 años  
 Subtítulos

**Backpropagation Details...**  
**...Part 2**  
  
 Backpropagation Details Pt. 2: Going bonkers with The...  
 StatQuest with Josh Starmer  
 55.593 visualizaciones  
 • hace 2 años  
 Subtítulos

Statistics Fundamentals ▶ Reproducir todo

These videos give you a general overview of statistics as well as a be a reference for statistical concepts.

**Histograms...**

**The Main Ideas behind Probability Distributions**

**The Normal Distribution...**

**Population and Estimated Parameters...**

**Calculating the Mean, Variance and Standard Deviation...**

**What is a mathematical model...**

## RECOMENDADOS



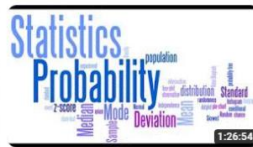




Modelo jerárquico Normal multinivel (parte 3)  
318 visualizaciones • hace 11 meses



Premio Fasescolda  
151 visualizaciones • hace 11 meses



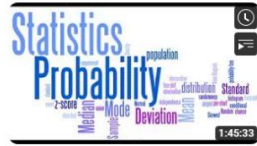
Variables aleatorias continuas  
139 visualizaciones • hace 11 meses



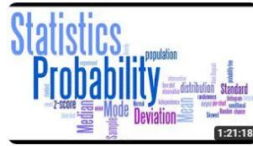
Modelo de bloques estocásticos  
101 visualizaciones • hace 11 meses



Modelo jerárquico Normal multinivel (parte 2)  
293 visualizaciones • hace 11 meses



Variables aleatorias discretas  
180 visualizaciones • hace 11 meses



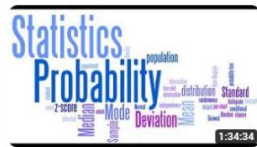
Teorema de Bayes e independencia estadística  
178 visualizaciones • hace 11 meses



Modelo jerárquico Normal multinivel (parte 1)  
413 visualizaciones • hace 11 meses



Modelos de grafos aleatorios exponenciales (parte 2)  
87 visualizaciones • hace 11 meses



Reglas de probabilidad  
175 visualizaciones • hace 11 meses



Modelos de grafos aleatorios exponenciales (parte 1)  
75 visualizaciones • hace 11 meses



Implementación del muestreador de Gibbs  
284 visualizaciones • hace 11 meses

## INTRODUCCION R

### Abrir mente en la CIENCIA DE DATOS

## Ecuación de la recta

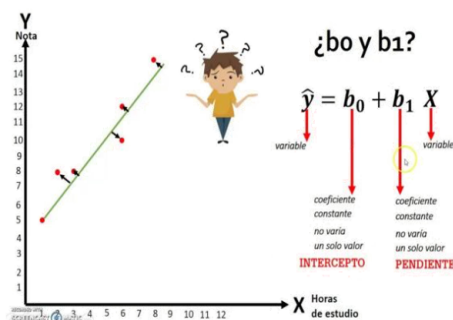
$$y - b = m(x - 0)$$

$$y - b = mx$$

$$y = mx + b$$

¿Por qué es relevante en la analítica predictiva?

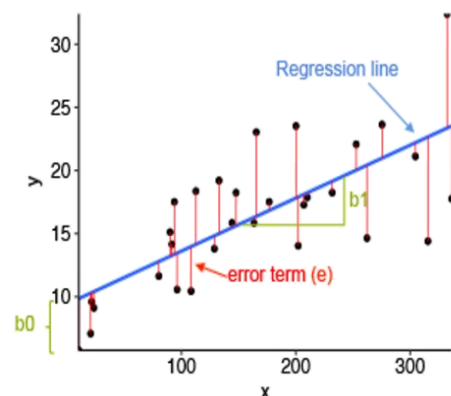
La ecuación de la recta, es la base matemática que permite definir un modelo lineal.



## Modelo de regresión lineal

El modelo de regresión lineal, en su forma simple, busca trazar una línea recta entre una nube de dispersión.

Edad en el eje X y Tasa de filtración glomerular en el eje Y



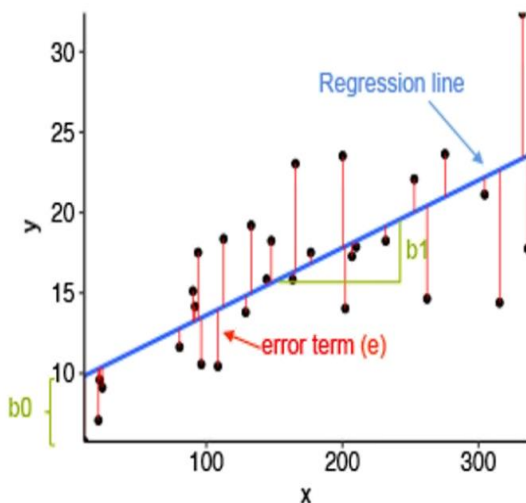
Nivel de satisfacción del cliente.

Niveles de datos generalizados

TARGET-----Objetivo

## Modelo de regresión lineal

- Es poco probable encontrar una única recta que una los diferentes puntos.
- No tener esa única unión de puntos nos genera el ERROR del modelo.



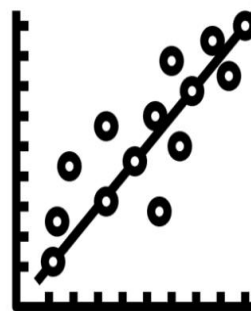
**EL MODELO PERFECTO NO EXISTE.**

R Cuadrado



## El error

- El error en cualquier modelo predictivo siempre esta presente.
- Hay que dudar de los modelos donde el error es  $= 0$ , la "perfección" es imposible de encontrar.



El error en el modelo lineal, es la distancia que se encuentra entre la línea recta calculada y cada uno de los puntos, la sumatoria de estas distancias nos dan como resultado final, el error del modelo.

## El error

- El calculo del error de un modelo se puede construir a partir de diferentes igualdades, todas estas se enmarcan en lo que conocemos como funciones de costo.
- En nuestro diario vivir, siempre queremos bajar los costos, para ser mas rentables, tener mayores utilidades.
- Las utilidades se traducen en capacidad, estabilidad del modelo.



## Sobre ajuste / Sub ajuste

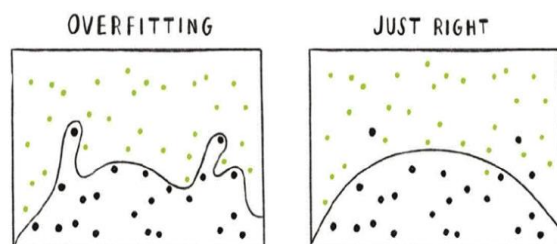
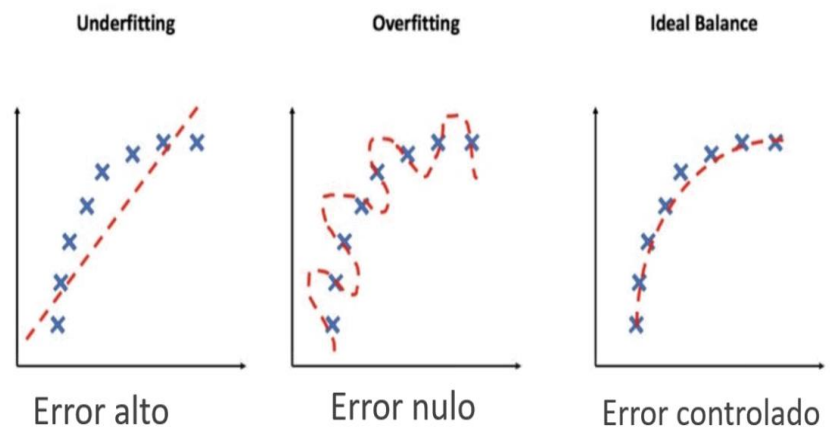
Trabajando sobre las funciones de costo de cualquiera de los modelos nos encontraremos con dos conceptos importantes:

- Sobre ajuste: en términos sencillos es el comportamiento de un modelo que dice tanta verdad, que finalmente son mentiras.



- Sub ajuste: a diferencia del anterior, el no tiene la capacidad de decir la verdad, razón por la cual desde un principio solo dice mentiras.

DIARIO VIVIR- error gestionar costo.

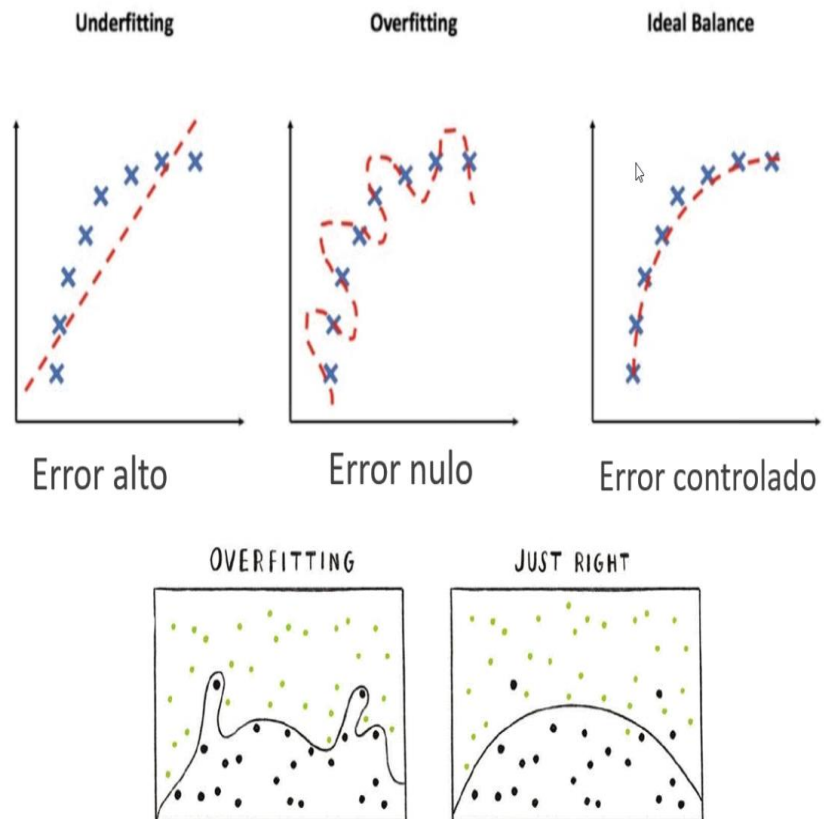




Costo mínimo > Rentabilidad

Subajuste. Talla 40 me gustaron los talla 37

Sobreajuste. Talla 40 compro 47

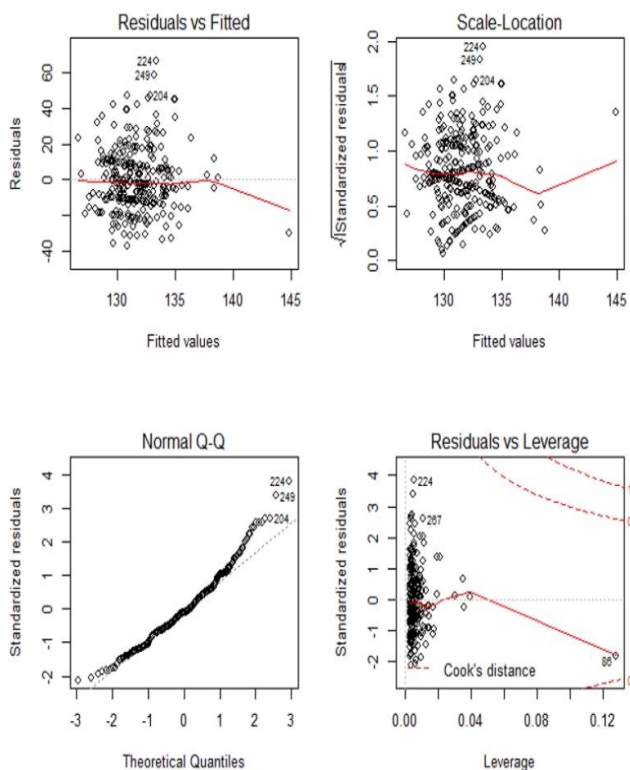


Listado de cliente. SAV

Nuevo jabón de la ropa. Sobre ajustado unos si otros no fuga de cliente.

Subajustado no aprendió. No sirvió.

# RESULTADOS



A partir de los graficos se pueden validar los supuestos de los modelos, los dos primeros graficos debe mostrar una simetría por encima y por debajo de la línea roja y una buena dispersión de los datos a lo largo del eje x.

El grafico de Normal Q-Q debe mostrar un comportamiento normal, es decir la mayoría de los datos deben estar sobre la línea recta para que cumpla el supuesto de normalidad.

El grafico Scale-Location es muy similar a los residuos frente a los ajustados, pero simplifica el análisis del supuesto de homocedasticidad. Toma la raíz cuadrada del valor absoluto de los residuos estandarizados en lugar de trazar los propios residuos. Recuerde que la homocedasticidad significa varianza constante en la regresión lineal.

Terminologías técnicas.

No es solo correr código. Capacidad del modelo.

## 1 Selección de variables

- Selección exhaustiva ( $2^p$ )
- Selección secuencial inteligente (*forward, backward, stepwise*)

## 2 Penalización

- Ridge
- Lasso
- otras penalizaciones

## 3 Reducción de dimensionalidad y transformación de variables

- Regresión de componentes principales
- Regresión de mínimos cuadrados parciales
- otras formas de transformación y reducción de variables



REDES NEURONALES. Disminuir mal aprendizaje.

Ayudas de la capacidad y gestionar un modelo. Para que sea mejor.

Penalización. Modelos lineales para mejores variables, intersección

## MEDIDAS DE CALIDAD DEL MODELO:

$R^2$

**Coeficiente de determinación**, es un estadístico usado en el contexto de un modelo estadístico cuyo principal propósito es predecir futuros resultados o probar una hipótesis. El coeficiente determina la calidad del modelo para replicar los resultados, y la proporción de variación de los resultados que puede explicarse por el modelo.

AIC

El criterio de información de Akaike (AIC) es una medida de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Como tal, el AIC proporciona un medio para la selección del modelo.

AUC

Area bajo la curva, medida estadística que permite observar el poder de clasificación de un modelo, donde la variable respuesta es de tipo categorica.

GINI

## Matriz de confusión

		Clasificación	
		Positivo	Negativo
Verdad Terreno	Positivo	Positivos Ciertos	Negativos Falsos
	Negativo	Positivos Falsos	Negativos Ciertos

“Si seleccionamos aleatoriamente dos items de una población, entonces estos deben ser de la misma clase y la probabilidad de esto es 1 si la población es pura”.

- 1.Variable objetivo categórica: “Success” o “Failure”
- 2.Solo divisiones binarias
- 3.A mayor valor de índice Gini, mayor la homogeneidad

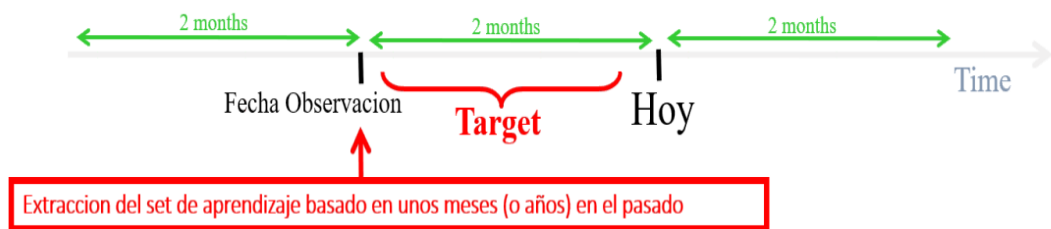
Evaluar capacidad del modelo.

Métricas de evaluación de los modelos.

**APRENDIZAJE POR RESFUERZO:** Tarea repetitiva (Redes Neuronales)  
Animal premio

Para crear un modelo predictivo:

- Un set de APRENDIZAJE (perfil al tiempo de observación)
- Un TARGET: los que no pagaron DESPUES del tiempo de observación



**El modelo debe trascender – Aprendizaje Automático.**

Ej. Qué tipo de moto debo vender 2022

Retroalimentar el modelo.

Proceso de calibración del modelo

Gestionar registros nuevos – ajustes de parámetros del modelo.

**CASO DE Recursos Humanos.**

Encuestas Estructuradas, satisfacción del colaborador, clima laboral.

¿Qué tan dispuesto está a irse a la Compañía? – Riesgo de fuga de colaboradores. Preventivas y correctivas. Sin nombre

**NO ASUSTAR ES EMOCIONAR. Implementación para  
basarnos**




Casos de negocios y herramientas

DATOS ESPACIALES.

Libros recomendados.

maths behind machine learning algorithms

X



Todo

Videos

Imágenes

Noticias


Shopping

Más

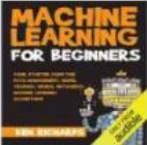
Herramientas

Cerca de 154.000.000 resultados (0,87 segundos)

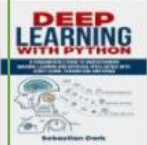
Anuncios · Comprar maths behind machine learning algorithms




Machine Learning...  
\$ 73.872,88 ...  
US\$ 14,95 ...  
Audible.com




Your Starter Guide for D...  
\$ 34.342,24 ...  
US\$ 6,95 m...  
Audible.com




Deep Learning Python: ...  
\$ 34.342,24 ...  
US\$ 6,95 m...  
Audible.com



Machine Learning...  
\$ 34.342,24 ...  
US\$ 6,95 m...  
Audible.com






Grokking Machine...  
\$ 123.286,1 ...  
US\$ 24,95 ...  
Audible.com



Algorithms: The Comple...  
\$ 98.679,53 ...  
US\$ 19,95 ...  
Audible.com

interpretable machine learning

X



Todo

Imágenes

Videos

Libros

Noticias

Más

Herramientas

Cerca de 18.100.000 resultados (0,60 segundos)

Anuncio · <https://www.mygreatlearning.com/>

MIT IDSS - 12 Week Program - Machine Learning Algorithms

Learn advanced Data Science & Machine Learning skills from world-renowned MIT faculty.

Data Science & ML Program

For Budding Data Scientists Recorded Lectures by MIT Faculty

What You'll Learn

Data Science Program Machine Learning Program

Program Benefits

Learn Data Science & ML Personalized Mentorship and Support

Program Faculty

Learn from MIT Program faculty Learn From the Best

Anuncio · <https://professionalonline2.mit.edu/>

Interpretable Machine Learning: A Guide for Making Black Box Models Explainable

Libro

¿Te gustó este libro?

Like

Dislike

Fecha de publicación original: 2019

Autor: Christoph Molnar

Obtener libro



## Machine Learning For Dummies

Libro



¿Te gustó este libro?



**Fecha de publicación original:** 10 de mayo de 2016

**Autores:** [Luca Massaron](#), [John Mueller](#)

Obtener libro



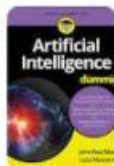
Busca en una biblioteca  
Buscar en WorldCat

Buscar

También se buscó



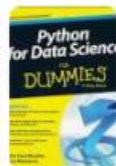
Deep  
Learning  
For Du...  
Luca Ma...



Artificial  
Intellig...  
For Du...  
Luca Ma...



Algorith...  
For  
Dummies  
Luca Ma...



Python  
for Data  
Scienc...  
John Mu...

Más información sobre Machine L... →

Comentarios

machine learning for dummies



Todo

Videos

Imágenes

Shopping

Libros

Más

Herramientas

Cerca de 1.810.000.000 resultados (0,62 segundos)