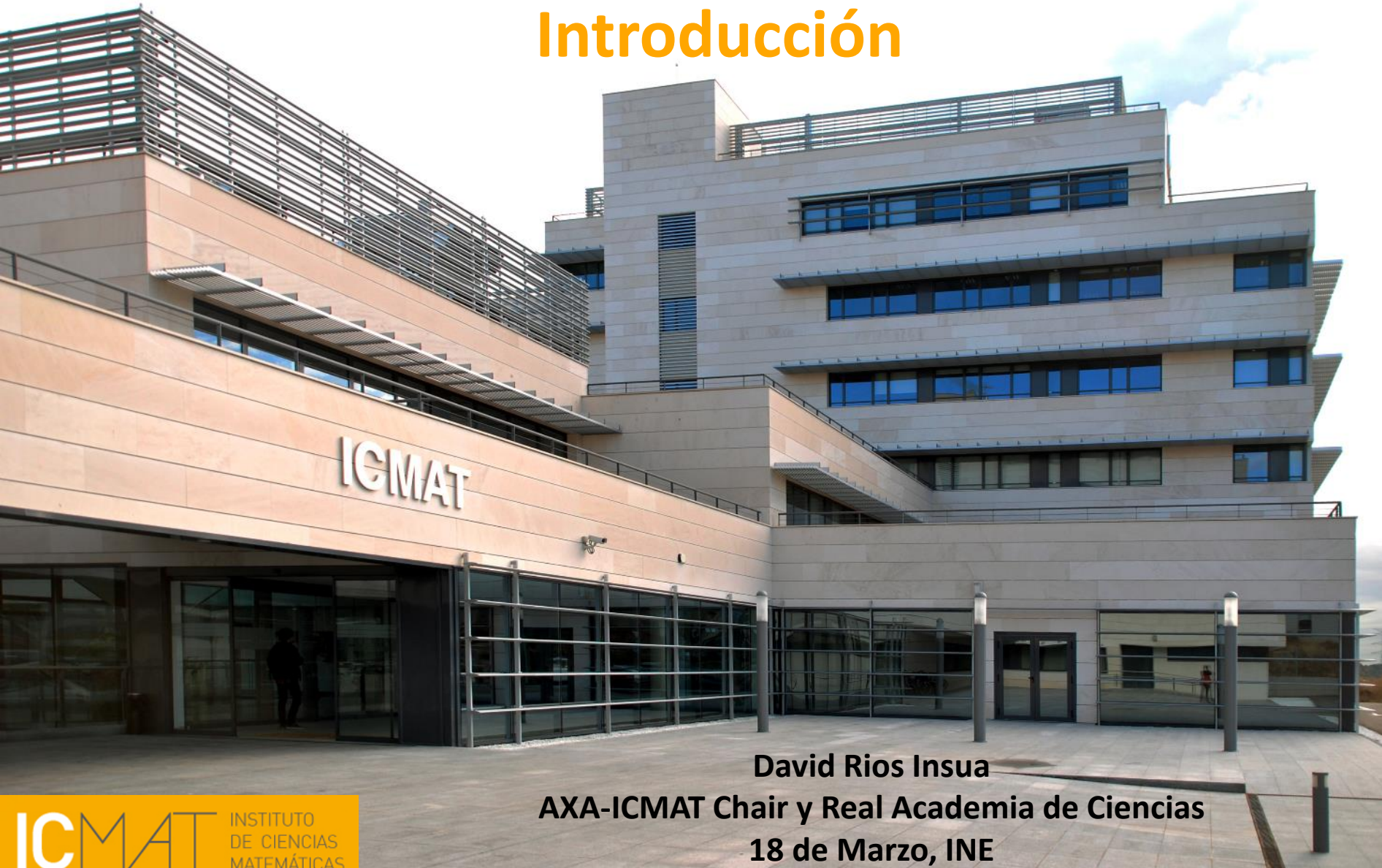


# Curso de Aprendizaje Automático para el INE Introducción



David Rios Insua

AXA-ICMAT Chair y Real Academia de Ciencias

18 de Marzo, INE

# Agenda

- Estadística vs Aprendizaje Automático?
- Contenidos
- Bibliografía básica
- Paquetes básicos a emplear
- Profesores

# Estadística vs Aprendizaje Automático?

- Aprendizaje automático, estadística de toda la vida revestida
- When you're fundraising, it's AI. When you're hiring, it's ML. When you're implementing, it's logistic regression
- Aprendizaje estadístico
- Aprendizaje automático es estadística glorificada

# Estadística vs Aprendizaje Automático

Deep Learning Book

Álgebra Lineal Numérica

Probabilidad y Teoría de la Información

Optimización

Conceptos básicos de AA

Sobreajuste, EMV, Estadística Bayesiana, SGD, Algoritmia, ...

Redes profundas

# Estadística vs Aprendizaje Automático

## Jerga diferente

Estadística	Aprendizaje automático
Estimación	Aprendizaje
Dato	Ejemplo
Regresión	Aprendizaje supervisado (r. continua)
Clasificación	Aprendizaje supervisado (r. discreta)
Respuesta	Etiqueta
Regresor	Rasgo (Variable predictora)
....	...

# Estadística vs Aprendizaje automático?

Algunas diferencias (matizables)

Estadística	Aprendizaje automático
Énfasis en inferencia	Énfasis en predicción
Cuantificar incertidumbre	Optimizar
Entender el principio subyacente	Predecir con precisión
Matemáticas	Computación
Mayor intervención humana	Mayor intervención de la máquina

Statistics in the Big Data era: Failures of the Machine

<https://www.sciencedirect.com/science/article/pii/S0167715218300737>

# Aprendizaje automático

“Un programa de ordenador aprende de la experiencia  $E$  con respecto a la clase de tareas  $T$  y la medida de comportamiento  $P$ , si su comportamiento con respecto a las tareas en  $T$ , medida según  $P$ , mejora con la experiencia  $E$ .”

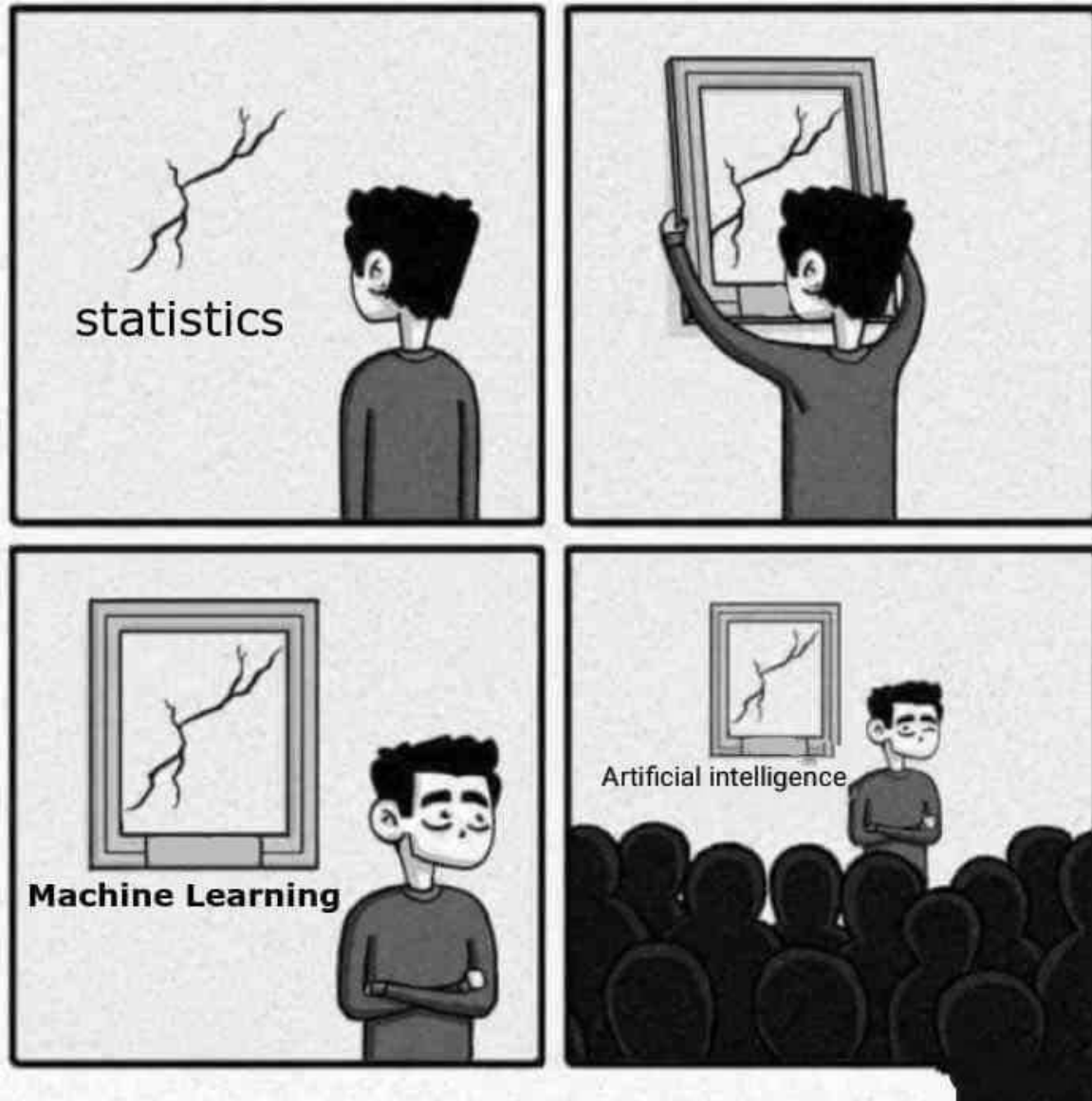
Representación-Evaluación-Optimización

# Aprendizaje automático: Tareas

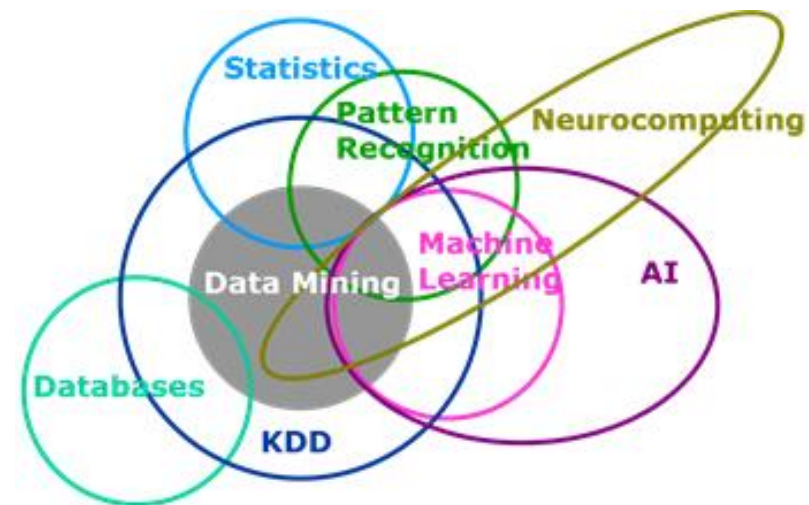
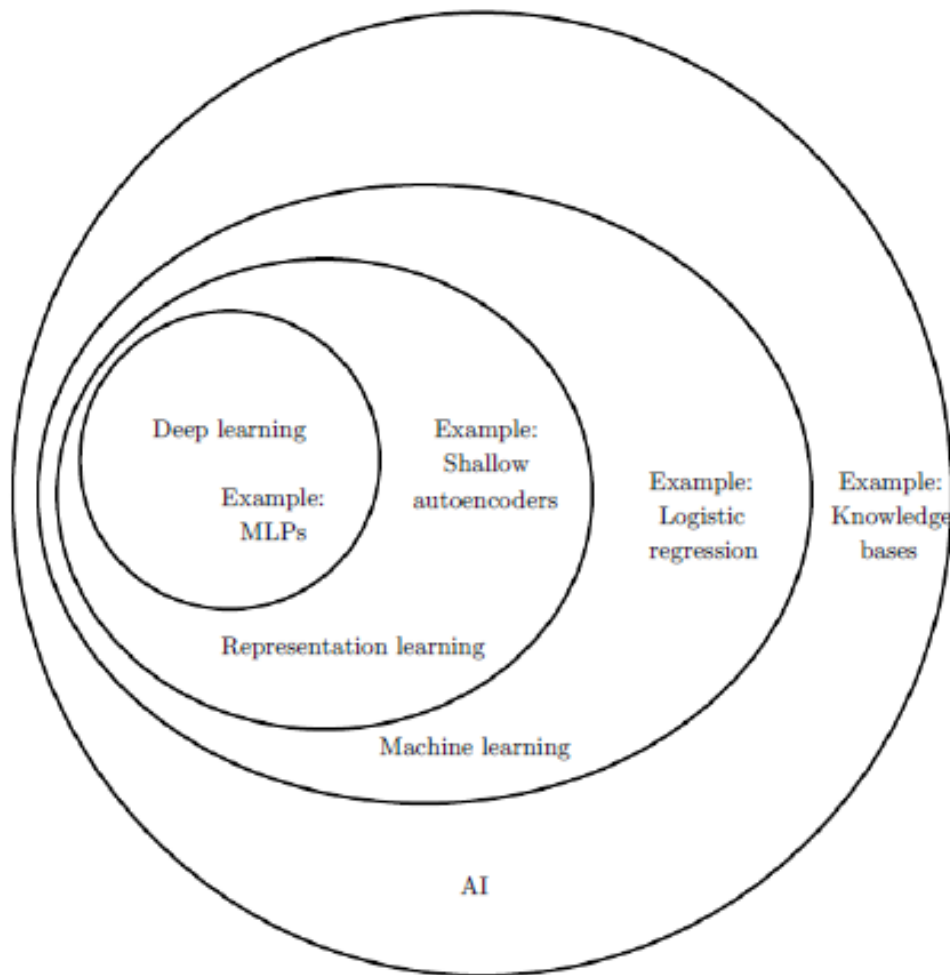
- Clasificación
- Clasificación con entradas faltantes
- Regresión
- Transcripción
- Traducción automática
- Salida estructurada
- Detección de anomalías
- Síntesis y muestreo
- Imputación
- Filtrado
- Estimación de densidades



# Estadística vs Aprendizaje Automático?



# Estadística vs Aprendizaje Automático?



# Estadística vs Aprendizaje Automático?

La Inteligencia Artificial fue definida por J. McCarthy en 1956, como *“la ciencia e ingeniería de hacer máquinas que se comporten de una forma que llamaríamos inteligente si el humano tuviese ese comportamiento”*. La IA es un área de la Informática y comparte algunas técnicas con otras disciplinas, como las Matemáticas y la Estadística o la Ciencia Cognitiva.

Instituto Nacional del Dato ...

# Aprendizaje automático y estadística oficial

<https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-017-0099-3>

<https://www.liebertpub.com/doi/pdfplus/10.1089/big.2013.0038>

[https://www.researchgate.net/profile/Siu\\_Ming\\_Tam/publication/280972848\\_Big\\_Data\\_Official\\_Statistics\\_and\\_Some\\_Initiatives\\_by\\_the\\_Australian\\_Bureau\\_of\\_Statistics/links/5783745408ae37d3af6bed61/Big-Data-Official-Statistics-and-Some-Initiatives-by-the-Australian-Bureau-of-Statistics.pdf](https://www.researchgate.net/profile/Siu_Ming_Tam/publication/280972848_Big_Data_Official_Statistics_and_Some_Initiatives_by_the_Australian_Bureau_of_Statistics/links/5783745408ae37d3af6bed61/Big-Data-Official-Statistics-and-Some-Initiatives-by-the-Australian-Bureau-of-Statistics.pdf)

<https://journals.sagepub.com/doi/pdf/10.1177/2053951714538417>

<http://mural.maynoothuniversity.ie/7276/1/stat>

[http://www.q2014.at/fileadmin/user\\_upload/ESTAT-Q2014-BigDataOS-v1a.pdf](http://www.q2014.at/fileadmin/user_upload/ESTAT-Q2014-BigDataOS-v1a.pdf)

# Contenidos

## Fundamentos aprendizaje automático

- Conceptos generales sobre aprendizaje supervisado
- Conceptos generales sobre aprendizaje no supervisado
- Conceptos generales de aprendizaje bayesiano y simulación Montecarlo

# Contenidos

## Aplicaciones prácticas aprendizaje automático

- Algoritmos de regresión
- Algoritmos de clasificación
- Árboles de decisión, bosques aleatorios, boosting
- Máquinas de soporte vectorial
- Análisis de cluster
- Modelos gráficos probabilísticos
- Regresión y clasificación bayesiana
- Modelos dinámicos lineales

# Contenidos

## Aprendizaje profundo

- Redes neuronales
- Redes neuronales profundas

# Bibliografía básica

- James, Witten, Hastie, Tibshirani. Introduction to Statistical Learning with R (ISLR)
- Bishop. Pattern Recognition and Machine Learning (Bishop)
- Hastie, Tibshirani, Friedman. Elements of Statistical Learning (ESL)
- Gelman, Carlin, Stern, Dunson, Vehtari, Rubin. Bayesian Data Analysis (BDA)
- French, Ríos Insua. Statistical Decision Theory (SDT)
- Ríos Insua, Martín, Jiménez. Simulación (SIM)



# Paquetes I

## **Importar y transformar datos**

- dplyr: filtrar filas, seleccionar columnas, crear nuevas columnas
- tidyr: transformar data.frames
- readr: importar datos
- stringr: manipular texto
- lubridate: fechas
- forcats: factores
- ggplot2: visualización

# Paquetes II

## **Herramientas de entrenamiento, validación cruzada, búsqueda de parámetros**

- caret: búsqueda de hyper-parametros, validación cruzada, partición train-test
- mlr: similar a caret
- rBayesianOptimization: bayesian optimization
- modelMetrics: métricas
- modelr: funciones auxiliares para modelizar
- broom: extraer y ordenar info de los modelos

# Paquetes III

## Modelos

- glmnet: penalized regression (Lasso, ElasticNet)
- glmnetUtils: interfaz de formula para glmnet
- MASS: ridge regression
- nnet: multinomial logistic regression
- ridge: ridge regression con selección automática de hiper parámetros
- rpart: arboles de decision
- randomForest: randomForest
- e1071: SVM
- gbm: gradient boosting
- xgboost: extreme gradient boosting
- kernlab: procesos gaussianos, reducción de dimensionalidad
- dlm: modelos dinámicos lineales

# Profesores



David Ríos Insua [david.rios@icmat.es](mailto:david.rios@icmat.es)



Victor Gallego [victor.gallego@icmat.es](mailto:victor.gallego@icmat.es)



Roi Naveiro [roi.naveiro@icmat.es](mailto:roi.naveiro@icmat.es)



Alberto Torres [alberto.torres@icmat.es](mailto:alberto.torres@icmat.es)

# Profesores

- <https://www.icmat.es/spor/>
- <https://datalab.icmat.es/>