

Schedule

State of the course

Lesson 4 Review

Challenge

Notebook + resources

madrid@saturdays.ai - [@aisaturdayses](https://twitter.com/aisaturdayses)

State of the course

- Lesson 1 Cleaning & Exploratory Data Pandas
- Lesson 2 Linear Regression & Decision Trees
- Lesson 3 FI, Random Forest Deep Dive
- Lesson 4 Unsupervised Learning
- Split
 - PCA
 - Clustering
 - Future...



Today!

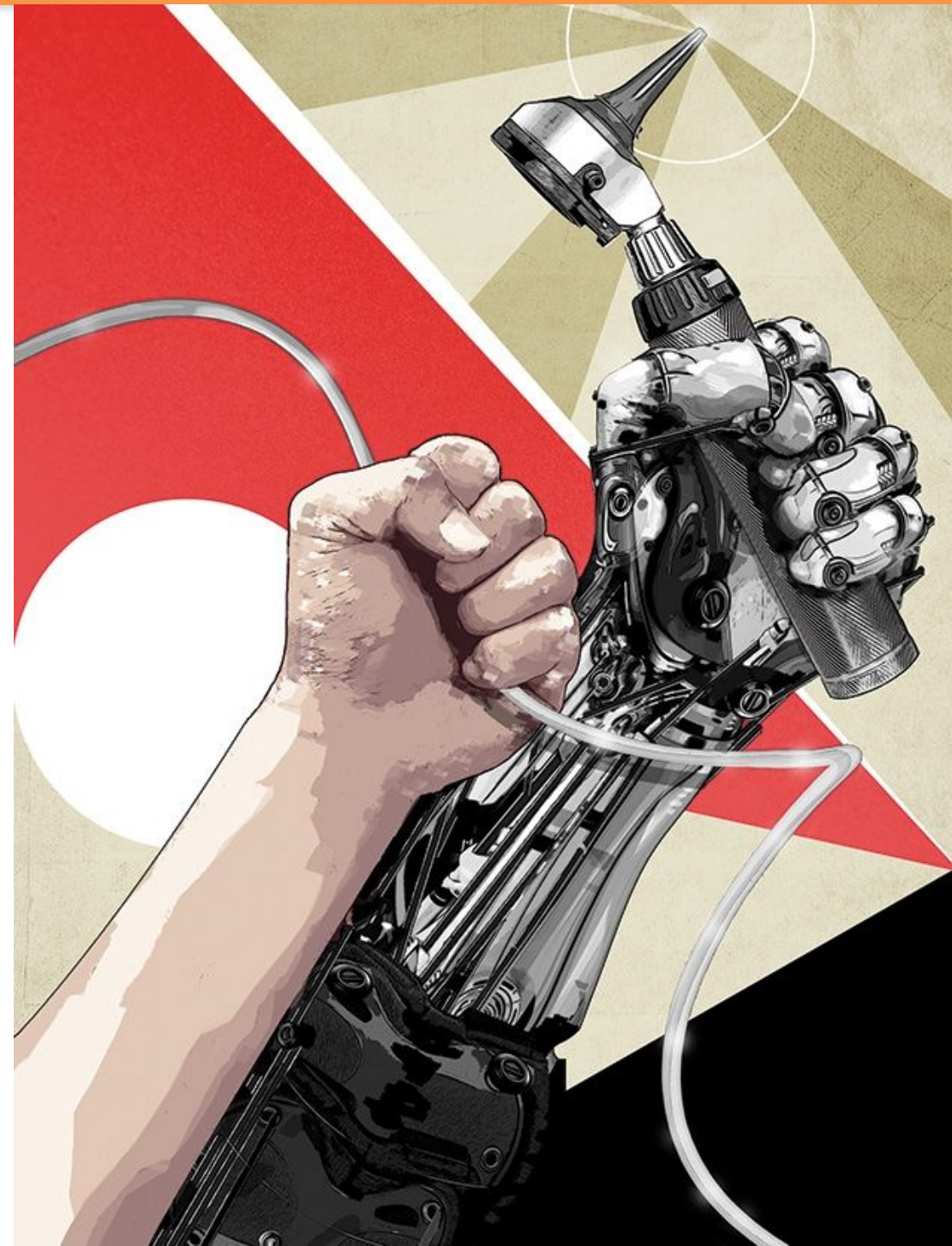
Agenda:

9:30 Machine Learning Lesson 4 + Work

11:30 Coffee break

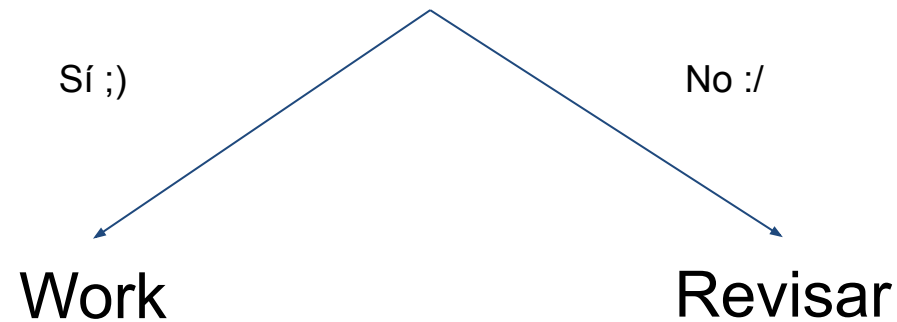
11:45 Work

13:30 ¿¿??



¿Cómo aprendemos?

Revisar y entender la documentación



Jugar con los
datos



Aprendizaje
cooperativo



Challenge

Cómo preparar la 3ª Sesión

Atentos a las indicaciones para preparar este sábado:

 Para ML, #3 - FI, Random Forest Deep Dive.

En esta tercera semana vamos a repasar conceptos de ML, y meternos de lleno en Random Forest (RF). Ya habéis experimentado, cada vez entenderéis más lo que estás practicando y, en concreto, RF es la mejor opción para continuar porque vamos a seguir trabajando sobre los mismos conceptos.

Para trabajar esta sesión vamos a ver las dos primeras lecciones de fast.ai. La idea es seguir el notebook que os dejamos mientras seguimos también ambos videos.

- [Notebook](#) para seguir junto con estos videos
- [Video 1](#)
- [Video 2](#)
- [Tutorial Feature Importance](#)

Tendremos preparados 3 ejercicios para este sábado y un Kahoot 

Slack

- Dudas Técnicas
- Seguimiento semanal
- Recursos
- Proyectos
- Contacto con la Comunidad

AI Saturdays Hi... ▾ 🔔

● Pablo Castañeda

🔍 Ir a...

Hilos de conversaciones

Canales +

- # aisaturdays_health
- # aleatorio
- 🔒 # ambassadorsmadrid
- # anaconda-support
- # cloud-setup
- # deep_learning
- 🔒 # facilitators_all
- # general
- 🔒 # hispanic_ambassadors
- # jobs
- # machine_learning
- # madrid_**
- 🔒 # notebooks_fastai_ml
- # presentations
- # proyectos_madrid
- # rl-studygroup

#madrid_

☆ | 👤 24 | 🗑️ 0 | ✎ Añadir un tema



🔍 Búsqueda



Scholarship
for women
in AI

Next cohort starting
October 14, Barcelona

APPLY NOW

Master
Arti
Inte



Pablo Castañeda 20:34

En cuanto a esta semana en ML. La idea es entender como se monta un modelo con sci-kit learn (libreria de ML)

Ver el problema de regresión de forma básica con regresiones lineales

Podéis ver primer cuales son los conceptos que afectan a un problema de regresión.

Captura de pantalla 2019-09-16 a las 20.35.16.png ▾



Mi recomendación, queréis entender bien esos conceptos, buscar algun video corto en youtube



Veréis que hay mucho código, intentad entenderlo. Ya sabéis como funciona, os ayudamos con notebooks para que lo estudies entre semana y el sábado os lanzamos un reto



Pablo Castañeda 21:28

Unsupervised
Learning

The future of machine learning is unsupervised learning.



*Supervised
learning is the
icing on the
cake*

*Unsupervised
learning is the
cake itself*

Humans learn mostly through unsupervised learning: we absorb vast amounts of data from our surroundings without needing a label.

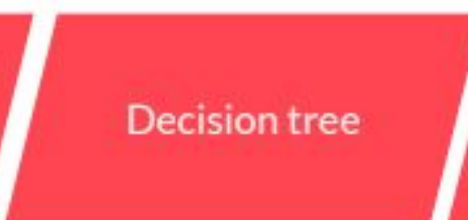
To reach true machine intelligence (i.e., a machine that thinks and learns for itself), ML needs to get better at **unsupervised** learning - it should learn without us having to feed it labels or explicit instructions.

We will have only scratched the surface in this class.



¿Y ahora?

Where are we?



We have finished discussing supervised algorithms and now we will introduce and discuss unsupervised algorithms. This will help lay the foundation for our discussion of natural language processing.



K-Means

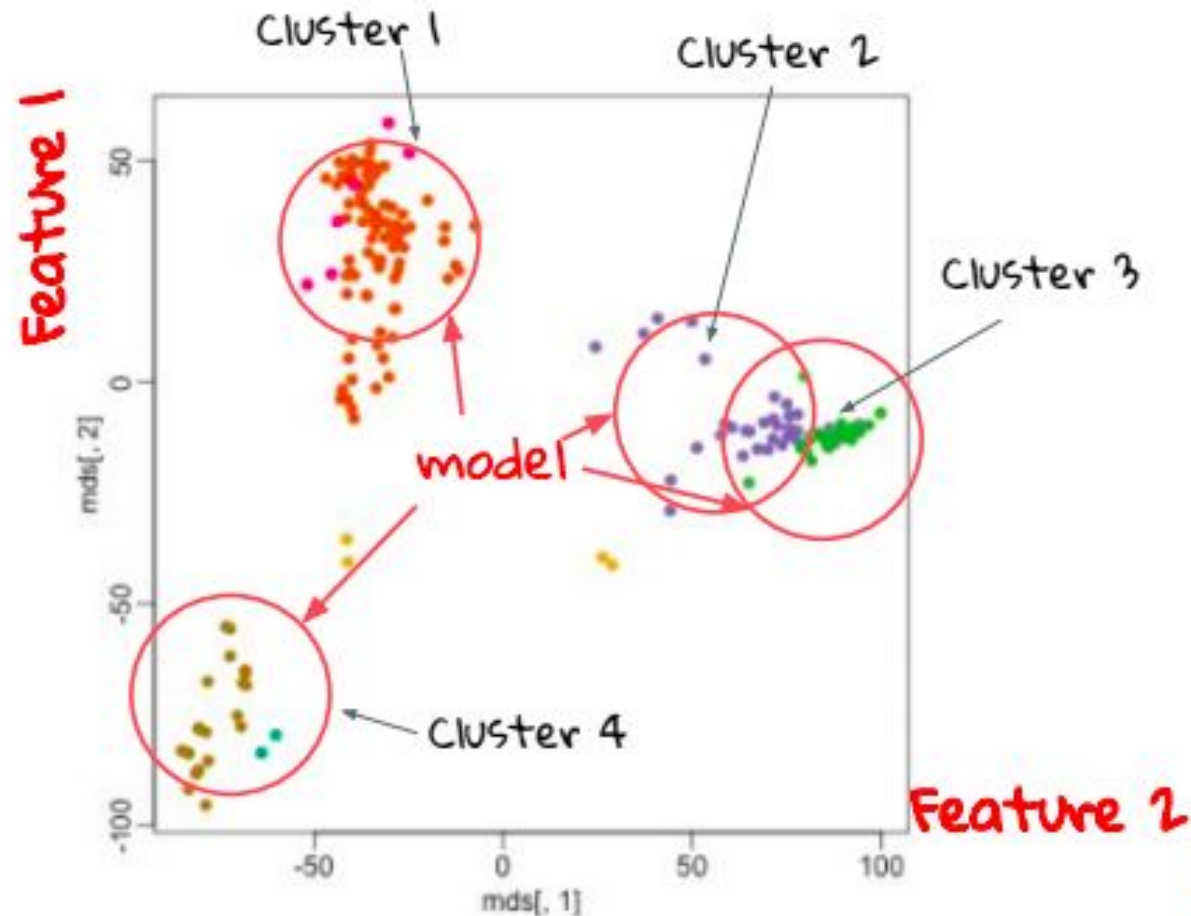
Clustering Task

Clustering is a powerful unsupervised algorithm that detects naturally occurring patterns in the data.

Clustering splits data in order to find out how observations are similar on a number of different features.

We are not predicting a true Y .

The clusters are the model. We decide the number of clusters, represented as K .



What is Principal Component Analysis?

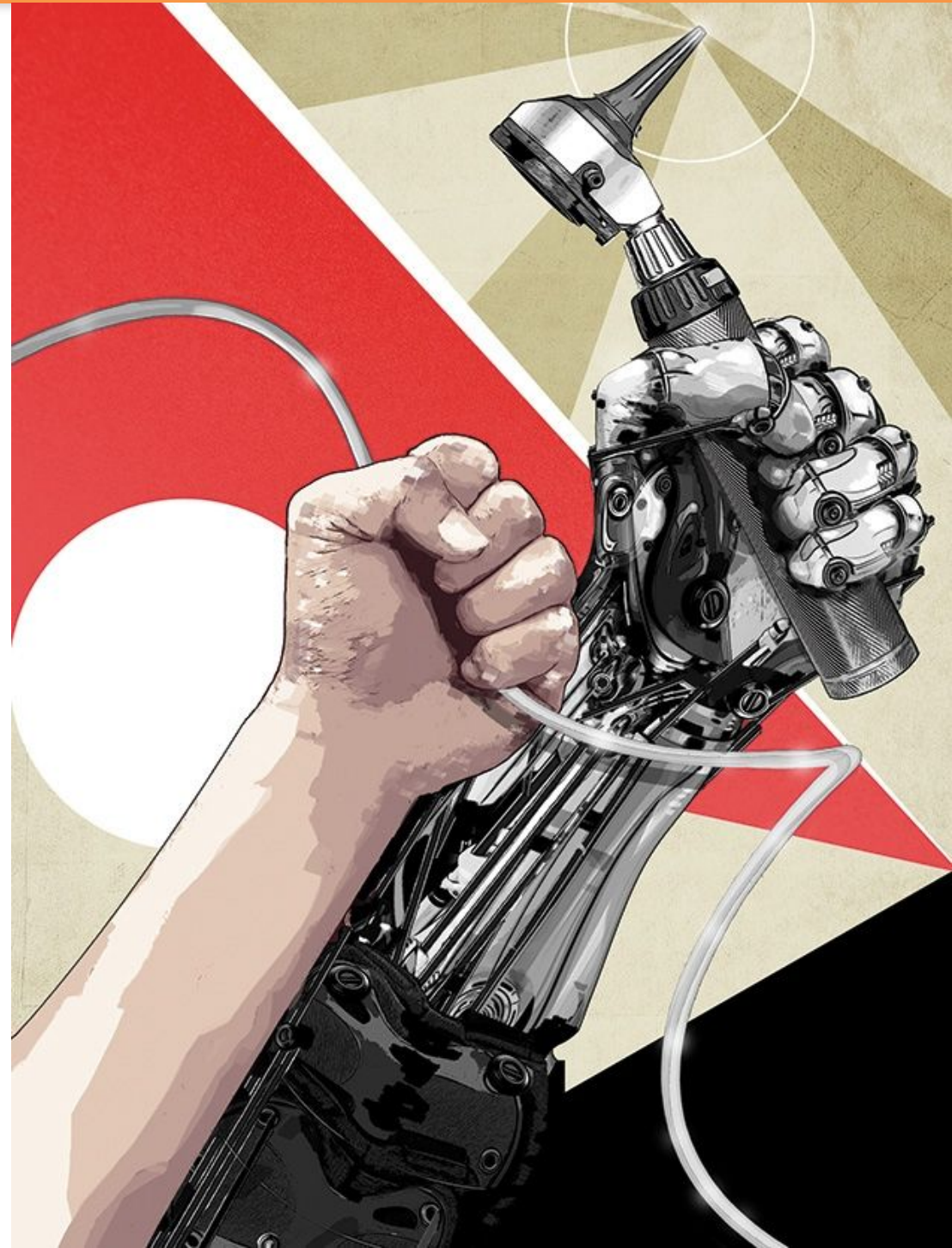
PCA finds which features are most correlated in a dataset, and removes them, leaving you with **the most “important” features** - i.e., the “principal components.”

PCA is helpful for feature selection and engineering



Work!

- Review the notebooks from the lecture.
- Challenge -> Apply your model
- Sigue los conceptos paso a paso



Other datasets

- Google Collab: instalar el paquete kaggle-cli

```
!pip install kaggle-cli  
# always use ! to run bash commands from Notebook
```

- Obtener los datos escribiendo:

```
!kg download -u <<Kaggle UserName>> -p <<Kaggle Password>> -c  
bluebook-for-bulldozers -f Train.zip
```

- Extraerlos y organizarlos

```
!mkdir -p data/bulldozers/  
!mv Train.zip data/bulldozers/  
!unzip data/bulldozers/Train.zip -d data/bulldozers/
```

- Google Collab: Utilizamos !wget para descargar el archivo de un repositorio, y !tar para descomprimirlo

```
!wget  
https://raw.githubusercontent.com/Giffy/Personal\_dataset\_repository/  
master/train.tar.gz  
!tar xvf train.tar.gz
```

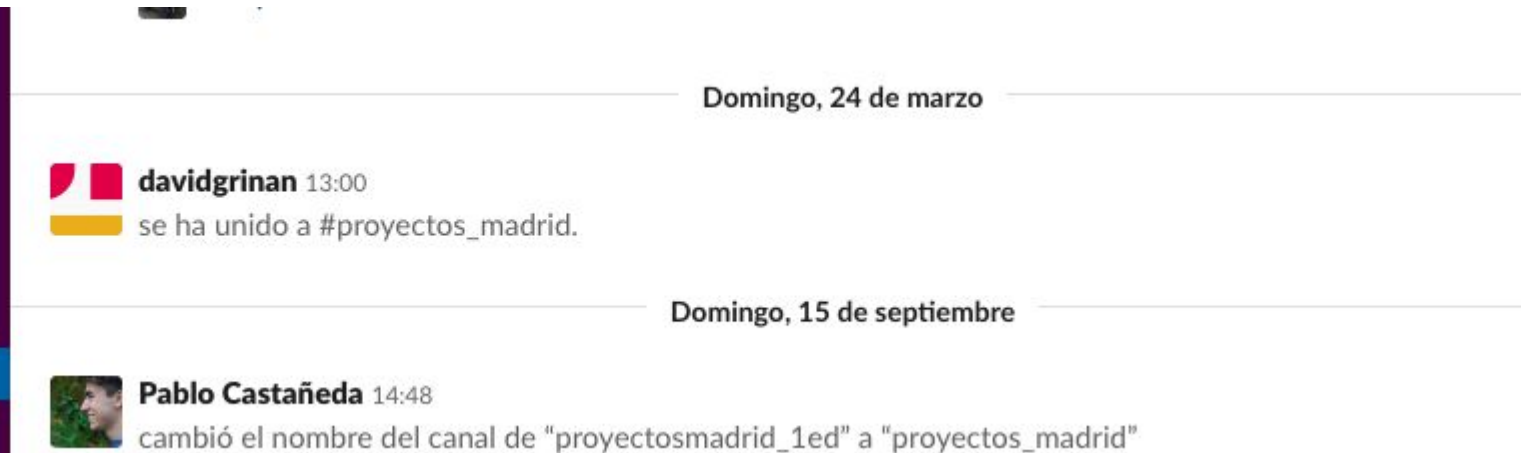
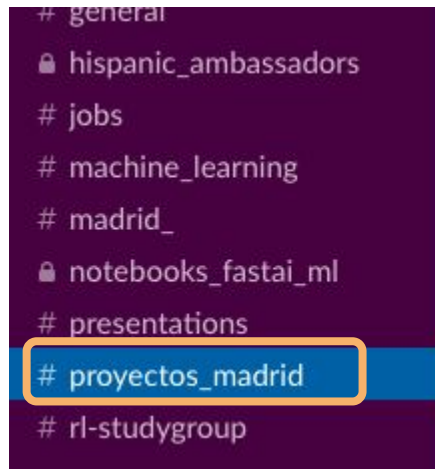
- Jupyter notebook: Descargas los archivos de Kaggle, decompreses el archivo en data/bulldozers (Debes crearla)

Your Projects

Hemos abierto ese canal para:

- Ideas
- Sugerencias
- Datasets
- etc.

Únete Ahora



BE SOCIAL!



@AISaturdaysES



@aisaturdays_madrid



@AISaturdaysES



Saturdays.AI



#AISaturdaysMadrid

madrid@saturdays.ai

LOOM_Guest -> Bienvenidos!
LOOM Princesa -> LoomPr1nc3sa

