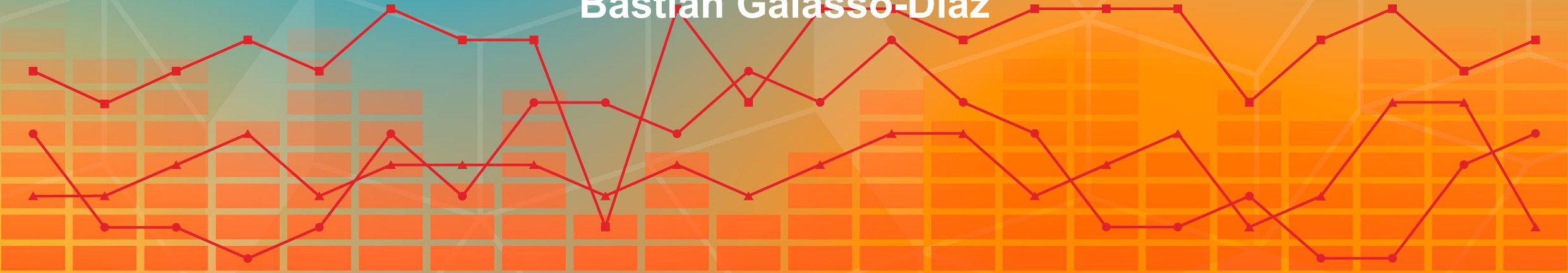


STATS STAGE

Intro al Machine Learning

Bastian Galasso-Díaz





PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS



STATS STAGE

Primero, ¡conozcámonos un poquito!





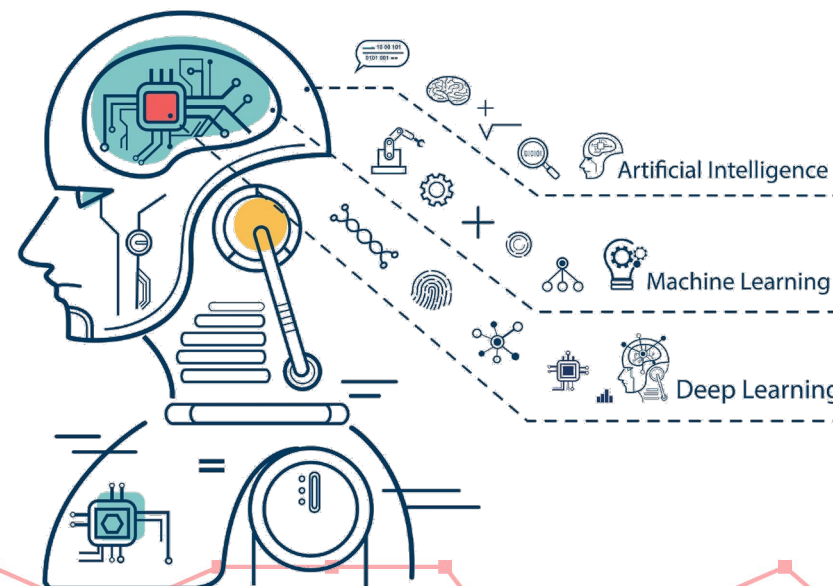
PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

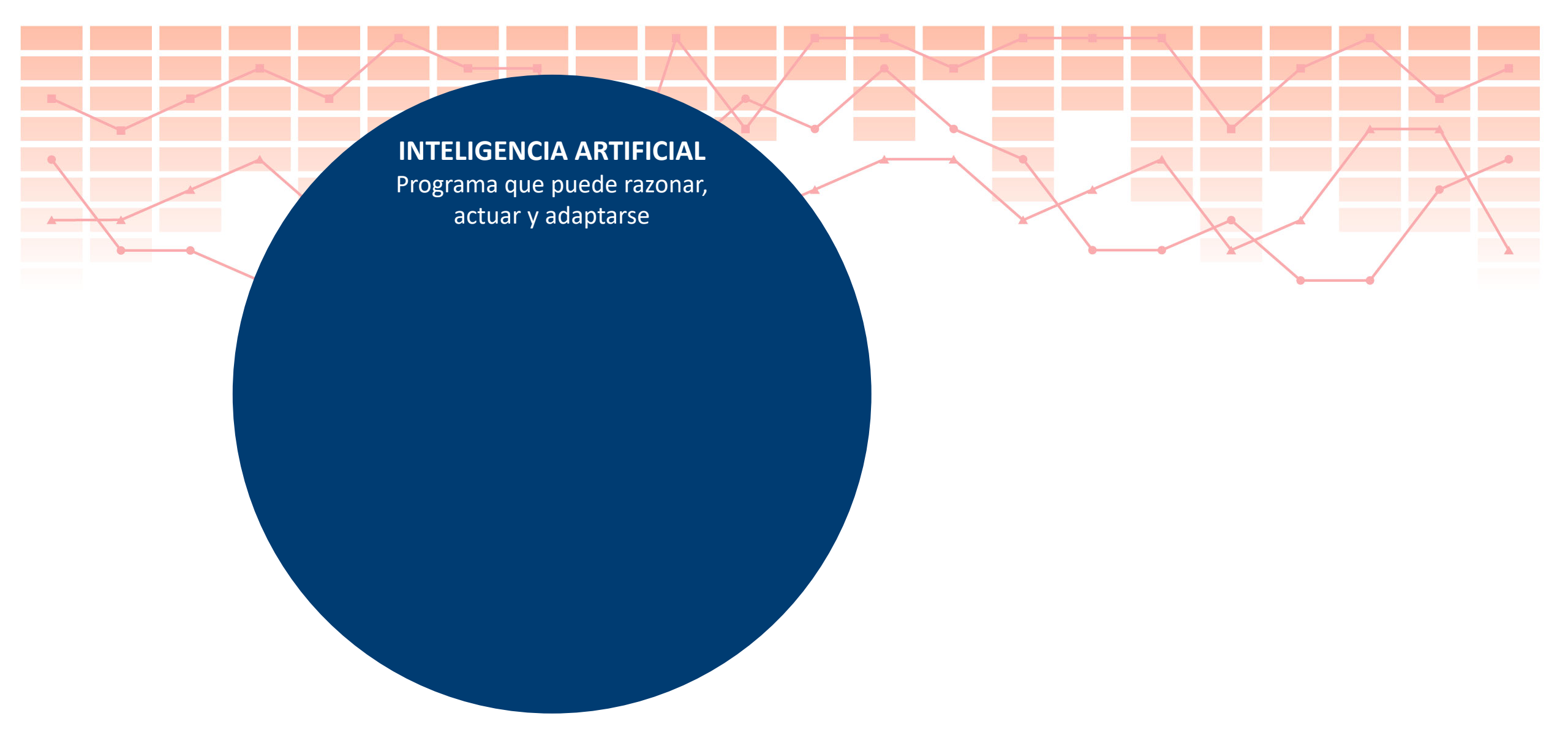


STATS STAGE

¿Qué es el Machine Learning?

El uso y desarrollo de máquinas que son capaces de aprender y adaptarse sin requerir instrucciones explícitas, usando algoritmos y modelos estadísticos para analizar y entregar inferencias a partir de patrones en los datos





INTELIGENCIA ARTIFICIAL

Programa que puede razonar,
actuar y adaptarse



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

DATAPALOOZA

STATS STAGE



INTELIGENCIA ARTIFICIAL

Programa que puede razonar,
actuar y adaptarse

MACHINE LEARNING

Algoritmos que van mejorando con el
paso del tiempo mientras van siendo
expuestos a más data.



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

DATAPALOOZA

STATS STAGE



INTELIGENCIA ARTIFICIAL

Programa que puede razonar, actuar y adaptarse

MACHINE LEARNING

Algoritmos que van mejorando con el paso del tiempo mientras van siendo expuestos a más data.

DEEP LEARNING

Subconjunto de machine learning que utiliza redes neuronales profundas



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



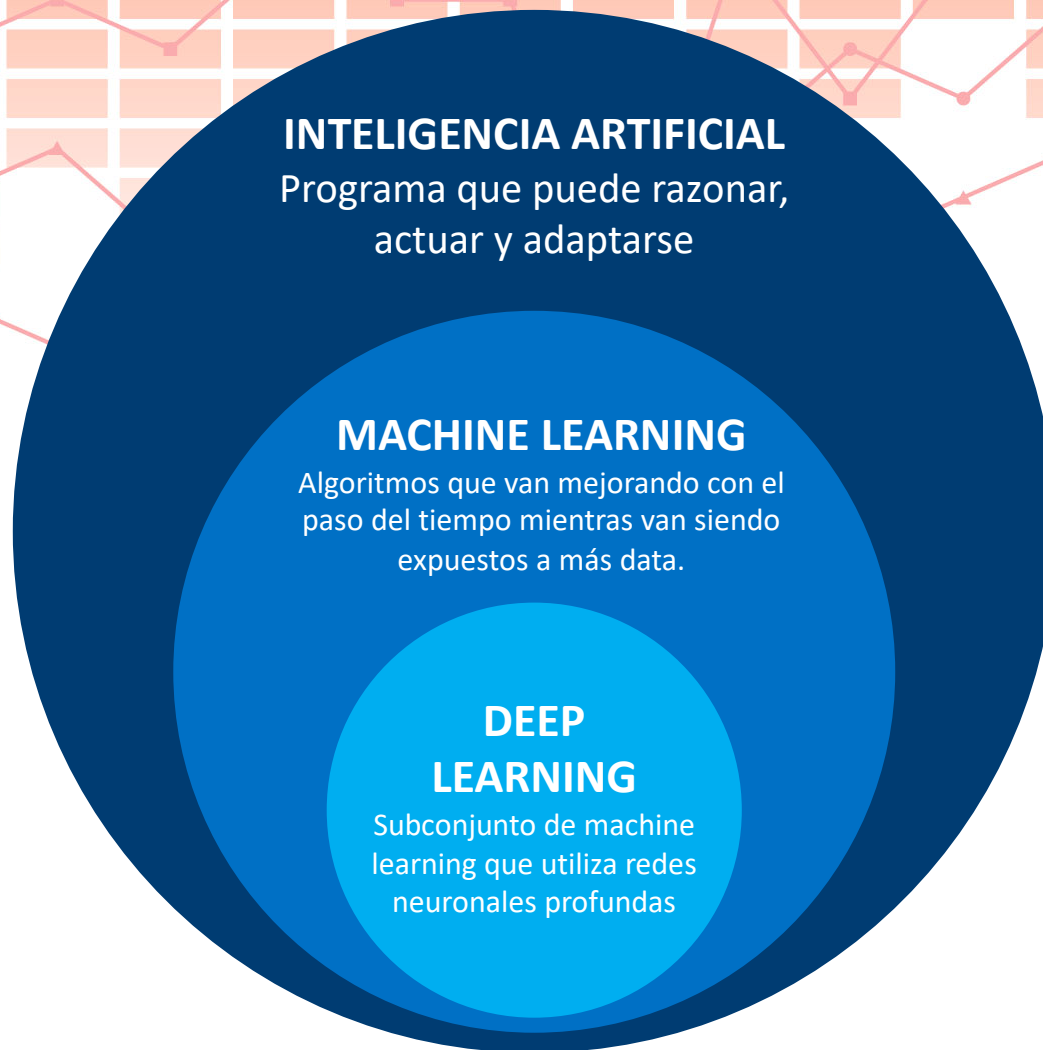
DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

DATA PALOOZA

STATS STAGE



¿Qué áreas del conocimiento se involucran?

- Ciencias de la computación
- Matemática
- Estadística



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



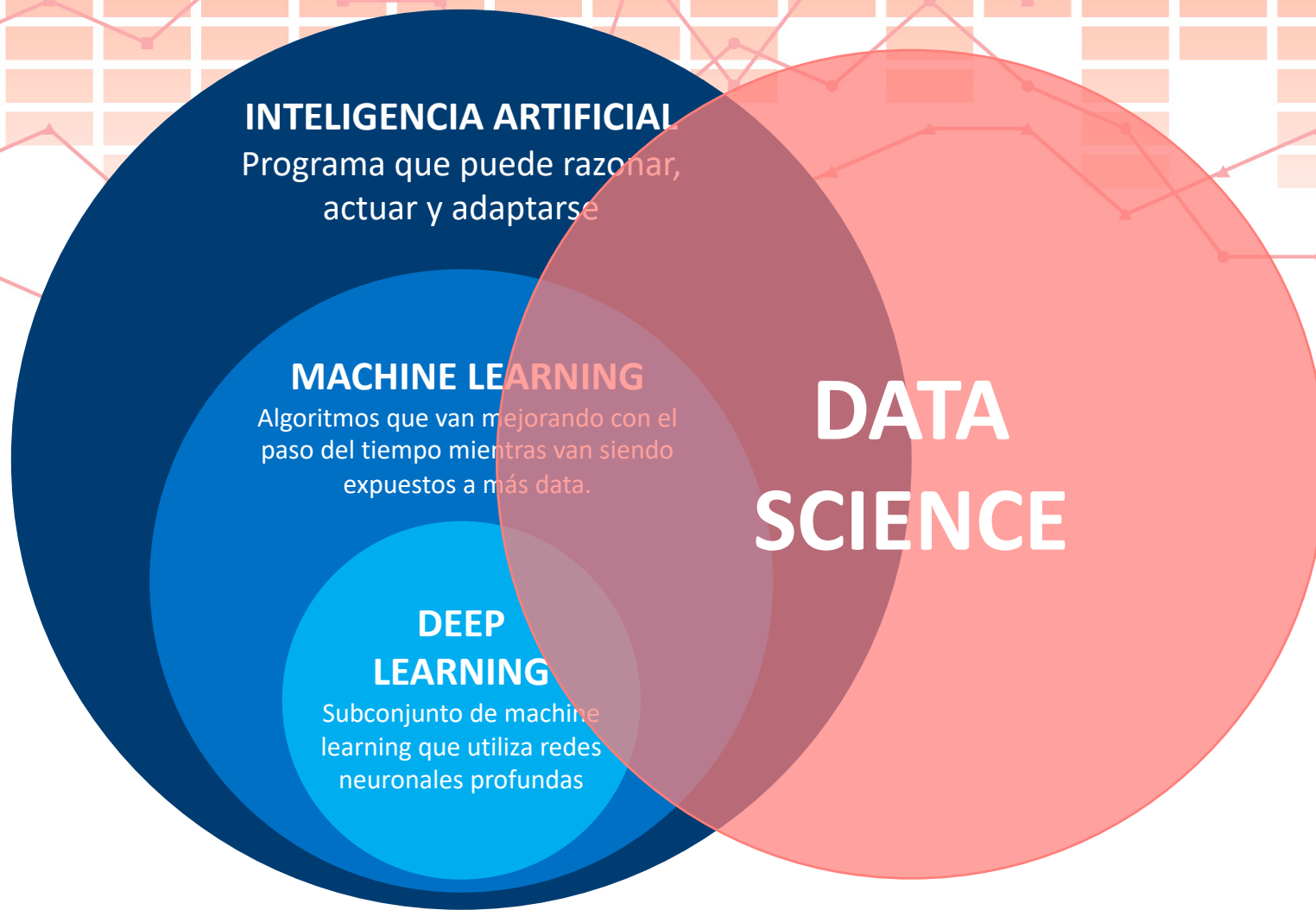
DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

DATAPALOOZA

STATS STAGE



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



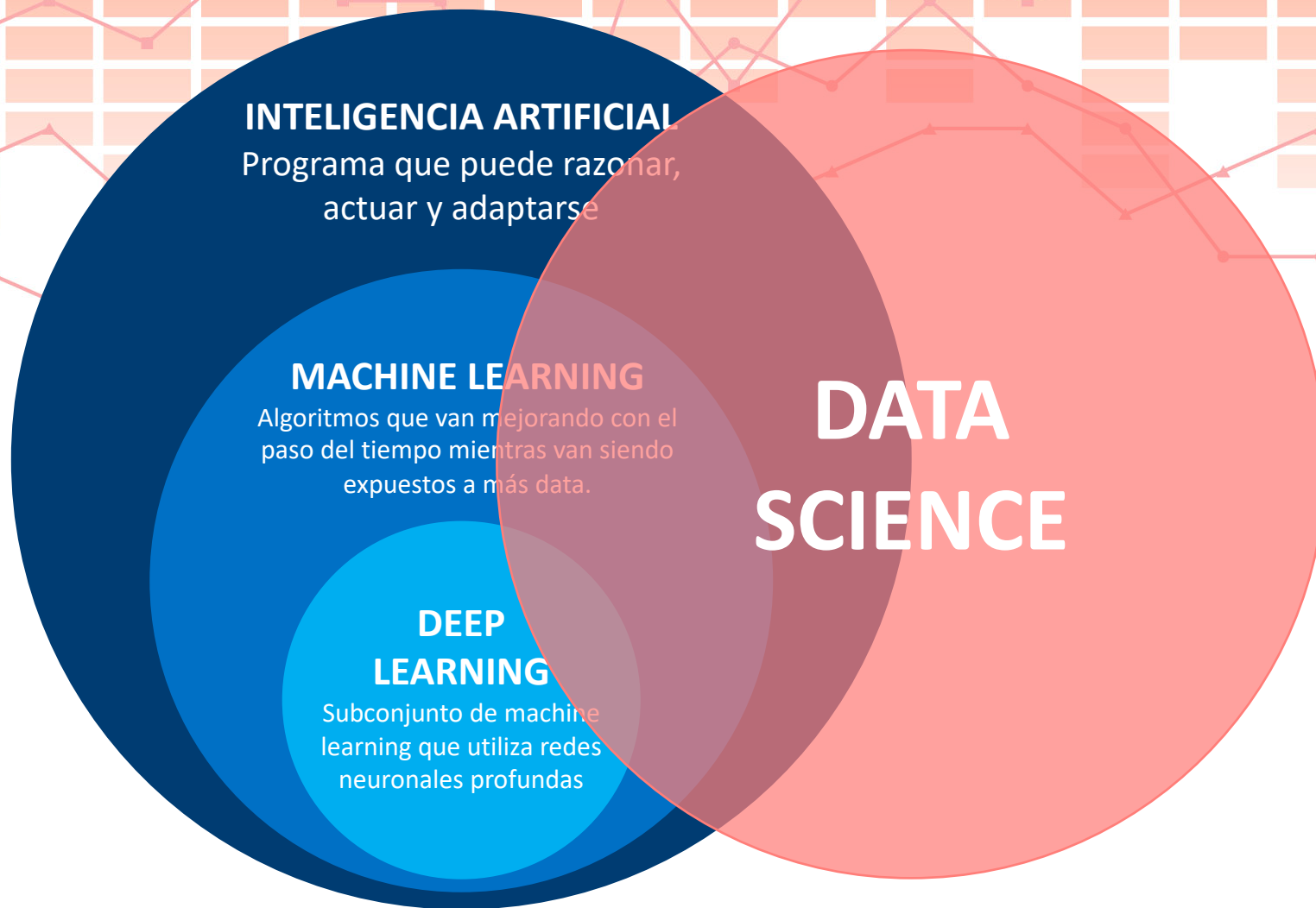
DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

DATAPALOOZA

STATS STAGE



- Conocimiento de negocio
- Solución de problemas reales
- Storytelling



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

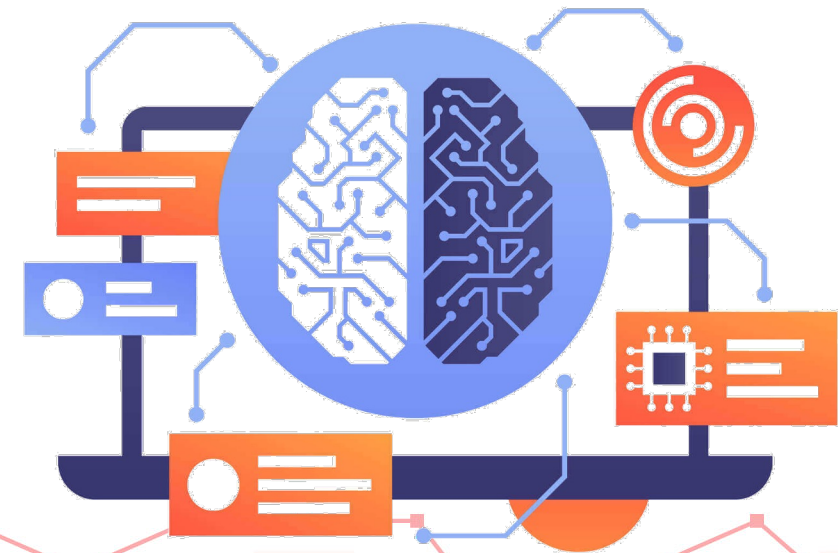
DATAPALOOZA

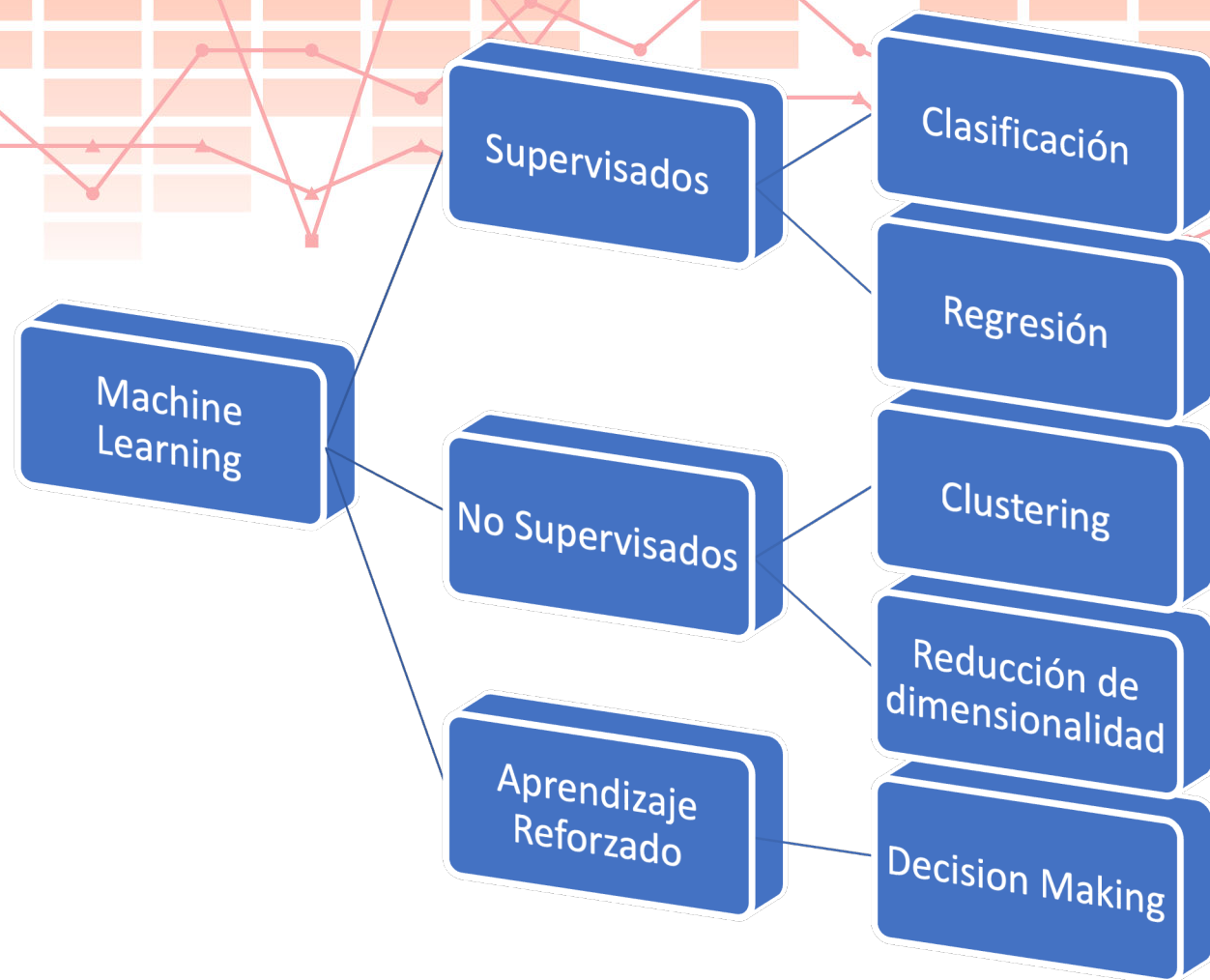
STATS STAGE

¿Qué tipo de algoritmos existen en ML?

En grueso, los algoritmos de Machine Learning se pueden clasificar en dos tipos

- **Supervisados:** Estos son modelos donde la variable que se desea explicar existe, es medible y tengo dichas mediciones en mis set de datos.
- **No Supervisados:** Son modelos donde no existe dicha variable y queremos inferir sobre ella sin conocerla.





PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



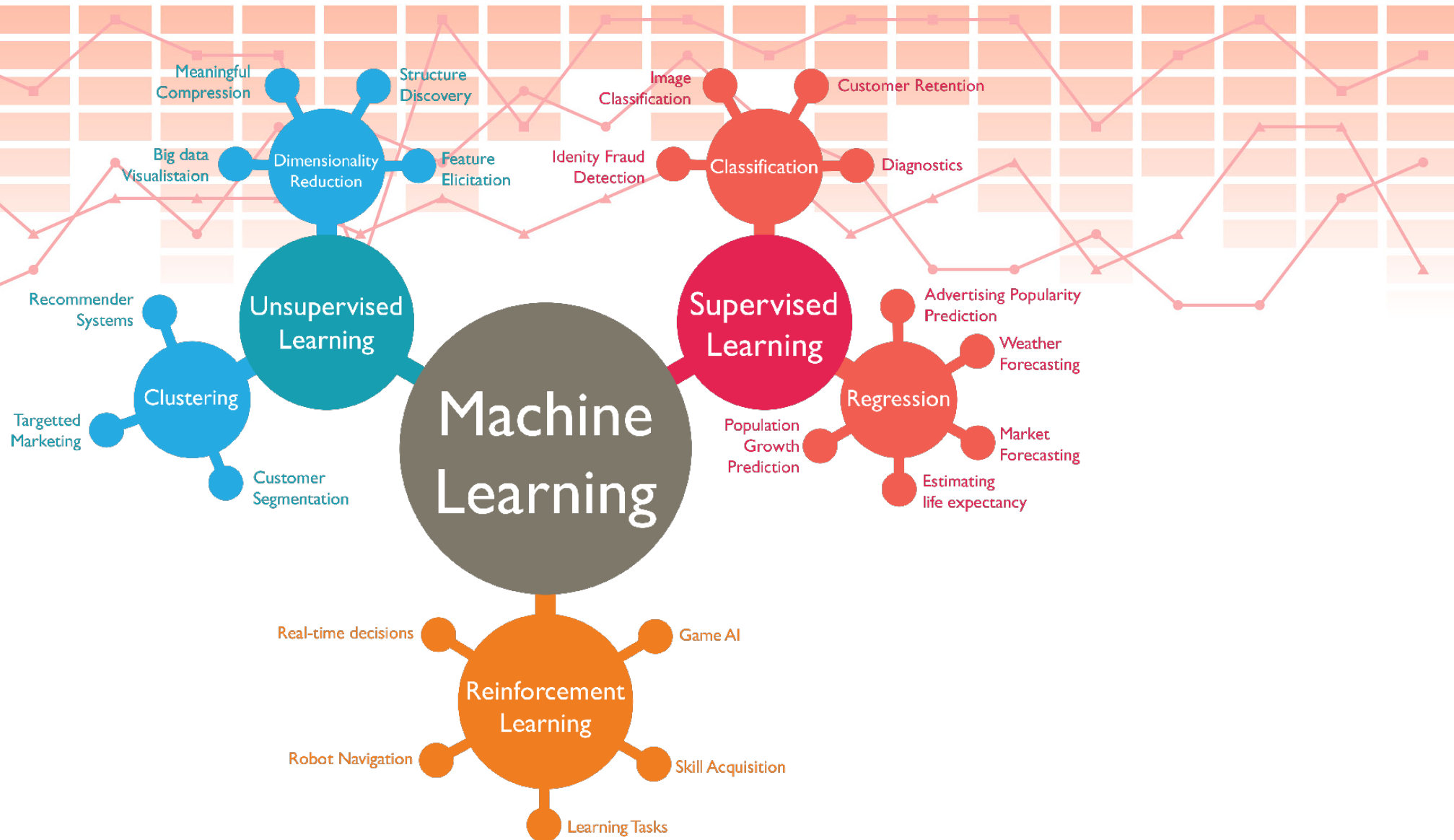
DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

DATA PALOOZA

STATS STAGE



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



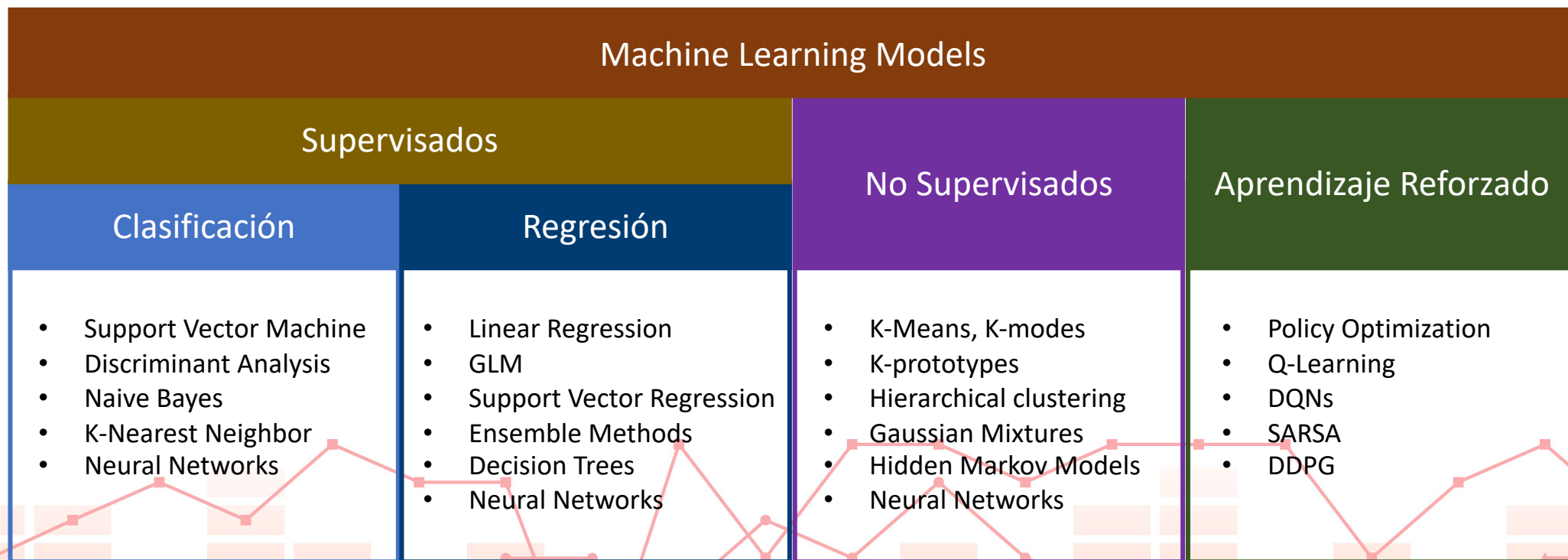
DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

DATAPALOOZA

STATS STAGE

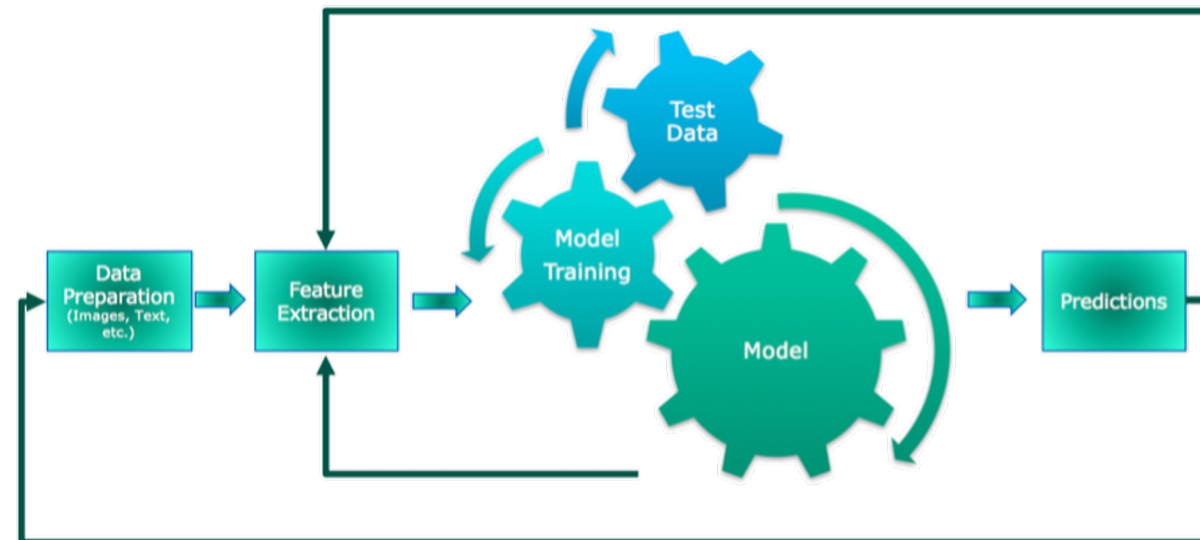


Algunos modelos



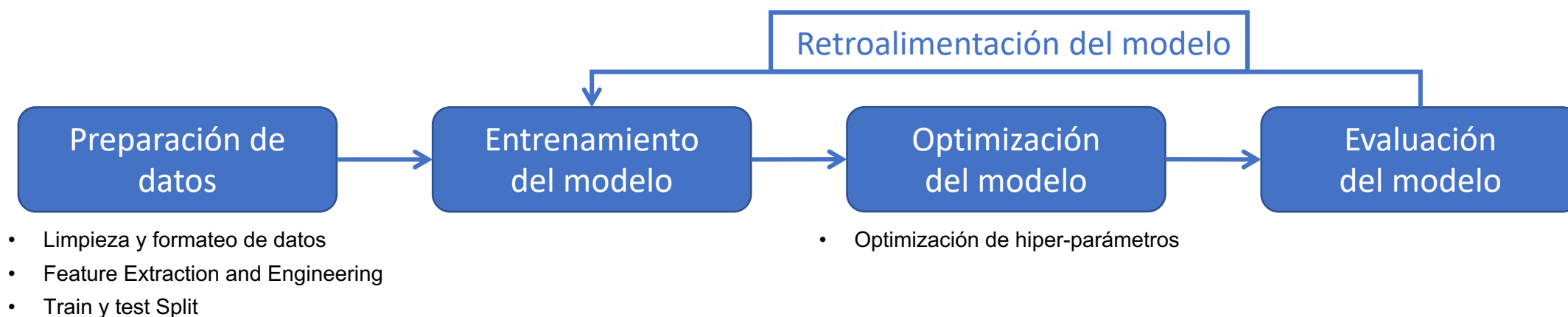
Machine Learning Pipeline

Para lograr generar un modelo de Machine Learning, es necesario conocer cada una de las etapas que considera este proceso, su importancia y también, en qué orden estas se llevan a cabo.



Machine Learning Pipeline

Estas etapas son:





Métricas de desempeño

Las métricas a utilizar dependen del tipo de modelo que estemos desarrollando

Modelos de Clasificación

- Accuracy
- Recall
- Precision
- F1 Score
- ROC - AUC

Modelos de Regresión

- MAE
- RMSE
- MSE
- R squared
- MAPE

Métricas de desempeño

Las métricas a utilizar dependen del tipo de modelo que estemos desarrollando

Modelos de Clasificación

		real	
predicción	No	No 37	Si 4
	Si	7	23

Accuracy = 0,84

Recall = 0,84

Precision = 0,77

Modelos de Regresión

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - Y_i)^2}$$



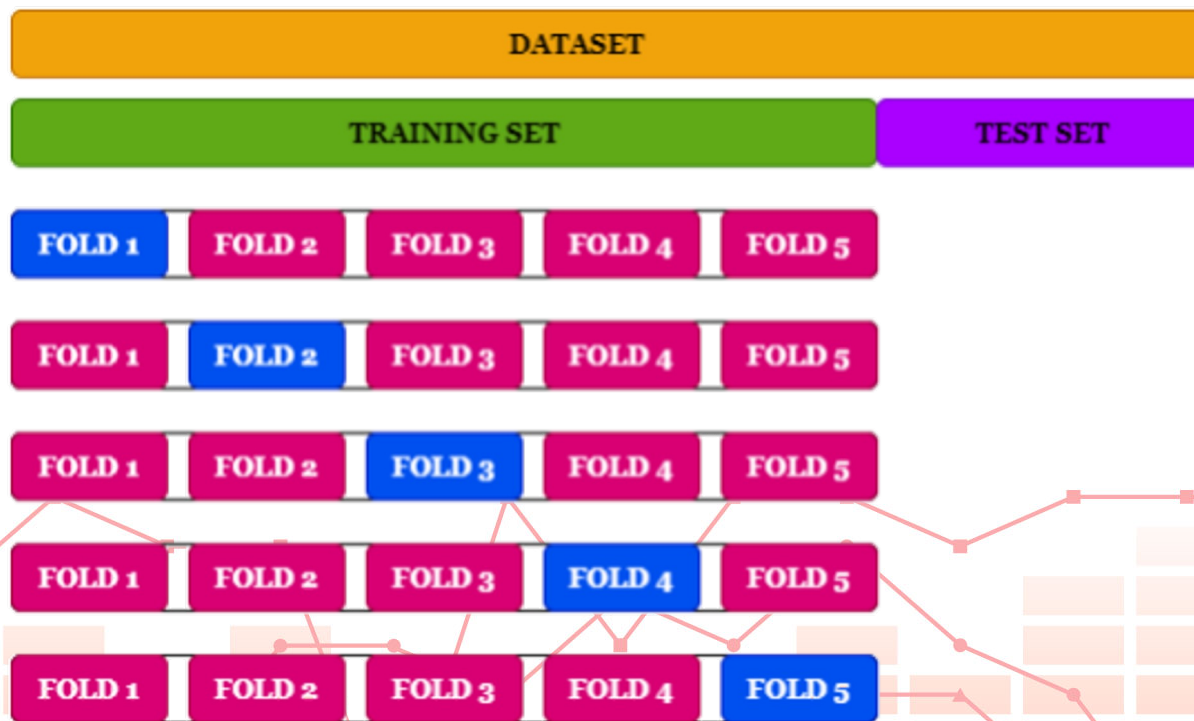
PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



STATS STAGE

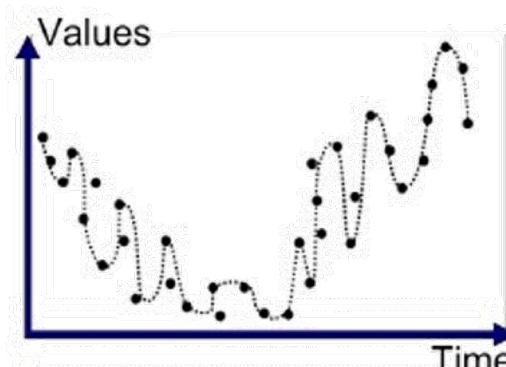
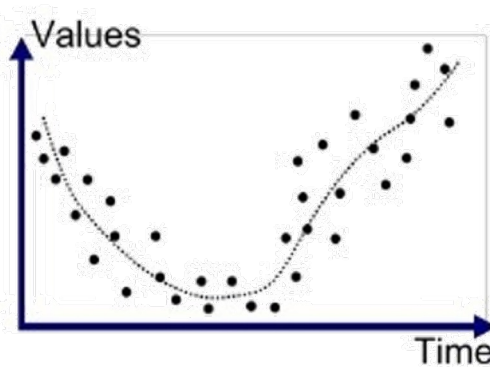
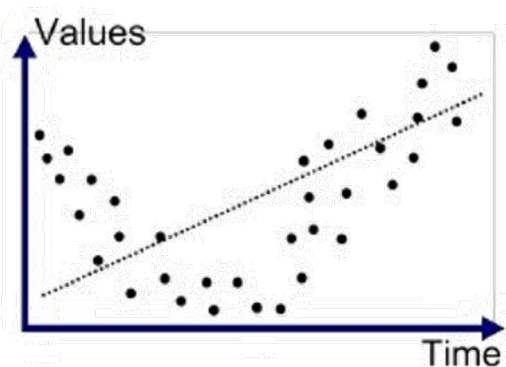
Validación cruzada

Para testear nuestro modelo o realizar optimización de hiper-parámetros, se utiliza la técnica de validación cruzada para evitar contaminación “cruzada” entre los datos de entrenamiento y testeo/optimización



Under y Over fitting

Este es uno de los problemas bien clásicos que uno puede enfrentarse al momento de entrenar modelos de machine learning



Underfitted

Good Fit/Robust

Overfitted



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



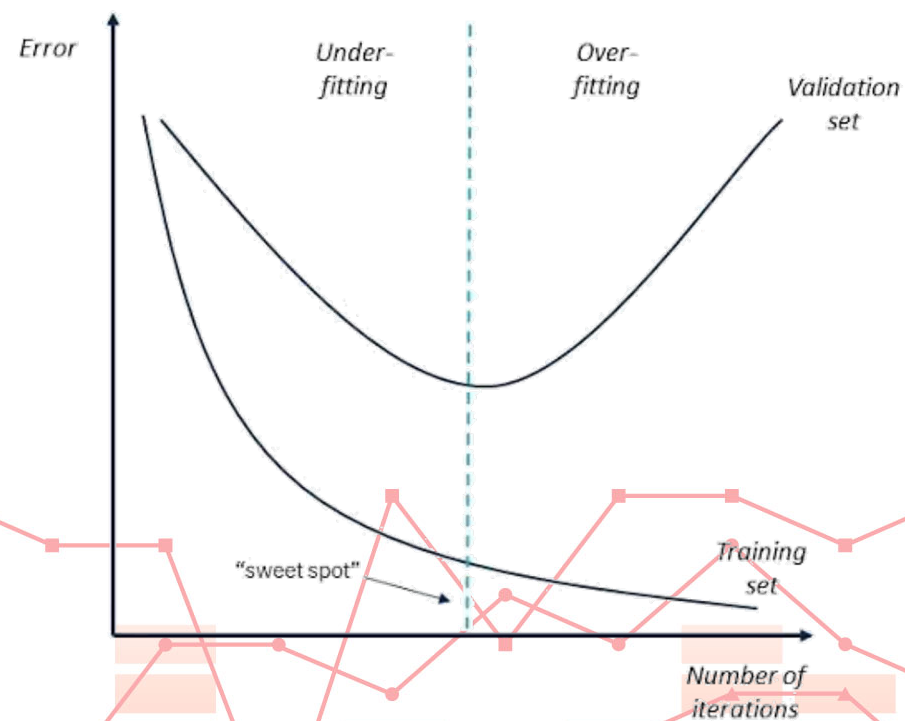
DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS



STATS STAGE

Under y Over fitting

¿Cómo lo detectamos?





PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



DIPLOMADO en
ESTADÍSTICA UC
FACULTAD DE MATEMÁTICAS



DIPLOMADO
DATA
SCIENCE UC
FACULTAD DE MATEMÁTICAS

DATAPALOOZA

STATS STAGE

¿Preguntas?



STATS STAGE

Intro al Machine Learning

Bastian Galasso-Díaz

