

## Programa del Curso

Agosto de 2021

Profesor: Adrián Soto Suárez  
adrian.soto@uai.cl  
Clases: J1 - J2

### Descripción

En la actualidad, el uso de *Machine Learning* es la base de muchas de las aplicaciones y productos tecnológicos que consumimos en el día a día. A su vez, esta herramienta se ha hecho indispensable para quienes quieren obtener valor desde grandes cantidades de datos. El propósito de este curso es introducir al alumno en las técnicas utilizadas en el área, tanto a nivel teórico como práctico.

### Objetivo General

Durante el curso el alumno aprenderá el flujo de trabajo que debe seguir un *data scientist* al involucrarse en un proyecto de *Machine Learning*. También se espera que el alumno obtenga conocimiento de los modelos y técnicas más famosas del área tanto a nivel teórico como práctico:

- Desde el punto de vista teórico, el alumno aprenderá la formalización de los modelos de aprendizaje desde un punto de vista estadístico. Además, el alumno será capaz de entender los fundamentos teóricos de las técnicas vistas en este curso.
- Desde el punto de vista práctico, el curso contempla el estudio de *frameworks* y librerías ampliamente utilizados en entornos de producción de proyectos de *Machine Learning*.

Finalmente, también se contempla estudiar técnicas para medir el desempeño de estos algoritmos y optimizar su funcionamiento. Al final del curso el alumno conocerá en qué contexto usar cada uno de los modelos, entendiendo sus ventajas y desventajas. Además podrá desarrollar algunos de estos modelos desde cero.

### Contenidos

#### Por qué *Machine Learning*

1. Motivación general.
2. Flujo de trabajo de un proyecto de *Machine Learning*.
3. Tipos de algoritmos de *Machine Learning*.

#### Conceptos previos

4. Conceptos de programación en Python.
5. Herramientas básicas: NumPy, SciPy, visualización en Python.

6. Análisis de Datos con la Librería Pandas.

#### Elementos básicos de *Machine Learning*

7. Clasificación binaria.
8. *Naive Bayes*
9. KNN.
10. Medidas de desempeño.
11. Clasificación multiclase.
12. *Trade-off* entre *bias* y varianza.

#### Regresión y entrenamiento

13. El problema de regresión.

14. Regresión lineal.

15. *Gradient Descent*.

16. Regresión polinomial.

17. Regresión logística.

## SVM

18. Clasificación lineal con SVM.

19. Clasificación no lineal con SVM.

20. Métodos de Kernel.

21. SVM y regresión.

## Árboles de decisión, *Ensemble Learning* y *Random Forests*

22. Árboles de decisión.

23. *Random forests*.

24. *Boosting*.

## Otros tópicos

25. *Feature engineering*.

26. Reducción de dimensionalidad.

27. Clustering.

28. Dimensión VC.

29. *Fairness en Machine Learning*.

## Introducción a redes neuronales

30. El perceptrón.

31. Introducción a *Deep Learning* y *backpropagation*.

32. *Redes convolucionales*

33. *Redes neuronales de grafos*

## Metodología

Cada semana el curso tendrá 2 segmentos que se detallan a continuación:

- Segmento teórico: el primer segmento enseñará los contenidos de la clase semanal, haciendo énfasis en la parte teórica de los contenidos del curso.
- Segmento práctico: el segundo segmento buscará que los alumnos desarrollen una actividad práctica guiada, que involucre programar código relacionado con el contenido de la semana.

Existe la posibilidad de no tener un segmento práctico en algunas semanas específicas.

## Evaluación

La evaluación se realizará en base a:

- Actividades semanales: el curso va a tener al menos seis actividades evaluadas, en las que se evaluará la componente práctica y la componente teórica. Si una actividad es evaluada o no se avisará con anterioridad. Además, se podrá borrar la actividad con la peor nota.
- Un proyecto final que busca poner en práctica los conocimientos adquiridos en un problema del mundo real.

El promedio de actividades será **PA** y el del proyecto será **PP**. El promedio final del curso será:

$$0,7 \cdot \mathbf{PA} + 0,3 \cdot \mathbf{PP}$$

En donde para aprobar, se debe cumplir que  $\mathbf{PA} \geq 4,0$  y  $\mathbf{PP} \geq 4,0$ . Si una de estas restricciones no se cumple, el promedio máximo va a ser 3,9.

## Bibliografía

- Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow (Aurélien Géron).
- Machine Learning (Peter Flach).
- An Introduction to Statistical Learning (James, Witten, Hastie & Tibshirani).
- Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python (Andrew Bruce, Peter C. Bruce, Peter Gedeck).
- Data Science from Scratch: First Principles with Python (Joel Grus).
- Foundations of Machine Learning (Afshin Rostamizadeh, Ameet Talwalkar & Mehryar Mohri).
- Diapositivas y lecturas de clase.
- Artículos que estarán disponibles en la página del curso.