

# Seminario de Aplicaciones Actuariales

## Seminario de Estadística I

### Aplicaciones de Ciencia de Datos con Python

El objetivo de este curso es preparar a los alumnos con las herramientas prácticas básicas para realizar un análisis de Ciencia de Datos utilizando el software Python. En particular, se utilizarán las bibliotecas Keras, Scikit-learn y TensorFlow para el desarrollo de los proyectos.

Profesor: Dr. Arrigo Coen Coria

El temario es el siguiente el cual está dividido en 5 módulos:

#### **1. Aspectos generales de Python para Ciencia de Datos**

- 1.1. Jupyter Notebook
- 1.2. Las herramientas Git y GitHub
- 1.3. Aritmética y variables
- 1.4. Condicionales y control de flujo
- 1.5. Funciones
- 1.6. El modelo CRISP-DM para realizar un proyecto

#### **2. Algoritmos de regresión**

- 2.1. Definición y conceptos generales
- 2.2. Regresión lineal simple
- 2.3. Regresión lineal múltiple
- 2.4. Modelos lineales regularizados (Ridge & LASSO Regression)
- 2.5. Regresión polinómica
- 2.6. Algoritmo de Descenso por gradiente
- 2.7. Curvas de aprendizaje

#### **3. Algoritmos de clasificación**

- 3.1. Definición y conceptos generales
- 3.2. Medidas de precisión y error
- 3.3. K-Nearest Neighbors (KNN)
- 3.4. Support Vector Machines (SVM)
- 3.5. Regresión logística
- 3.6. Curva ROC
- 3.7. Análisis del error
- 3.8. Generalizaciones de algoritmos de clasificación
- 3.9. Múltiples clasificadores

#### **4. Árboles de decisión**

- 4.1. Definición y conceptos generales
- 4.2. Generación de árboles de decisión para regresión y clasificación

- 4.3. Algoritmo CART
- 4.4. Estrategias de podado y cultivo de árboles
- 4.5. Hiperparámetros de árboles
- 4.6. Bosques aleatorios
- 4.7. Algoritmo ADA

## **5. Redes Neuronales**

- 5.1. Perceptron
- 5.2. Funciones de activación
- 5.3. Profundidad de redes y algoritmo Backpropagation
- 5.4. Entrenamiento y ajuste de redes
- 5.5. Algoritmos de aceleramiento de redes

## **Evaluación**

El curso será evaluado de la siguiente manera:

- 75% Proyectos por módulo: Para cada uno de los cinco módulos habrá una tarea correspondiente.
- 25% Proyecto final: El proyecto final es el análisis de un conjunto de datos utilizando las técnicas de Ciencia de Datos adecuadas para su análisis. Este proyecto consta de un trabajo escrito y una exposición oral del mismo.

## **Calendario**

- Módulo 1: 30 ago – 3 sept
- Módulo 2: 6 sept – 24 sept
- Módulo 3: 27 sept – 15 oct
- Módulo 4: 18 oct – 12 nov
- Módulo 5: 15 nov – 26 nov