



### Aprendizaje Estadístico I

<b>Clave</b>	: 1EST17	<b>Créditos</b>	: 3
<b>Tipo</b>	: Obligatorio	<b>Semestre</b>	: 2023-2
<b>Horario</b>	: Sábado 2 p.m. – 5 p.m.	<b>Requisitos</b>	: No tiene
<b>Profesor/a</b>	: Mg. Enver Tarazona Vargas		

#### 1. Sumilla

En este curso se presentan los conceptos fundamentales y métodos del aprendizaje estadístico, considerando tanto el aprendizaje supervisado como el no supervisado. Se considera el uso de programa estadísticos especializados en el manejo de bases de datos. Se presentan conceptos básicos para el aprendizaje estadístico, técnicas de análisis exploratorio de datos, diversas técnicas de agrupamiento y de clasificación, haciéndose énfasis en la evaluación y selección de tales métodos.

#### 2. Objetivos de aprendizaje

El desarrollo del curso permitirá a los participantes de este curso,

- Conocer de manera formal el marco teórico de los diferentes modelos del aprendizaje estadístico que le permitirán la comprensión y el análisis de las relaciones o patrones que puedan existir entre las variables que conforman diferentes bases de datos.
- El estudio específico de los modelos que corresponden al aprendizaje no supervisado y supervisado.

#### 3. Contenido

##### **Introducción**

Modelos y fundamentos del aprendizaje estadístico. Ideas generales. Modelos. Aprendizaje supervisado y no supervisado.

##### **Aprendizaje Supervisado:**

Modelos paramétricos y no paramétricos. Evaluación de modelos. Compromiso sesgo - varianza. Evaluación por remuestreo.

Modelos de clasificación y regresión: La regla ingenua. La regla de los k vecinos más cercanos. El clasificador de Bayes. Clasificación de dos poblaciones normales multivariantes. Discriminante lineal. La regresión lineal. La regresión logística. El gradiente descendente y la regla delta.

Modelos de clasificación basados en árboles. Árboles de clasificación. Particionamiento recursivo. Medidas de impureza. Evaluación de la performance. Reglas de parada y poda. Medidas para evitar el sobreajuste. Ventajas y desventajas. Árboles de regresión. Predicción. Medidas de impureza. Evaluación de la performance. Métodos agregados: Bagging, Bosque aleatorio y AdaBoost. Ventajas y desventajas.

Máquinas de vectores de soporte para la clasificación. Caso perfectamente separable y no separable, caso de margen blando. Ejemplos.

Redes Neuronales. Introducción. Conceptos y arquitectura de la red. Elección de la arquitectura de la red. Entrenamiento de la red. Retropropagación de errores. El gradiente descendente y la Regla Delta para el entrenamiento de la red. Algunos tipos de redes neuronales. Análisis de los resultados. Ventajas y desventajas.

#### **Aprendizaje no Supervisado:**

Álgebra Matricial y vectores aleatorios. Conceptos sobre matrices. Álgebra vectorial. Matrices definidas positivas. Autovalores y autovectores. Descomposición espectral. Raíz cuadrada de una matriz Teoremas de maximización. Vectores aleatorios. Matrices de varianza y covarianza.

Componentes principales. Introducción. Componentes principales poblacionales. Componentes principales para muestras. Interpretación de componentes principales. Aplicaciones. Introducción al Análisis factorial exploratorio.

Análisis de conglomerados. Introducción. Distancia y similaridad. Métodos de conglomerados jerárquicos. Métodos de K-medias. Otros métodos. Aplicaciones.

Reglas de asociación. Introducción. Descubrimiento de reglas de asociación en las bases de datos relacionadas con transacciones comerciales. El algoritmo de asociación Apriori. Medidas para la elección de reglas útiles de asociación. Confianza y soporte. Interpretación de los resultados. Significancia estadística de las reglas.

#### **4. Metodología**

Se dictarán clases teóricas y prácticas a través de la plataforma Paideia, durante las cuales se mostrarán aplicaciones de los conceptos aprendidos, usando R.

#### **5. Sistema de evaluación**

La realización de los objetivos propuestos se evaluará a través de indicadores basados en una evaluación continua, y examen parcial y final.

Fórmula de Evaluación

$$NF = 0.4EC + 0.30EP + 0.30 EF$$

donde: NF = Promedio Final del Curso

EC = Promedio de evaluación continua (cuestionarios o lista de ejercicios)

EP = Nota en el Examen Parcial

EF = Nota en el Examen Final

## 6. Bibliografía

Véliz, C, (2020) Aprendizaje automático. Fondo editorial de la Pontificia Universidad Católica del Perú. Lima Perú.

Gareth G., Witten D., Hastie T., Tibshirani R. (2021) An Introduction to Statistical Learning. Springer. USA

Géron A. Hands On Machine Learning with Scikit-Learn – Tensor Flow. (2017). O'Reilly. USA.

Herramientas Matemáticas para el Aprendizaje Automático. Notas preparadas por el profesor C. Véliz. 2021

## 7. Cronograma

Semana	Unidad, tema o capítulo
1	Introducción. Modelos y fundamentos del aprendizaje estadístico
2	Modelos y fundamentos del aprendizaje estadístico
3	Modelos de clasificación y regresión
4	Modelos de clasificación y regresión
5	Medidas de evaluación. Aplicaciones en R.
6	Remuestreo
7	Modelos de clasificación basados en árboles
8	Modelos ensamblados
9	Examen parcial
10	Otros modelos de clasificación y regresión
11	Aprendizaje no supervisado (I): Componentes principales
12	Aprendizaje no supervisado (II): Conglomerados
13	Aprendizaje no supervisado (I): Conglomerados
14	Aprendizaje no supervisado (I): Reglas de asociación
15	Redes neuronales
16	Examen final
17	Libre