



DIPLOMADO DATA SCIENCE

26

VERSIÓN

MACHINE LEARNING E INTELIGENCIA ARTIFICIAL, DEEP LEARNING



MODALIDAD ON-LINE

SOFTWARE: PYTHON, R, SPARK, SQL*

Nuevos contenidos sobre la Inteligencia Artificial para lenguaje natural de tipo ChatGPT

*Se incluye acceso a la plataforma cloud de Amazon Web Services (AWS).

PRESENTACIÓN DEL DIPLOMADO

Miércoles 5 de Noviembre 2025
 19:00 hrs.

INICIO DE CLASES

Lunes 10 de Noviembre 2025
 19:00 hrs.

FECHA LÍMITE PARA MATRICULARSE

4 de Noviembre de 2025 o hasta completar cupo máximo.

PROFESORES

HAMDI RAISSI

PhD Universidad de Lille, Francia, Profesor Adjunto PUCV.

MARIO GUZMÁN

Magister en Estadística PUCV, Data Scientist.

PATRICIO VIDELA

Jefe de Docencia del Instituto de Estadística, Profesor Auxiliar PUCV.

HORARIO

Lunes y Miércoles
 de 19:00 a 22:00 hrs.

CONTACTO

diplomado.estadistica@pucv.cl

CLASES

Noviembre 2025	10	12	17	19	24	26			
Diciembre 2025	1	3	10	15	17	22	29		
Enero 2026	5	7	12	14	19	21	26	28	
Marzo 2026	2	4	9	11	16	18	23	25	30
Abril 2026	1	6							

El programa considera 96 hrs. Cronológicas.

Todas las clases son de 3 horas y empiezan a las 19 hrs. en modalidad "online"***

TEMAS BÁSICOS

1. ESTADÍSTICA DESCRIPTIVA

- Estadísticas descriptivas y su visualización.
- Tipos de variables en los datos.

2. TOMA DE DECISIÓN EN UN ENTORNO ALEATORIO

- Test estadístico.
- Intervalos de confianza para pronósticos.

3. ANÁLISIS DE ASOCIACIÓN DE VARIABLES

- Estrategias para medir la correlación entre variables: Pearson, Spearman o Kendall?
- Modelos lineales simples: Estimación MCO, Diagnóstico de bondad, Test de normalidad.
- One way ANOVA y two way ANOVA, razón de correlación.

4. MÉTODOS MULTIVARIADOS EN ESTADÍSTICA

Análisis por componentes principales (ACP).

TEMAS AVANZADOS

1. MODELOS LINEALES MÚLTIPLES

- Estimación MCO, diagnóstico de bondad (t-test, test de Fisher) y tipos de predicción (individual y del fenómeno estudiado).
- Test de homogeneidad poblacional de Chow.
- Identificación de las variables pertinentes (Cp de Mallows, Criterios de selección forward, stepwise y backward). Como introducir las variables categóricas en un modelo lineal.
- Problema de colinealidad y soluciones (regresión PCR, PLS, regresión Ridge, LASSO, LASSO por grupo y elastic net).
- Datos outliers (atípicos): detección y diagnóstico (leverages, residuos studentizados, distancia de Cook, DFBETAS). Solución con la estimación robusta de Theil-Sen y Siegel, estimación M.
- Heteroscedasticidad y autocorrelación: diagnóstico (test de Durbin Watson, tests de Breusch-Pagan) y estimación MCG. Modelos AR, MA y ARMA ajustados a residuos.

2. MÉTODOS NUMÉRICOS DE ALTO NIVEL COMPUTACIONAL

- Introducción a EC2 de AWS.
- Métodos bootstrap: paramétrico, no paramétrico y aplicados a modelos lineales.

3. MODELOS PARA DATOS TEMPORALES

- Modelamiento univariado de datos temporales con modelos SARIMA.
- Identificación: Autocorrelaciones (ACF), Autocorrelaciones parciales (PACF), Criterios de información.
- Diagnóstico y predicción.
- Regresión falsa y test de raíz unitaria.

4. INTRODUCCIÓN A SQL

- Comandos SQL y tipos de datos.
- Modelos relacionales.
- Rutinas de comandos en SQL Server.
- Depuración de datos para resolución de problemas.
- Conexión a SQL Server.

5. INTRODUCCIÓN A SPARK

- Tratamiento de data frame.
- Análisis descriptivo.
- Categorización de bases.
- Rutinas de Pyspark.

6. ALGORITMO DE K-MEDIAS

- Medidas de similitudes.
- Algoritmo K-medias.
- Clustering Jerárquico.
- Métricas de validación.
- Aplicaciones.

7. ÁRBOLES DE DECISIÓN

- Clasificación del árbol.
- Requisitos y supuestos de los datos.
- Interpretación de los resultados.
- Predicción y Evaluación.
- Aplicaciones.

8. RANDOM FOREST

- Introducción al Random Forest.

- Entrenamiento de un modelo Random Forest.
- Evaluación de out-of-bag error.
- Evaluación del rendimiento del modelo Random Forest.
- Aplicaciones.

9. MODELO DE REGRESIÓN LOGÍSTICA

- Presentación del modelo e interpretación.
- Ajuste del Modelo e interpretación de resultados.
- Estudio de caso real en Python: evaluación y construcción.

10. MÁQUINAS DE VECTORES DE SOPORTE

- Definición de hiperplano de separación.
- Clasificador de margen máximo.
- SVM para clasificador linealmente separable.
- SVM para clasificador linealmente no separable.
- Extensión de las máquinas de vectores de soporte.
- Métricas de validación.
- Aplicaciones.

11. REDES NEURONALES

- Arquitectura de una red.
- Perceptrón.
- Función de activación.
- Back-propagation.
- Métricas de validación.
- Aplicaciones.

12. TEXT MINING

- Homologación de textos en base a cercanía de textos.
- Arquitectura del web scraping.

- Aplicaciones de web scraping y cercanía de textos en Python.

13. MANEJO DE HERRAMIENTAS DE AWS

- Introducción a S3.
- Gestión de permisos con IAM.
- Redes virtuales en la nube VPC.
- Introducción a SageMaker.
- Rutinas de modelos de ML en SageMaker con Python.

14. SISTEMAS DE RECOMENDACIÓN

- Filtros colaborativos.
- Sistema basado en usuarios e items.
- Aplicaciones de sistemas de recomendación.

15. DEEP LEARNING

- Introducción al Deep Learning.
- Redes convolucionales (CNN).
- Arquitectura Alexnet.
- Aplicaciones de CNN con framework torch en Python.

16. INTELIGENCIA ARTIFICIAL PARA LENGUAJE NATURAL

- IA como modelos generativos.
- LLM (Large Language Models) desde una perspectiva Estadística.
- Características de un LLM: tamaño muestral, ventana de contexto, ingeniería de prompts.
- Oportunidades, limitaciones y riesgos en el uso de LLM.
- Caso de uso: usando un modelo de tipo ChatGPT.

TEMARIO

* No se necesita conocimientos previos de los software dado que una introducción será hecha para cada software ocupado. Los códigos listos para el uso y comentados en la clase.

** Datos reales o simulados.