

**APRENDIZAJE DE
MÁQUINA
2017-2**

I. INFORMACIÓN GENERAL

CURSO	APRENDIZAJE DE MÁQUINA
CLAVE	1INF02
CRÉDITOS	3.5
HORAS DE DICTADO	CLASE: 3 Semanal LABORATORIO: EXAMEN:
HORARIO	TODOS
PROFESORES	FELIX ARTURO ONCEVAY MARCOS CÉSAR ARMANDO BELTRÁN CASTAÑÓN

II. PLANES CURRICULARES DONDE SE DICTA EL CURSO

ESPECIALIDAD	ETAPA	NIVEL	CARÁCTER	REQUISITOS
INGENIERÍA INFORMÁTICA	PREGRADO EN FACULTAD	0	ELECTIVO	INF265 APLICACIONES DE CIENCIAS DE LA COMPUTACIÓN [07]

Tipos de requisito

- 04 = Haber cursado o cursar simultáneamente
- 05 = Haber aprobado o cursar simultáneamente
- 06 = Promedio de notas no menor de 08
- 07 = Haber aprobado el curso

III. DESCRIPCIÓN DEL CURSO

En ciencia y análisis de datos, una de las tareas más comunes que se desarrollan son la predicción y el aprendizaje de máquina. El aprendizaje de máquina es la ciencia enfocada en conseguir que las computadoras actúen sin ser explícitamente programadas. En el presente curso, los alumnos comprenderán los conceptos básicos y técnicas más efectivas de aprendizaje de máquina, y sus respectivas aplicaciones, para el descubrimiento de nuevo conocimiento en los datos.

El curso otorga una introducción al área de aprendizaje de máquina y reconocimiento de patrones estadístico. Entre los temas se encuentran: aprendizaje supervisado (regresión, clasificación), aprendizaje no supervisado, aplicaciones (reducción de dimensionalidad) y tópicos complementarios (aprendizaje profundo, aprendizaje semi-supervisado).

IV. SUMILLA

En Ciencia y Analítica de Datos una de las tareas más comunes que se desarrollan son la predicción y el aprendizaje de máquina. El aprendizaje de máquina es la ciencia de conseguir que las computadoras actúen sin ser explícitamente programadas. En el presente curso aprenderá las técnicas más efectivas de aprendizaje de máquina y sus respectivas aplicaciones.

El curso otorga una introducción al aprendizaje de máquina, minería de datos y reconocimiento de patrones estadístico. Comprende temas como: aprendizaje supervisado, aprendizaje no supervisado y, aplicaciones y tópicos avanzados.

V. OBJETIVOS

El objetivo principal del curso es estudiar distintas técnicas de aprendizaje de máquina que permitan la búsqueda de patrones y nuevo conocimiento dentro de la información para un mejor uso de la misma. Asimismo, este curso permite:

- Conocer los principales conceptos de técnicas algorítmicas de aprendizaje de máquina y sus respectivas aplicaciones en problemas de ciencia y analítica de datos.
- Aplicar técnicas algorítmicas de aprendizaje de máquina a problemas reales.
- Analizar el estado del arte con respecto a las recientes aplicaciones de aprendizaje de máquina usando grandes volúmenes de información.
- Resolver un problema práctico mediante la aplicación de las técnicas vistas en el curso, orientada al uso de los algoritmos de aprendizaje de máquina.

Estos objetivos contribuyen al logro de los siguientes seis Resultados del Programa:

- Aplica los conocimientos relacionados a las matemáticas, ciencias e ingeniería.
- Diseña sistemas, componentes o procesos que satisfagan las necesidades presentadas.
- Identifica, formula y resuelve problemas de ingeniería.
- Reconoce la necesidad y se compromete con el aprendizaje a lo largo de toda la vida.
- Utiliza las técnicas, estrategias y herramientas de la ingeniería moderna necesarias para la práctica de la misma.
- Aplica los conocimientos relacionados a los lenguajes de programación, modelado de sistemas de información, construcción de software de calidad y administración de recursos tecnológicos.

VI. PROGRAMA ANALÍTICO

UNIDAD 1 INTRODUCCIÓN (2 horas)

Objetivo: Conocer los objetivos y metodología del curso. Entender que el aprendizaje de máquina se ha constituido en una de las tareas más comunes que realiza un científico de datos y/o analista de datos.

Contenido: Conceptos básicos. Aplicaciones más recientes en aprendizaje de máquina.

PRIMERA PARTE: APRENDIZAJE SUPERVISADO

Objetivo: Estudiar el concepto de aprendizaje supervisado y sus principales técnicas.

UNIDAD 2 REGRESIÓN (3 horas)

Contenido: Regresión lineal para la predicción de un valor real basado en determinados valores de entrada, función modelo y costo, método de gradiente descendiente para aprendizaje, aplicaciones de regresión lineal.

UNIDAD 3 CLASIFICACIÓN (7 horas)

Contenido: Conceptos básicos. Algoritmos de clasificación: KNN, Árboles de decisión, Regresión lineal logística, Regularización, representación y aprendizaje de Redes Neuronales.

UNIDAD 4 KERNELS Y SVM (3 horas)

Contenido: Clasificación por máquinas de vectores de soporte, uso de kernels, aplicaciones de SVM.

UNIDAD 5 CLASIFICACIÓN POR ENSAMBLE (6 horas)

Contenido: Clasificación por esquemas de ensamble. Adaboost. Random Forest. Esquemas de boosting, bagging y votación.

SEGUNDA PARTE: APRENDIZAJE NO SUPERVISADO

Objetivo: Estudiar el concepto de aprendizaje no supervisado y sus principales técnicas

UNIDAD 6 CLUSTERING Y MODELOS BASADOS EN CENTROIDES (2 horas)

Contenido: Introducción al análisis de clustering, algoritmo de K-medias, K-medianas, Ventajas y desventajas, optimización de k-medias.

UNIDAD 7 CLUSTERING JERÁRQUICO O BASADO EN CONECTIVIDAD (2 horas)

Contenido: Algoritmo de clustering jerárquico, corte de árbol, clustering máximo, medio y promedio, validación de clusters, correlación de clusters.

UNIDAD 8 OTROS MODELOS DE CLUSTERING (2 horas)

Contenido: Algoritmos de clustering basados en distribución (Expectation Maximization), y densidad (DBSCAN, OPTICS)

TERCERA PARTE: TÓPICOS COMPLEMENTARIOS

UNIDAD 9 TRANSFORMACIÓN Y SELECCIÓN DE DATOS (6 horas)

Contenido: Detección de anomalías. Reducción de dimensionalidad: Principal Component Analysis (PCA).

UNIDAD 10 . OTROS ESQUEMAS DE APRENDIZAJE (6 horas)

Contenido: Aprendizaje semi-supervisado y ¿Active learning¿. Introducción al aprendizaje profundo (¿Deep learning¿). Introducción al aprendizaje por refuerzo (¿Reinforcement learning¿).

VII. METODOLOGÍA

El curso se basa en clases expositivas en donde se darán los fundamentos teóricos y se desarrollarán aplicaciones y ejercicios en clases. Las aplicaciones prácticas y evaluaciones se realizarán en laboratorio, empleando principalmente el lenguaje de programación Python.

Dentro de la forma de evaluación se considerará el desarrollo de trabajos grupales.

VIII. EVALUACIÓN

Sistema de evaluación

Nº	Codigo	Tipo de Evaluación	Cant. Eval.	Forma de aplicar los pesos	Pesos	Cant. Eval. Eliminables	Consideraciones adicionales	Observaciones
1	Pb	Práctica tipo B	5	Por Promedio	Pb=2	0		
2	Ta	Tarea académica	1	Por Evaluación	Ta1=3			
3	Ex	Examen	2	Por Evaluación	Ex1=2 Ex2=2			

Fórmula para el cálculo de la nota final

$$(2Pb + 3Ta1 + 2Ex1 + 2Ex2) / 9$$

Aproximación de los promedios parciales Redondeado a 1 decimal

Aproximación de la nota final Redondeado a 0 decimales

Consideraciones adicionales

Se aplica la Modalidad de Evaluación 2.

TA = actividades académicas designadas por el profesor. Estas pueden ser: monografías, exposiciones en clase, trabajos especiales, informes, controles de lectura, seminarios, participación en clase, etc. El puntaje obtenido por el alumno por este concepto será consignado como nota TA.

IX. BIBLIOGRAFÍA

Referencia obligatoria

Referencia complementaria

- Libro
C. Bishop
2006
Pattern Recognition and Machine Learning
Springer-Verlag, New York
- Libro
Duda, Richard O.
2001
Pattern classification
New York : Wiley, 2001.

[https://pucp.ent.sirsi.net/client/es_ES/campus/search/detailnonmodal/ent:\\$002f\\$002fSD_ILS\\$002f0\\$002fSD_ILS:282135/one](https://pucp.ent.sirsi.net/client/es_ES/campus/search/detailnonmodal/ent:$002f$002fSD_ILS$002f0$002fSD_ILS:282135/one)

- Libro
Hastie, Trevor.
2001
The elements of statistical learning : data mining, inference, and prediction
New York : Springer, 2001
[https://pucp.ent.sirsi.net/client/es_ES/campus/search/detailnonmodal/ent:\\$002f\\$002fSD_ILS\\$002f0\\$002fSD_ILS:282278/one](https://pucp.ent.sirsi.net/client/es_ES/campus/search/detailnonmodal/ent:$002f$002fSD_ILS$002f0$002fSD_ILS:282278/one)
- Libro
I.H. Witten and E. Frank
2005
Data Mining: practical Machine learning Tools and Techniques, (Second Edition)
Morgan Kaufmann, 2005
- Libro
L. Costa and R. Cesar Jr.
2000
Shape Analysis and Classification: Theory and Practice
CRC Press, Boca Raton
- Libro
M.A. Nielsen
2015
Neural networks and deep learning
Determination Press
- Libro
Mitchell, Tom M.
1997
Machine learning
New Delhi : McGraw Hill Education, 1997.
[https://pucp.ent.sirsi.net/client/es_ES/campus/search/detailnonmodal/ent:\\$002f\\$002fSD_ILS\\$002f0\\$002fSD_ILS:592828/one](https://pucp.ent.sirsi.net/client/es_ES/campus/search/detailnonmodal/ent:$002f$002fSD_ILS$002f0$002fSD_ILS:592828/one)
- Libro
R. Sutton and A. Barto
1998
Reinforcement Learning: An Introduction
MIT Press
- Libro
S. Theodoridis, K Koutroubas
1998
Pattern recognition
Academics press, San Diego, 1998

X. POLÍTICA CONTRA EL PLAGIO

Para la corrección y evaluación de todos los trabajos del curso se va a tomar en cuenta el debido respeto a los derechos de autor, castigando severamente cualquier indicio de plagio con la nota CERO (00). Estas medidas serán independientes del proceso administrativo de sanción que la facultad estime conveniente de acuerdo a cada caso en particular. Para obtener más información, referirse a los siguientes sitios en internet

www.pucp.edu.pe/documento/pucp/plagio.pdf