

PROGRAMA DE CURSO

Código	Nombre			
MA5203	Aprendizaje de Máquinas Probabilístico			
Nombre en Inglés				
Probabilistic Machine Learning				
SCT	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
	10	3	2	5
Requisitos			Carácter del Curso	
MA3401 o MA3403				
Resultados de Aprendizaje				
Este curso enseña los fundamentos del enfoque probabilístico para aprendizaje de máquinas, su importancia en relación a otras ramas que requieren análisis de datos y su relación con métodos determinísticos. Al final del curso, el alumno deberá conocer las propiedades de distintos métodos de aprendizaje probabilístico, cómo máxima verosimilitud, máximo a posteriori, y estimación Bayesiana, tanto para modelos paramétricos como no-paramétricos. El alumno también aprenderá la diferencia entre inferencia exacta y aproximada desde puntos de vista teóricos y aplicados. Finalmente, el alumno deberá implementar las herramientas aprendidas a datos reales en problemas de regresión, predicción, clasificación, selección de modelos y reducción de dimensionalidad.				

Metodología Docente	Evaluación General
La metodología del curso comprenderá los siguiente ítems: • Clases expositiva de 90 minutos • Demostraciones de las métodos aprendidos • Clases auxiliares donde se guiará a los alumnos a implementar las técnicas vistas en cátedra • Realización de un proyecto individual	La evaluación será en aspectos tanto teóricos como prácticos durante el desarrollo del curso e incluirá: • Participación en clase • Presentaciones y discusiones grupales en base a material de grupo de lectura • Tareas • Proyecto

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
1	Contexto	1.5
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
a) Introducción - Inteligencia Artificial, fundamentos filosóficos, agentes [IAMA, MA] - Sistemas experto versus automatización [IAMA] - Definición de aprendizaje [LFD, ISL] - Ejemplo: Ajuste de curvas polinomiales - Aplicaciones reales y motivación b) Definición de Aprendizaje de Máquinas [IAMA cap18] - Taxonomía: Aprendizaje supervisado, no-supervisado, semi-supervisado, reforzado [LFD] - Sobreajuste / subajuste [LFD, ISL] - Entrenamiento, validación y test [LFD, ISL] - Navaja de Occam y sesgo de muestreo [LFD] c) Aprendizaje mediante inferencia probabilística [BDA] - Tres pasos: modelo, distribución posterior, evaluación. - Inferencia bayesiana - Probabilidad como medida de incertidumbre - Modelos generativos - Ejemplo: un modelo de un parámetro - Nota histórica: Bayes y Laplace	- Conocer los fundamentos del aprendizaje de máquinas en relación a la computación clásica y a la inteligencia artificial - Entender la relación entre aprendizaje de máquinas y razonamiento probabilístico - Comprender cómo extraer información y manejar incertidumbre de datos con un enfoque Bayesiano	Ver notas de referencia en cada contenido

Número	Nombre de la Unidad	Duración en Semanas
2	Regresión: inferencia exacta y aproximada	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
- Modelo lineal [ISL] - Regularización mediante elección de distribución <i>a priori</i> - Regresión no lineal: polinomial, escalones, bases, <i>splines</i> [ISL] - Modelos lineales generalizados	- Habilidad para formular un problema de regresión y lo resolverlo de forma determinística o probabilística	Ver notas de referencia en cada contenido

[MLPP, ISL] - Modelos jerárquicos [BDA] - Altas dimensiones [ISL] y <i>curse of dimensionality</i> [MLPR] - Regresión logística [MLPP] - Aproximación de Laplace - Nota sobre métodos de Monte Carlo [MLPP]	- Resolver modelos intratables mediante métodos de Monte Carlo - Realizar predicción en base a datos observados tomando en cuenta conceptos de incertidumbre y sobreajuste	
--	---	--

Número	Nombre de la Unidad	Duración en Semanas
3	Selección y promedio de modelos [MSMA]	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
- Selección: Akaike IC, Bayesian IC, Negative log-predictive distribution [BDA, ESL (ch7)] - Promedio: Maximizar versus integrar: enfoque bayesiano versus determinista [ITILA]	- Evaluar, seleccionar y combinar distintas estructuras de modelos en base a índices de desempeño, incluso cuando la cantidad de estructuras es infinita	Ver notas de referencia en cada contenido

Número	Nombre de la Unidad	Duración en Semanas
4	Clasificación	2.5
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
- K-vecinos más cercanos [ISL] <i>Naïve Bayes</i> [ISL] - Teoría de Aprendizaje estadístico: La dimensión de Vapnik-Chervonenkis [ISL, LFD, ESL] - Máquinas de soporte vectorial [LWK]	-Representar probabilísticamente el problema de clasificación y relacionarlo con enfoques deterministas. -Conocer la teoría de aprendizaje estadístico y su relación con las máquinas de soporte vectorial -Implementar máquinas de soporte vectorial	Ver notas de referencia en cada contenido

Número	Nombre de la Unidad	Duración en Semanas
5	Redes neuronales [ISL, MLPR, DL]	2
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía

- Definición, perceptrón, adaline - Backpropagation, tipos de redes (CNN, RNN, FFNN). - Heurísticas - Autoencoders y redes generativas adversariales <i>Deep Learning</i> <i>Bayesian Deep Learning</i> [UDL]	-Conocer las estructuras estándar de redes neuronales y técnicas de entrenamiento -Implementar estructuras actuales de redes neuronales usando toolboxes disponibles -Nociones sobre el estado del arte en redes bayesianas	Ver notas de referencia en cada contenido
--	---	---

Número	Nombre de la Unidad	Duración en Semanas
6	Procesos gaussianos [GPML]	2
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
- Construcciones: suma infinita de funciones bases, perceptrón de una capa infinitamente ancho, teorema de consistencia de Kolmogorov - Diseño de un GP: Función de media y kernel de covarianza - Representación espectral - Entrenamiento y complejidad computacional - Sparse GPs	-Entender el proceso gaussiano (GP) como el límite infinito-dimensional de modelos paramétricos -Diseñar un GP e implementarlo con datos reales -Comprender y resolver los costos computacionales de GPs	Ver notas de referencia en cada contenido

Número	Nombre de la Unidad	Duración en Semanas
7	Reducción de dimensionalidad	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
- Análisis de componentes principales: lineal, probabilístico y de kernel - LDA/ANOVA/ICA	-Conocer la importancia de las técnicas de reducción de dimensionalidad para preprocesamiento de datos, visualización e interpretación. -Dominar distintas técnicas de reducción de dimensionalidad y saber cuándo usarlas	Ver notas de referencia en cada contenido

Número	Nombre de la Unidad	Duración en Semanas
8	Clustering	2
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
• K-medias • DBSCAN: <i>Density-based spatial clustering of applications with noise</i> • <i>Kernel density estimation</i> [MLPP] • Mezcla de gaussianas: El algoritmo <i>expectation maximisation</i> [MLPP] • Nota sobre inferencia variacional [VI, MLPP]	-Entender la diferencia entre métodos paramétricos y no-paramétricos -interpretar <i>clustering</i> como el diseño de un modelo generativo -ocupar un enfoque variacional para resolver modelos intratables (inferencia aproximada)	Ver notas de referencia en cada contenido

Bibliografía General
[IAMA] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 1994. Prentice Hall. [MA] P. Domingues, The Master Algorithm, 2015. Basic Books. [LFD] Y. Abu-Mostafa, M. Magdon-Ismail and H. Lin, Learning from data, 2012. AMLBook [ISL]: G. James, D. Witten, T. Hastie and R. Tibshirani, Introduction to statistical learning, 2013. Springer. [BDA]: A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, Bayesian Data Analysis, 2013. CRC Press. [ITILA]: D. MacKay, Information theory, inference and learning algorithms, 2003. Cambridge. [MSMA]: G. Claeskens and N. L. Hjort, Model selection and model average, 2008. Cambridge. [PRML]: C. Bishop, Pattern recognition and machine learning, 2006. Springer. [GP4ML]: C. Rasmussen and C. Williams, Gaussian processes for machine learning, 2006. MIT. [MLPP]: K. Murphy, Machine learning: A probabilistic perspective, 2012. MIT. [ESL]: T. Hastie, R. Tibshirani, J. Friedman, Elements of statistical learning, 2009. Springer. [LWK]: B. Schölkopf and A. Smola, Learning with kernels, 2002. MIT. [UDL]: Y. Gal, Uncertainty in Deep Learning, 2015. http://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf [VI]: D. M. Blei, A. Kucukelbir and J. D. McAuliffe, Variational Inference: A Review for Statisticians, 2015. https://arxiv.org/abs/1601.00670.

Vigencia desde:	
Elaborado por:	Felipe Tobar
Revisado por:	