

CICLO DE COMPLEMENTACIÓN CURRICULAR

LICENCIATURA EN CIENCIA DE DATOS

PROGRAMA

00191 – Inteligencia Artificial

Lic. G. Sebastián PEDERSEN

1. Cantidad de horas semanales y totales

Carga Horaria Total: 96 horas.

- Teoría: 48 horas.
- Práctica: 48 horas.

Carga Horaria Semanal: 6 horas.

2. Nombres de las/los integrantes del equipo docente

Docente: Lic. G. Sebastián Pedersen

3. Fundamentación

Al llegar a este curso los y las estudiantes ya poseen conocimientos básicos de Análisis Matemático, Álgebra y Probabilidades y Estadística. Además ya han pasado por materias como Inferencia Estadística y Reconocimiento de Patrones, Análisis Multivariado y Programación Avanzada, las cuales les dotan de herramientas suficientes en las cuales poder construir conocimiento más avanzado y firme sobre Aprendizaje Automático.

Siendo esta una materia del segundo cuatrimestre del tercer año, es un buen momento para que les estudiantes utilicen todo el conocimiento previo adquirido en materias anteriores, y puedan desde esa base potenciar tanto el mejor entendimiento de temas ya vistos como profundizar en temas fundamentales aún no vistos.

Esta materia se propone dar un cierre a las técnicas de Aprendizaje Automático (Machine Learning) contemporáneas, a la vez que terminar de potenciar la independencia de criterio y habilidades de les estudiantes, brindando una sólida base sobre la cual recae casi cualquier problemática actual de Ciencia de Datos.

4. Programa sintético

Trade-off entre sesgo y varianza. Regularización. Selección de Modelos. Riesgo empírico. Train/Valid/Test split, cross-validation.

Regresión logística como red neuronal. Introducción a las redes neuronales. Redes neuronales, back propagation. Descenso por gradiente con variantes. Redes neuronales Profundas. Redes Convolucionales. Árboles de decisión. Bagging, boosting, ensembles.

Aprendizaje no supervisado. Clustering por K-means, mezcla de gaussianas, expectation-maximization.

5. Objetivos

Al finalizar el curso el o la estudiante estará en capacidad de:

- Abordar un problema de Machine Learning y entender la parte teórica.
- Abordar un problema de Machine Learning y, partiendo desde haber entendido la parte teórica, poder realizar una implementación práctica programando alguna solución.
- Identificar los posibles escenarios adecuados para utilizar las variadas herramientas que brinda el Machine Learning y el Deep Learning.
- Entender las dificultades y beneficios relacionadas con los datos, y su impacto a la hora de resolver un problema de Machine Learning o Deep Learning..
- Proponer mejoras o modificaciones a modelos predictivos actuales del machine/deep learning.
- Ser capaz de leer y entender de forma independiente un artículo científico de vanguardia de machine/deep learning.
- Interpretar adecuadamente las diferentes modificaciones en cuanto a hiperparámetros se refiere, que los modelos (e implementaciones de los mismos) de machine/deep learning poseen.
- Realizar un informe de carácter profesional respecto al análisis tanto de un problema como de su solución e implementación práctica, de machine/deep learning.
- Realizar una exposición científica de un problema y su solución, de machine/deep learning.

6. Propósitos de la enseñanza

El curso es teórico-practico, donde el estudiante debe desarrollar problemas en grupo e individualmente, que permita poner en práctica los conocimientos adquiridos.

Las prácticas serán definidas por parte del profesor el cual dará únicamente los lineamientos generales a seguir. El o los problemas escogidos y su solución aplicada deben ser volcados a un informe y expuestos por los grupos respectivos.

7. Contenidos

Los contenidos del programa estarán resumidos en los siguientes módulos:

- Selección de modelos: trade-off sesgo/varianza. Regularización. Riesgo empírico. Train/Valid/Test split y cross-validation. Implementación en Python.
- Redes neuronales clásicas: regresión logística como red neuronal. Intro redes neuronales. Back Propagation. Descenso por gradientes (varias variantes). Implementación en Python.
- Redes Neuronales profundas: clásicas fully-connected. Redes Convolucionales. Implementación en Python.
- Clasificación no supervisada: K-means. Mezcla de gaussianas. Expectation-Maximization. Implementación en Python.
- Árboles de decisión: bagging. Boosting. Ensamblados. Random forest.

8. Bibliografía y recursos audiovisuales

Bibliografía Obligatoria:

Hastie, T.; Tibshirani, R; Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2nd. ed.

Ian Goodfellow, Yoshua Bengio, Aaron Courville. (2016). *Deep Learning*. MIT Press.

Bibliografía Complementaria:

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Ng, Andrew. (2018). *Machine Learning Yearning*. Deeplearning.ia.

Marc P. Deisenroth, A. Aldo Faisal, Cheng Soon Ong. (2020). *Mathematics for Machine Learning*. Cambridge University Press.

9. Metodología

La cátedra estará compuesta por clases teórico-prácticas presenciales, ejercitaciones y trabajos prácticos grupales.

Las clases teórico-prácticas, estarán divididas del siguiente modo:

- Exposición de los conceptos teóricos y debate sobre los temas dados.
- Ejercitaciones para fijar los contenidos.

Trabajos prácticos:

- Desarrollo de trabajos prácticos de investigación.
- Trabajos individuales y grupales
- Seguimiento de trabajos prácticos desarrollados en lenguaje Python.

10. Uso del campus virtual e integración de TIC en la propuesta pedagógica

La transformación de la educación durante los últimos tiempos, nos apremia a poner foco y maximizar el uso de TIC's para implementar la enseñanza "virtual" o "a distancia", preparándonos para la educación "post-pandemia".

El uso del "Campus Virtual" continuando la propuesta pedagógica, será el portal de conocimiento en donde los alumnos podrán la propuesta educativa y sus recursos. Complementando su uso, con otras tecnologías para mejorar la participación e involucramiento de los estudiantes, tales como, clases video-presenciales, herramientas on-line para la resolución de problemas, y recursos mediáticos para reforzar la propuesta educativa.

Finalmente, utilizar el medio para continuar y reforzar el trabajo de los alumnos en grupos, ya que sin un "aula física" la formación de "grupos", e interacciones entre los estudiantes es reducida.

11. Evaluación

La evaluación corresponderá a una nota global entre:

- Entrega de ejercicios.
- Dos trabajos prácticos (con derecho a sendos recuperatorios)

12. Cronograma de actividades / Planificación de clases

A continuación se presenta el cronograma de clases correspondiente al primer cuatrimestre de 2022. Las clases se desarrollan los Lunes de 18 a 22 hs.

Fecha	Temas	Hitos
lun 22/ago	Sesgo/varianza; Regularización; selección de modelos.	Lanzamiento del TP1
lun 29/ago	Sesgo/varianza; riesgo empírico.	---
lun 05/sep	Regresión Logística como Red Neuronal; intro a las Redes Neuronales.	---
lun 12/sep	Consultas/Repaso	---
dom 18/sep	---	Límite de entrega del TP1
lun 19/sep	Regresión Logística como Red Neuronal; intro a las Redes Neuronales.	Lanzamiento del TP2
lun 26/sep	Redes Neuronales; back propagation; mejoras.	---
lun 03/oct	Redes Neuronales; back propagation; mejoras.	Lanzamiento de la búsqueda del paper a exponer
lun 10/oct	FERIADO	FERIADO
lun 17/oct	Aprendizaje No supervisado; k-means; mezcla de gaussianas; Expectation-Maximization.	---
lun 24/oct	Aprendizaje No supervisado; k-means; mezcla de gaussianas; Expectation-Maximization.	---
lun 31/oct	Consultas/Repaso	---
dom 06/nov	---	Límite de entrega del TP2
lun 07/nov	---	Exposición del paper
lun 14/nov	---	Exposición del paper
lun 21/nov	FERIADO	FERIADO
lun 28/nov	---	Exposición del paper
lun 05/dic	---	Semana de recuperatorios
sáb 10/dic	---	Último día de clases.