



APRENDIZAJE ESTADÍSTICO

Carrera: *Maestría en Inteligencia de Datos orientada a Big Data*

Carga Horaria: 64 Hs.

Año 2019

Profesor a Cargo: *Dra. Laura Lanzarini,
Dr. Alejandro Rosete, Lic. Facundo Quiroga*

OBJETIVO

Brindar al alumno los conocimientos necesarios para construir modelos estadísticos y probabilísticos a partir de datos. Se espera que, quienes aprueben el curso, estén capacitados para seleccionar los modelos adecuados para el tratamiento y solución correcta de un problema dado.

Se presentarán modelos predictivos básicos: regresión múltiple y logística, y las técnicas estimación bayesiana y de máxima verosimilitud. También se analizarán algunos modelos probabilísticos como las redes bayesianas o de markov.

MODALIDAD DE EVALUACION

Se realizarán ejercicios prácticos durante la cursada, con una evaluación final con modalidad examen o trabajo final.

PROGRAMA

Parte I – Enfoque estadístico al aprendizaje automático.

- Breve repaso de probabilidad, variables y vectores aleatorios, probabilidad condicional, teorema de Bayes, estimación puntual, distribuciones discretas y continuas.
- Generalización, sobreajuste y sobregeneralización, separación lineal, regularización, capacidad de un clasificador. Sesgo y varianza. Funciones de pérdidas (loss). Riesgo empírico y esperado. Elementos de la teoría de Vapnik–Chervonenkis. Dimensión VC.



- Regresión lineal simple y múltiple. Estimación de mínimos cuadrados. Regresión logística. Regresión logística multinomial. Modelos Generales Lineales (General Linear Models, GLM). Estimación bayesiana.
- Evaluación de modelos. Diagnósticos de influencia. Indicadores de ajuste y comparación. Métodos gráficos: curvas ROC, gráficos de ganancia y de riesgo. Validación cruzada. Técnicas bootstrapping, jackknifing y pruebas de permutación.
- Métodos de selección de variables.

Parte II – Modelos probabilísticos

- Factores. Distribuciones. Normalización. Evidencia Factorización de una distribución. Distribuciones asociadas. Marginalización y maximización. Regla de la cadena.
- Los tres problemas asociados a un modelo probabilístico: aprendizaje (estimación de parámetros), muestreo, e inferencia (consultas de probabilidades como marginalización y maximización). Soluciones de fuerza bruta.
- Modelos gráficos probabilísticos. Independencias en un modelo gráfico probabilístico. Mapas de independencia. Mapas perfectos. Mapas minimales. Equivalencia de independencias.
- Modelos de independencia: Naive Bayes, Redes Bayesianas, Redes Markovianas, Condición de Markov, Modelos de Markov, Modelos Ocultos de Markov.
- Aprendizaje: Estimación de parámetros y aprendizaje de estructura. Estimación de independencias.
- Algoritmos de muestreo: muestreo para adelante, muestreo por importancia, muestreo de gibbs.
- Algoritmos de inferencia: suma-producto y máximo-producto. Algoritmos aproximados y exactos.

BIBLIOGRAFIA

- Friedman, J., Hastie, T., & Tibshirani, R. *The elements of statistical learning*. New York: Springer series in Statistics, 2009.
- Abu-Mostafa, Yaser S., Malik Magdon-Ismael, and Hsuan-Tien Lin. *Learning from Data: A Short Course*. New York, NY, USA. AML Book, 2012.



- Guyon, Isabelle, and André Elisseeff. “An introduction to variable and feature selection”. *Journal of machine learning research* 3.Mar (2003): 1157-1182.
- Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- Barber, David. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.
- MacKay, David JC. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
- Koller, Daphne, and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- Wainwright, Martin J., and Michael I. Jordan. "Graphical models, exponential families, and variational inference." *Foundations and Trends in Machine Learning* 1.1–2 (2008): 1-305.