



Universidad Nacional Autónoma de
México
Temario Estadística Computacional y
Machine Learning
Prof. Jimmy Hernández Morales
Porf. Adjunto Rodrigo Quijón
Hipolito

El curso pretende ser una introducción a lo que actualmente se conoce como Machine Learning, que son un conjunto de técnicas estadísticas que junto con el desarrollo de algoritmos computacionales han tenido mucho auge en los últimos años. Todos los temas serán estudiados desde el punto de vista estadístico formal, pero haciendo mucho énfasis en la implementación computacional. El lenguaje principal que usaremos será Python o Java y como auxiliar R. Las bibliotecas esenciales que usaremos son Matplotlib, Scipy, Numpy, Pandas, Scikit Learn y MLlib. Se calificará mediante tareas y proyectos.

Temario

1. Regresión Lineal y Logística

- 1.1 Regresión Lineal Simple
- 1.2 Regresión Múltiple
- 1.3 Selección de Modelos
 - 1.3.1 Criterios AIC, BIC y divergencia de Kullback-Leibler
- 1.4 Regresión Logística

2. Aprendizaje Supervisado (Clasificación)

- 2.1 Clasificadores Lineales y Gaussianos.
- 2.2 Regresión Lineal y Logística
- 2.3 Discriminante Lineal (LDA)
- 2.3 Naive Bayes
- 2.4 Máquinas de soporte vectorial
- 2.5 Kernelización
- 2.6 Ejemplos, caso de estudio: redes neuronales.

3. Modelos aditivos, árboles y Boosting

- 3.1 Modelos aditivos Generalizados.
 - 3.2 Árboles de regresión
 - 3.3 Árboles de clasificación
 - 3.4 Métodos boosting
 - 3.5 Ajuste boosting y modelos aditivos.
-

3.5 Árboles boosting

4. Aprendizaje no supervisado(Clustering)

4.1 Reglas de asociación

4.2 Analisis de Cluster

4.3 Componentes principales.

5. Estadística de Altas Dimensiones y Big Data

5.1 Algoritmos de optimización

5.1.1 Gradiente descendiente

5.1.1 Coordenadas descendientes

5.2 Regresión Contraída

5.2.1 Regresión Lasso

5.2.1 Regresion Ridge

5.3 Correlación Espuria

5.4 Endogeneidad incidental

5.4.1 GMM

5.4.2 FGMM

5.5 Acumulación de ruido

Bibliografía

- [1] HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J The Elements of Statistical Learning. Data Mining, Inference, and Prediction, 2nd ed., Springer.
- [2] G. JAMES, D. WITTEN, T. HASTIE AND R. TIBSHIRANI An Introduction to Statistical Learning, with Application in R (Springer, 2013).
- [3] DIGGLE, P Statistical Learning with Sparsity. The lasso and generalizations. Chapman and Hall.
- [4] T. W. ANDERSON An Introduction to Multivariate Statistical Analysis, John Wiley and Sons Inc; Edición: 3rd Edition .
- [5] MICHIE, SPIEGELHALTER Y TAYLOR y PETRICK J. HEAGARTY , Machine Learning, Neural and Statistical Classification.
- [6] ALAN AGRESTI , Categorical Data Analysis, *Wiley-Interscience Publication*.
- [7] THOMAS W. YEE, Vector Generalized Linear and Additive Models with implementations in R, *Springer*.
- [8] ANDREAS MULLER, SARAH GUIDO, Introduction to Machine Learning with Python, a guide for data Scientist *Springer*.