

Profesor: Alvaro J. Riascos Villegas

Contacto: e-mail: ariascos@uniandes.edu.co, Skype: alvaro.riascos

Oficina: Bloque W, Oficina 918.

Horario de clase: Martes 2PM-3:20PM

Horario de atención a estudiantes: 11:30AM – 12:30PM, con cita previa o por skype.

Página en Internet del curso: www.alvaroriascos.com

1. Objetivos de la materia

Este seminario introduce los estudiantes en los pilares teóricos fundamentales de la teoría de aprendizaje estadístico como marco teórico de la minería de datos (el problema de aprendizaje o *machine learning*, el compromiso entre sesgo y varianza, aproximación y error, riesgo, consistencia, regularización, complejidad, etc.) Paralelamente se van aprender las principales técnicas de minería de datos (método de vecindades, redes neuronales, redes bayesianas, árboles, *boosting*, *cross validation*, máquinas de vectores de soporte, *clustering*, reglas de asociación, etc.) a través de ejemplos y problemas que los estudiantes deberán implementar y resolver haciendo uso de un computador y, en lo posible, resolviendo problemas que sean de su interés (análisis de datos de redes sociales, reconocimiento de caracteres, extracción de señales, pronósticos, clasificación de clientes, alertas de fraude, anomalías, etc.).

Este curso es muy práctico y requiere de la participación constante de los estudiantes, haciendo presentaciones de diferentes técnicas, talleres, haciendo un proyecto final y participando de un concurso en Kaggle (<http://www.kaggle.com/>). Para facilitar la aplicabilidad de las técnicas aprendidas habrá una introducción corta a R (<http://www.r-project.org/>) y parte de las técnicas se aprenderán de forma paralela con su implementación en este lenguaje de programación y sus respectivos paquetes.

2. Contenido

1. **Introducción**
2. **Introducción a R (Miguel Bernal)**
3. **Aprendizaje Estadístico I: Modelos, Conceptos y Resultados**
[LS]
[JWHT] - Capítulo 1 y 2.
4. **Regresión Lineal (G1)**
[JWHT] - Capítulo 3.
5. **Clasificación (G2)**
[JWHT] - Capítulo 4.

ENTREGA TALLER I

6. **Resampleo (G3)**
[JWHT] – Capítulo 5
7. **Minería de Texto y Redes I (Miguel Bernal)**
8. **Minería de Texto y Redes II (Miguel Bernal)**
9. **Modelos Lineales y Selección de Modelo (G4)**
[JWHT] – Capítulo 6
10. **Twitter y el Mercado de Valores (Diego Jara)**
11. **Semana Santa**
12. **Árboles (G5)**
[JWHT] – Capítulo 8
- ENTREGA PROPUESTA DE PROYECTO**
13. **Aprendizaje Profundo (Sergio Camelo)**
- ENTREGA TALLER II**
14. **Máquinas de Vectores de Soporte**
[JWHT] – Capítulo 9
15. **Aprendizaje No Supervisado (G6)**
[JWHT] – Capítulo 10
16. **Minería de Datos en la DIAN (Carlos Patiño)**

3. Metodología

Este curso es muy práctico y requiere de la participación intensa de los estudiantes para su desarrollo. Los estudiantes tendrán que formar grupos (de 3 o máximo 4) para hacer cinco entregas:

- Taller 1 (30% de la nota para los que no hayan presentado a la fecha y 15% para quienes hayan presentado siendo el otro 15% de estos últimos la presentación realizada)
- Taller 2 (20% de la nota para los que no hayan presentado a la fecha y 10% para quienes hayan presentado siendo el otro 10% de estos últimos la presentación realizada)
- Propuesta proyecto (10% de la nota): Máximo dos páginas incluyendo tablas, gráficos, bibliografía, etc.
- Proyecto final (30% de la nota): Máximo seis páginas incluyendo tablas, gráficos, bibliografía, etc.
- Proyecto Kaggle (10% de la nota): <http://www.kaggle.com/>¹

¹ El mejor grupo en el concurso tendrá una nota de 5, el Segundo mayor de 4.5, el tercer 4.0, el cuarto 3.5 y el quinto 3.0.

4. Sistema utilizado para aproximar la nota definitiva

El sistema de notas definitivas es el siguiente: las notas totales con decimales en 0 o en .5 no se modificarán. Las notas totales con decimales entre .25 a .49 y entre .75 a .99, se aproximarán a la nota definitiva siguiente. Las notas con decimales entre .01 a .24 y entre .51 a .74, se aproximarán a la nota definitiva anterior.

5. Bibliografía

Referencias obligatorias

[LS]: Luxburg, U., B. Scholkopf. 2008. Statistical Learning Theory: Models, Concepts and Results.

[JWHT]: Introduction to Statistical Learning with Applications in R.

Descargable de: <http://www-bcf.usc.edu/~gareth/ISL/>

Referencias Adicionales

[HTF]: Hastie, T., Tibshirani, R. y J. Hastie. 2009. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Segunda Edición. Springer