

Profesor: Alvaro J. Riascos Villegas

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

Oficina: Bloque W, Oficina 918.

Horario de atención a estudiantes: Martes 3:30PM – 4:30PM, 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, minería de texto, 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 activa 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

Semana	Tema	Expositor	Referencias
1	Introducción a la Minería de Datos	Alvaro Riascos	[JWHT]: Capítulo 1. [HTF]: Capítulo 1.
2	Aprendizaje Estadístico I: Modelos, Conceptos y Resultados	Alvaro Riascos	[LS] [JWHT]: Secciones 2.1, 2.2.
3	Aprendizaje Estadístico II: Modelos, Conceptos y Resultados	Alvaro Riascos	[LS] [JWHT]: Secciones 2.1, 2.2.
4	Métodos lineales de clasificación Aplicaciones Entrega Taller I	Alvaro Riascos	[JWHT]: Capítulo 3,4. [HTF]: Capítulo 3,4.

5	Métodos no lineales y regularización	Alvaro Riascos	[HTF]: Capítulo 5,6
6	Métodos no lineales y regularización II Aplicaciones Entrega Taller II	Alvaro Riascos	[HTF]: Capítulo 5,6
7	<i>Arboles, Boosting</i> de Árboles y <i>Random Forests</i> I Última Semana 30% Notas	Alvaro Riascos	[HTF]: Capítulos 9, 10 y 15
8	<i>Arboles, Boosting</i> de Árboles y <i>Random Forests</i> II Aplicaciones	Alvaro Riascos	[HTF]: Capítulos 9, 10 y 15
9	Aprendizaje Estadístico III: Selección de Modelos e Inferencia	Alvaro Riascos	[HTF]: Capítulos 7 y 8 [JWHT]: Capítulo 5
10	Minería de Texto I	Francisco Barreras	
11	Minería de Texto II Entrega Taller III	Francisco Barreras	
12	SEMANA SANTA		
13	Análisis no supervisado Primera Evaluación Concurso Kaggle	Alvaro Riascos	
14	Análisis no supervisado	Felipe Suarez	
15	Redes Neuronales y Aprendizaje Profundo y Aplicaciones Segunda Evaluacion Concurso Kaggle	Monica Ribero	Redes Neuronales y Aprendizaje Profundo y Aplicaciones Primera Evaluacion Concurso Kaggle
16	Máquinas de Vectores de Soporte Propuesta de Proyecto Final	Fabian Latorre	[HTF]: Capítulos 12 y 16 [JWHT]:

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 mínimo persona a máximo cuatro personas) para hacer las siguientes entregas:

- Taller 1 (15% de la nota)
- Taller 2 (15% de la nota)
- Taller 3 (15% de la nota)
- Proyecto Kaggle primera evaluación (10% de la nota): <http://www.kaggle.com/>¹
- Proyecto Kaggle segunda evaluación (10% de la nota): <http://www.kaggle.com/>²
- Propuesta de proyecto final (10% de la nota).
- Proyecto final (25% de la nota): Máximo seis páginas incluyendo tablas, gráficos, bibliografía, etc.

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

Las referencias principales del curso son:

[LS]: Luxburg, U., B. Scholkopf. 2008. Statistical Learning Theory: Models, Concepts and Results.
<http://arxiv.org/abs/0810.4752>

[JWHT]: Introduction to Statistical Learning with Applications in R.
<http://www-bcf.usc.edu/~gareth/ISL/>

[HTF]: Hastie, T., Tibshirani, R. y J. Hastie. 2009. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Segunda Edición. Springer
http://web.stanford.edu/~hastie/local.ftp/Springer/OLD/ESLII_print4.pdf

Las referencias de las aplicaciones se irán subiendo a la página web del curso a lo largo del semestre.

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

² El mejor grupo en el concurso tendrá una nota de 5, el segundo mejor de 4.5, el tercer 4.0, el cuarto 3.5 y el quinto 3.0.