

**Exposiciones por parte de los alumnos (EEA2015):**

Duración aproximada de cada exposición: 30 minutos más 10 para preguntas.  
En total son 40 minutos por grupo (30+10)

**FECHAS DE PRESENTACIONES: 14/11 y 5/12**

TEMAS	Grupos
Datos desbalanceados (Pablo)	6 y 5
Aplicaciones de curvas ROC (Pablo)	8 y 10
Principal Components Regression y Partial Least Squares (Daniel)	1 y 9
Datos faltantes (Daniel)	3 y 4
Regresión Ridge y Lasso (Silvia)	2 y 7
Modelos Lineales Generalizados en DM (Silvia)	11 y 12

<b>Grupo 1:</b> BRESCIANI Juan Pablo CAVALLO Juan COPPOLILLO Nestor Raul INZERILLI Carla SOKIL Juan Pablo	<b>Grupo 2:</b> AGUIRRE Augusto Ariel DELL'ERA Diego Mario FARIAS Damián Gustavo MARAFIOTI Pablo Ezequiel
<b>Grupo 3:</b> BESIO Mariano Andres GARCIA Gonzalo Daniel MUSCHITIELLO Nicolas Miguel Felix VAINMAN Daiana NOVIDELSKY	<b>Grupo 4:</b> MOREIRA PERA Martín Andrés OJEDA Daniel RODRIGUEZ HERNANDEZ Jhonny ROMERO QUISHPE Adriana Lucía VELEZ DIAZ Alvaro Andres
<b>Grupo 5:</b> AVELLANEDA MEJIA Leandro MONTERO BERMUDEZ Eduardo José NARVAEZ GOMEZ Giovanni Alberto TORRES REYES Jenny Rocío VELEZ LUNA Roberto Daniel	<b>Grupo 6:</b> CHOI Da Woon MARTINEZ Pablo Witold RUSCONI Ivo
<b>Grupo 7:</b> ANDREONI Pablo Facundo GONZALEZ Fernando Ezequiel	<b>Grupo 8:</b> BEATI María Paula BIANCO Javier Antonio FABIANO Ariel Alejandro GAVILAN Sebastián Aníbal GAY Rafael Lucas HERZ Gerardo Sebastian
<b>Grupo 9:</b> DE RACO Sergio Andrés GALBAN Sebastian Federico MATOS Jorge Luis RIVERO GOYTIA Matias WATFI Pablo	<b>Grupo 10:</b> CORAS AVILA Nilo Adolfo GUEVARA LENIS Jorge Eduardo OYOLADiego Hernán SPAIRANI Julio Augusto VERALLI Eduardo Damian
<b>Grupo 11:</b> ACOSTA CASTAÑEDA Diego Alberto GASKAMaria GOMEZ BERET Flavia Matilde JIMENEZ FORERO Jairo Andrés LIBERMAN Gastón Elías	<b>Grupo 12: ;?</b>

## **Desarrollo:**

El trabajo se enfoca a estudiar el tema propuesto, armando un documento con una explicación del tema. Deben además relevar software disponible para aplicar la técnica, presentar un problema-ejemplo y resolverlo. Puede ser un ejemplo de libro, no es necesario que sea original sino que sepan explicarlo.

La presentación oral se hará en una clase frente al resto de los alumnos y docentes.

Los temas estarán presentes en la evaluación del examen integrador.

## **Temas propuestos**

### **Principal Components Regression y Partial Least Squares: Comparación de métodos (DANIEL)**

Objetivo: presentar el modelo, contando el significado teórico. Luego en un ejemplo mostrar el ajuste con algún software.

SW posible: SPSS + JMP + E. Miner

Bibliografía posible:

- Handbook of Statistical Analysis and Data Mining Applications (Lisbet, Elder, Miner)
- Hastie, Trevor; Tibshirani, Robert y Friedman, Jerome. "The Elements of Statistical Learning – Data Mining, Inference, and Prediction". Springer, Nueva York, 2001
- Regression using JMP - R. Freund, R. Littell, L. Creighton (2003)

### **Datos faltantes (DANIEL)**

Objetivo: recordar los inconvenientes relacionados con la presencia de gran cantidad de valores ausentes (missings) que aparecen en los conjuntos de datos a analizar en Data Mining. Plantear alternativas univariadas, bivariadas y multivariadas para imputar. Explicar con ejemplos las ventajas y desventajas de las diferentes formas de imputación.

SW posible: SPSS, Enterprise Miner, Weka, R

Bibliografía posible:

- Statistical Analysis with missing data. Little & Rubin
- F. Medina, M. Galván (2007) - Imputación de Datos: Teoría y Práctica (CEPAL))
- D. Stekhoven, P. Bühlmann (2011) – MissForest Non-parametric missing value imputation or mixed-type data (BIOINFORMATICS)
- Xiaofeng Zhu, Shichao y ZhangZhi Jin(2001) - Missing Value Estimation for Mixed-Attribute Data Sets

### **Datos desbalanceados (PABLO)**

Objetivo: plantear los problemas que se presentan al trabajar con datos desbalanceados, y mostrar algunas de las estrategias útiles para combatirlos. Explicar con ejemplos cuáles son algunos de los métodos que funcionan y cuáles no.

SW posible: SPSS, Enterprise Miner, R, MODELER

Bibliografía posible:

Applied Multivariate Statistics With Sas Software. Ravindra Khattree,Dayanand N. Naik  
Jmp 8 Statistics and Graphics Guide, Second Edition. SAS Institute,SAS Publishing  
Nathalie Japkowicz (2002). The Class Imbalance Problem: A Systematic Study

### **Aplicaciones de curvas ROC (PABLO)**

Objetivo: Explicar beneficios y problemas que puede generar utilizar estos Roc o accuracy al medir los resultados de los modelos. ¿Los algoritmos para clasificación que utilizan accuracy pueden tener algún inconveniente? Plantear alternativas.

SW posible: R, SPSS, MODELER

F. Provost, T. Fawcett (1997). Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions. *Third International Conference on Knowledge Discovery and Data Mining* (KDD-97)

T. Fawcett (2004). ROC graphs: Notes and practical considerations for researchers.

T. Fawcett (2006). An introduction to ROC analysis

T. Fawcett, P. Flach (2005). A Response to Webb and Ting's *On the Application of ROC Analysis to Predict Classification Performance under Varying Class Distributions*

### **Modelos Lineales Generalizados en DM. (SILVIA)**

Objetivo: presentar el modelo, por ejemplo Regresión Poisson, contando el significado teórico. Luego en un ejemplo mostrar el ajuste con algún software.

SW posible: R, SAS, SPSS..

Bibliografía posible:

Tufféry, S. (2011). Data Mining and Statistics for Decision Making. UK. John Wiley & Sons

Montgomery, Peck & Vining

### **Regresión Ridge y Lasso (SILVIA)**

Objetivo: presentar el modelo, contando el significado teórico. Luego en un ejemplo mostrar el ajuste con algún software.

SW posible: SPSS + SAS

Bibliografía posible:

Applied Regression (Draper) + Regresión Lineal (Montgomery) + Alternative Methods of Regression (Birkes, Dodge) + Classical and Modern Regression (Myers) + Modern Regression Methods (Ryan)

Hastie, Trevor; Tibshirani, Robert y Friedman, Jerome. "The Elements of Statistical Learning – Data Mining, Inference, and Prediction". Springer, Nueva York, 2001.