

**MAESTRÍA EN MATEMÁTICA**  
**Seminario Álgebra Lineal Aplicada y Manejo de Datos 2015**

**"Imputación de Datos Faltantes"**

Lic. María Cristina Werenitzky Curia

**Resumen:** La falta de datos o datos erróneos en una base de datos dificulta el análisis de los mismos. Debido a esto, se aplican todo tipo de métodos iterativos o heurísticos que permitan la imputación de datos faltantes en las bases de datos para luego poder realizar análisis y/o modelos de los datos que sean de utilidad. (Schneider, 2001). En este seminario se presenta el método iterativo Expectation Maximization (EM) para realizar la imputación de datos faltantes en una Base de Datos Ionosféricos.

El Algoritmo EM es un proceso iterativo que se basa en el Análisis de Componentes Principales (PCA) - o Descomposición en Valores Singulares (SVD). Se utiliza en el caso de que exista una fuerte correlación lineal entre las variables involucradas. Este algoritmo se beneficia de dicha correlación para encontrar los valores de relleno. El método completa la matriz con valores iniciales, y en cada iteración, calcula la descomposición SVD de la matriz, para luego rearmar la matriz con un número determinado de valores singulares y obtener así una nueva estimación de los valores faltantes. La elección de la cantidad de valores singulares utilizados para recomponer la matriz no es un detalle menor y representa un problema a la hora de automatizar el método. Se presenta una variante del método EM que permitiría mejorar la precisión de las estimaciones y permitir la automatización del mismo.