

Aspectos teóricos:

1.- CRISP-DM es una metodología robusta y ampliamente probada en la industria. Su finalidad es brindarnos un enfoque para resolver un proyecto de minería de datos. Su enfoque es iterativo, por lo que estas etapas pueden repetirse según sea necesario.

Etapas:

1.- La primera etapa es la del entendimiento del negocio. En esta etapa se establecen objetivos y criterios de negocio, los cuales habrán de ser resueltos con minería de datos.

2.- La segunda etapa es la que se conoce como “Data Understanding”, en esta etapa se exploran los datos que ya se tienen, se verifica la calidad de los datos y se hacen algunos experimentos para poder llegar a comprender mejor los datos que ya se tienen.

3.- La etapa tres consiste en preparar los datos, en caso de que se necesiten limpiar los datasets, se hace en esta etapa, si se van a agregar o incluir datos adicionales en esta etapa es donde esto se lleva a cabo.

4.- La cuarta etapa consiste en construir el modelo que utilizaremos para correrlo en nuestros datos. En esta etapa se decide si utilizaremos, por ejemplo, redes neuronales, o cual algoritmo de clasificación, por ejemplo, es mejor para nuestro dataset. Esto tomando en cuenta, el tiempo que toma entrenar al modelo, su eficacia (“accuracy”) etc.

5.- En la quinta etapa se evalúan los resultados de todas las etapas anteriores y se toman decisiones sobre si se hará alguna otra iteración o si ya se implementa el modelo.

6.- La ultima etapa es la de implementación, aquí se construye un plan de implementación, y se implementa el proyecto. Adicionalmente se monitorean los datos y que el modelo esté funcionando como se esperaba. Se documentan los resultados.

## 2.- Tipos de aprendizaje

Dependiendo del tipo de problema que queramos resolver, se puede optar por el aprendizaje supervisado, el no supervisado o el aprendizaje por refuerzo.

El que hemos estado viendo en la clase hasta ahora ha sido el aprendizaje supervisado, es decir que tenemos datos de cierta forma estructurados y con sus respectivas etiquetas. El objetivo final es predecir ciertas salidas, ejemplo de ello pudiera ser, si un estudiante va a incumplir su crédito, predecir las ventas del próximo año con base a ciertos factores, determinar a que especie pertenece una planta etc.

Con el aprendizaje no supervisado, se trabajan con datos sin etiquetar, la finalidad principal de este tipo de aprendizaje es descubrir patrones subyacentes en los datos. Algunas de las aplicaciones pueden ser, por ejemplo, recomendaciones en videos de YouTube, detección de anomalías, agrupaciones como por ejemplo de clientes con intereses similares, clasificación de documentos etc.

Por último, en el del aprendizaje por refuerzo, se pueden resolver problemas como los automóviles que aprenden a manejar de manera autónoma, estos tampoco

requieren de supervisión. Y nos pueden ayudar a resolver algunos problemas de exploración por ejemplo.

### 3.- Explica el Batch Gradient Descent:

El BSG se utiliza en conjunto con la técnica de regresión logística.

Por cada parámetro (feature) que se quiere predecir, su valor debe ser multiplicado por una beta, esta es la función de la regresión logística.

A su vez, podemos definir una función de costo, para determinar que tan preciso es nuestro modelo.

El algoritmo de la gradiente descentente, nos ayuda a minimizar nuestra función de costo de manera iterativa y a encontrar el mínimo local de nuestra función.

Basicamente el algoritmo lo que hace es, calcula la pendiente de nuestra función (haciendo uso de la primera derivada) para posteriormente moverse en la dirección opuesta de la pendiente (usando el learning rate) y poder calcular una gradiente. Esto se debe hacer de manera iterativa hasta encontrar un punto de convergencia, especificando un threshold. Cuando el movimiento de la gradiente sea menor al threshold, se llega al punto de convergencia.

4.-

$$4: P = \frac{1}{1 + e^{-\vec{\beta} \cdot \vec{x}}} \quad J = - \sum_{i=1}^n y_i [\ln(P_i(\vec{\beta})) + \ln(1 - P_i(\vec{\beta}))]$$

$$\frac{\partial J}{\partial \beta_j} = \sum_{i=1}^n \left[ y_i \frac{1}{P_i(\vec{\beta})} \frac{1}{P_i(\vec{\beta})} \frac{\partial P_i(\vec{\beta})}{\partial \beta_j} - \frac{1}{1 - P_i(\vec{\beta})} \frac{\partial P_i(\vec{\beta})}{\partial \beta_j} \right]$$

$$\frac{\partial P_i(\vec{\beta})}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \left( \frac{1}{1 + e^{\vec{\beta} \cdot \vec{x}}} \right) = -x_{ij} e^{\vec{\beta} \cdot \vec{x}}$$

$$\frac{\partial J}{\partial \beta_j} = \sum_{i=1}^n \frac{x_{ij} e^{\vec{\beta} \cdot \vec{x}}}{(e^{\vec{\beta} \cdot \vec{x}} + 1)^2} \left[ \frac{y_i}{P_i(\vec{\beta})} - \frac{1}{1 - P_i(\vec{\beta})} \right]$$

Escaneado con CamScanner

5.- El algoritmo ID3 nos ayuda a generar un árbol de decisión, este algoritmo es un algoritmo recursivo, que lo que hace es va generando ramas con base en el feature con el que puedes discriminar más. Y así de manera recursiva (quitando la feature n-1) hasta que se llegue a un caso base. Esto para todas las ramas del árbol.