



Inteligencia Analítica

Conceptos preliminares

Las organizaciones se encuentran en una etapa de transformación con respecto a la cantidad de datos que está generando un mundo que es cada vez más digital. La revolución del BigData brinda oportunidades a futuros profesionales además de importantes ventajas competitivas a las organizaciones. La inteligencia analítica (analytics en Inglés) se eleva más allá de la tradicional inteligencia de negocio. En este campo no solo es conocer al cliente a través de la integración de su información, es ir más allá; encontrar patrones no triviales, modelación matemática sofisticada, predicción. Todo lo anterior basado en la nutrición que los datos proveen.

Inteligencia Analítica

La inteligencia analítica se refiere a la disciplina del conocimiento referente a obtener información y conocimiento no trivial y no evidente de grandes volúmenes de datos. La inteligencia analítica es la intersección de tres disciplinas: Matemáticas Aplicadas, Bases de Datos y Planeación estratégica.

Un componente adicional es el diseño e implementación de nuevos algoritmos de predicción o clasificación.

Pasar de una vista de inteligencia de negocio donde lo más importante es medir, a una vista analítica donde lo relevante es entender las causas, futuros estados y descripciones de los fenómenos que nos rodean no es una tarea fácil. Incluso, se requiere formación específica en estadística, análisis multivariante, investigación de operaciones, entre muchas otras técnicas que fungan como el vehículo de acceso al conocimiento, el combustible de dicho vehículo son los datos que tengamos disponibles.

La Inteligencia analítica requiere de una estructura especial de datos para entrenar los modelos. En el siguiente tema revisaremos los detalles.

Metodología de modelado

Para poder realizar un modelo matemático, se requiere una estructura especial de datos. Antes de adentrarnos en el tema, es necesario saber que existirán dos clasificaciones fundamentales para entrenamiento de modelos:

- Aprendizaje supervisado: El modelo aprende con base en ejemplos “etiquetados”, es decir, cuenta con variable objetivo a predecir/clasificar.
- Aprendizaje no supervisado: El modelo aprende de observaciones “sin etiqueta”, carece de variable objetivo. Su función principal es clasificar.

Una vez entendido lo anterior, expliquemos la metodología fundamental de armado de información para modelado.



El supuesto principal de los modelos predictivos en inteligencia analítica es que el futuro se comportará como el pasado reciente. “Reciente” dependerá del contexto en el que nos encontremos y el negocio a analizar. La información histórica del pasado reciente se utilizará para entrenar un modelo para poder predecir futuros estados del sistema. Por ejemplo, para predecir si en los próximos 6 meses un cliente incumplirá el pago de un crédito, podemos determinar con base en un modelo logístico alimentado con la información de los últimos 12 meses la probabilidad de que dicho cliente incumpla. Esto se ejemplifica en el siguiente diagrama:

Observación												Desempeño					
t-11	t-10	t-9	t-8	t-7	t-6	t-5	t-4	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4	t+5	t+6

La tabla de datos que será presentada al modelo deberá contener la siguiente estructura: sean x_i un conjunto de variables explicativas y sea Y la variable objetivo, se busca encontrar un operador matemático f tal que $Y = f(x_i)$. Dichos operadores serán discutidos en la siguiente sección. Cada renglón de la tabla representa un vector de entrada en un hiperespacio, donde tras la presentación sucesiva de dichos vectores al algoritmo de aprendizaje versus la variable objetivo completarán el entrenamiento del modelo.

Aprendizaje máquina

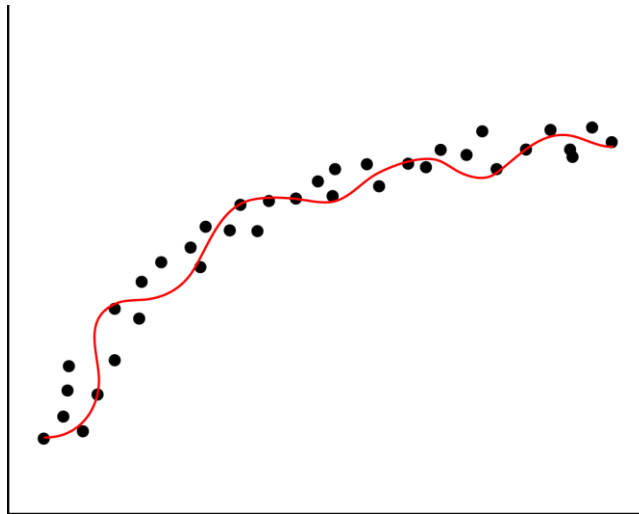
Conocido en inglés como Machine Learning, es un método de análisis de datos que automatiza la construcción de modelos analíticos a través del uso de algoritmos que iterativamente aprenden de los datos. El aprendizaje máquina permite a las computadoras encontrar patrones careciendo de programación explícita que dirija la búsqueda.

El aprendizaje máquina se ha ido transformando con el paso de los años, si bien muchos de los algoritmos llevan un largo tiempo vigentes, la habilidad para aplicar automáticamente cálculos matemáticos complejos a grandes volúmenes de datos una y otra vez cada vez con mayor velocidad es un desarrollo reciente y es considerado como tecnología de punta. Algunas aplicaciones recientes del aprendizaje máquina se enlistan a continuación:

- Los sistemas de conducción automática de autos como los de Tesla y Google
- Recomendación en línea como Amazon, Netflix, etc.
- Análisis de redes sociales. ¿qué dicen los clientes de mi compañía?
- Prevención de fraudes.
- Reconocimiento de imágenes
- Aplicaciones militares

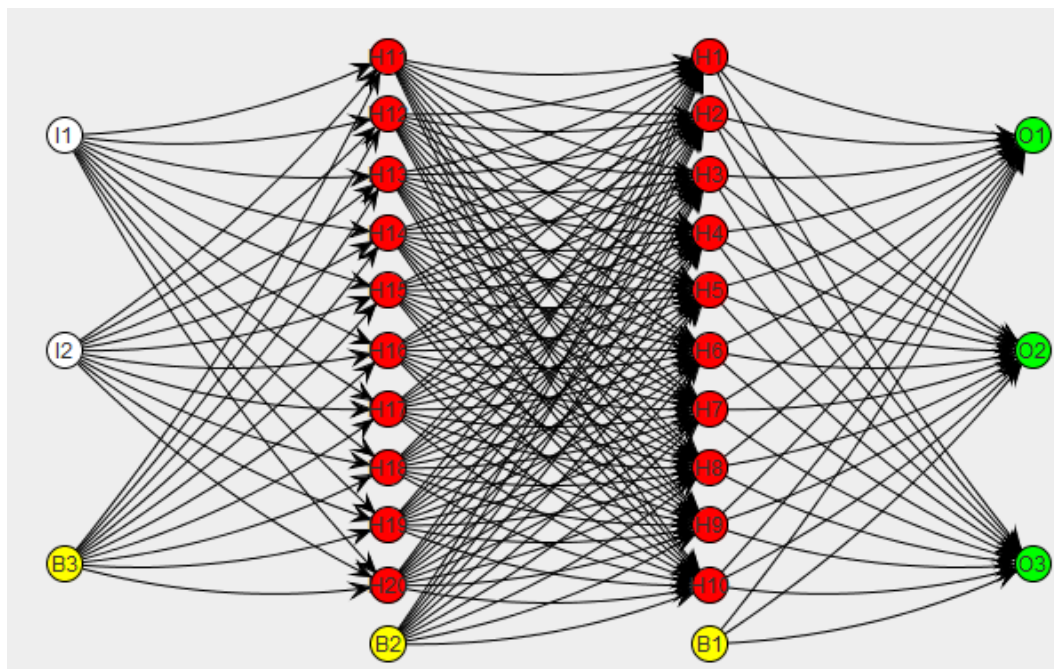
Veamos un esbozo de cada algoritmo que revisaremos:

Modelo que permite encontrar una expresión matemática que relaciona las variables explicativas con la variable objetivo. El entrenamiento de dichos modelos se realiza mediante técnicas clásicas de ajuste de parámetros como mínimos cuadrados o máxima verosimilitud.



Redes neuronales

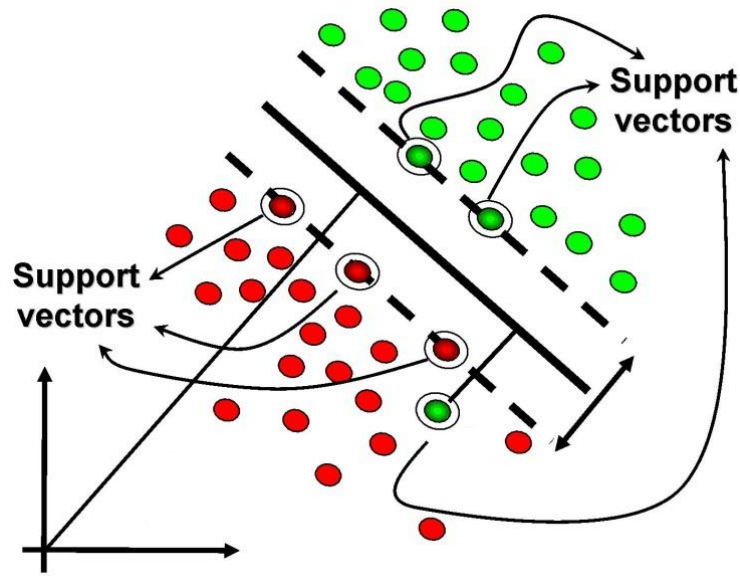
Son un modelo que imita a los sistemas naturales de aprendizaje de los seres vivos. A partir de un sistema natural observable, se construye un modelo con un conjunto de elementos simples de cómputo (neuronas) en un sistema altamente interconectado que tratará de reproducir fenómenos complejos. Pueden ser vistas como una especie de regresión no lineal altamente flexible.





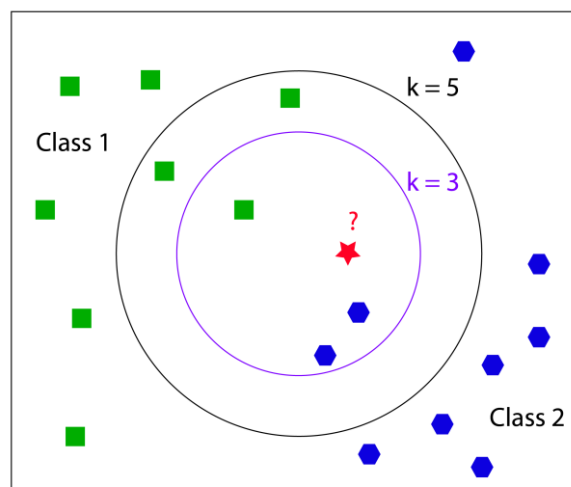
Maquinas vector soporte

Son un clasificador discriminante, formalmente se definen como un hiperplano de separación. Dado un conjunto de entrenamiento, las máquinas vector soporte proporcionarán el hiperplano óptimo que clasifique correctamente nuevas observaciones.



K vecinos más cercanos (K-Nearest Neighbors)

Es un algoritmo no paramétrico, es decir, no hace suposiciones sobre las distribuciones de las variables involucradas. Es un algoritmo perezoso, lo cual quiere decir que no existe entrenamiento explícito o éste es mínima. KNN asume que los datos se encuentran en un "espacio de características" es decir, un espacio métrico donde exista la noción de distancia. La decisión del algoritmo se basará en la similitud de una observación con sus vecinos más cercanos.





Gradiente Descendente Estocástico

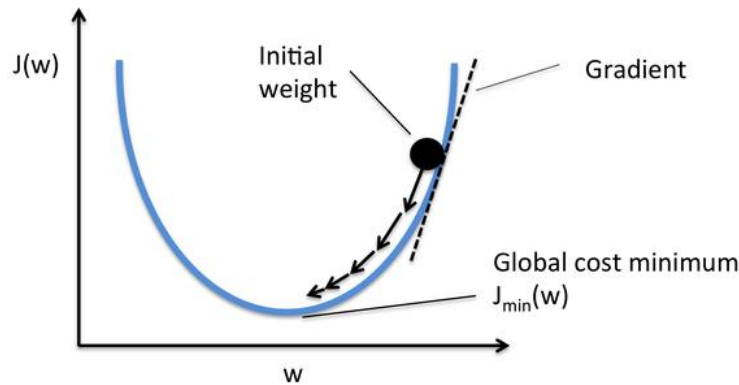
El algoritmo estándar actualiza los parámetros θ de la función objetivo $J(\theta)$ tal que:

$$\theta = \theta - \eta \nabla_{\theta} E[J(\theta)]$$

Donde el valor esperado en la ecuación anterior se aproxima al evaluar el costo del gradiente sobre el conjunto de entrenamiento completo. Este método simplemente hace el cómputo del gradiente de los parámetros usando unas cuantas observaciones de entrenamiento y la actualización estará dada por:

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x^{(i)}, y^{(i)})$$

Para una tupa $x^{(i)}, y^{(i)}$ del conjunto de entrenamiento.



Clasificador ingenuo de Bayes

Es un clasificador basado en el Teorema de Bayes que nos permite conocer la probabilidad de pertenencia de una observación en el conjunto de entrenamiento dadas sus características asumiendo independencia entre las mismas. El clasificador modela la probabilidad de pertenencia a una clase C basado en las predictoras independientes X_i , como sigue:

$$P(C|X_1, X_2, \dots, X_n)$$

Si aplicamos el Teorema de Bayes, tenemos:

$$P(C|X_1, X_2, \dots, X_n) = \frac{P(C)P(X_1, X_2, \dots, X_n|C)}{P(X_1, X_2, \dots, X_n)}$$



De donde sabemos que la probabilidad posterior es proporcional a la verosimilitud presentada en el numerador, en consecuencia:

$$P(C|X_1, X_2, \dots, X_n) \propto P(C) \prod_{i=1}^n P(X_i|C)$$

De donde:

$$\hat{C} = \arg \max_C P(C) \prod_{i=1}^n P(X_i|C)$$

Ensamblados de modelos

Basados en el principio: “Dos cabezas piensan mejor que una”, o como es nuestro caso, dos modelos clasificarán mejor juntos que por separado, los ensambles de modelos combinan las predicciones de un conjunto de algoritmos con la intención de mejorar la generalización/robustez que tendría un estimador individual. Los ensambles se dividen en dos grupos fundamentalmente:

- Ensamblados por promedio (Averaging)
- Ensamblados por impulso (Boosting)

Respectivamente, los primeros basan la estimación en el promedio de las estimaciones individuales, mientras que los segundos son contruidos de forma secuencial buscando reducir el sesgo del estimador combinado, en resumen, se contruyen varios modelos débiles para producir un ensamble poderoso.

Los ejemplos más comunes de ensambles se muestran a continuación:

Bosques Aleatorios

Se contruyen árboles de decisión mediante muestreo con reemplazo del conjunto de entrenamiento

Árboles Extremadamente Aleatorizados

Es similar a los bosques aleatorios con la salvedad de que los umbrales de partición son aleatorios para cada característica y el mejor de ellos es utilizado como regla de partición.

AdaBoost (Adaptative Boosting), Impulso Adaptativo

Fue introducido en 1995 por Freund and Schapire, la base del algoritmo es la construcción de predictores débiles (ligeramente mejores que el azar) en versiones de los datos repetidamente modificadas. Las predicciones de cada uno de ellos es combinada mediante una mayoría ponderada de voto.



Impulso de árboles gradiente

Es una generalización del impulso con funciones diferenciables arbitrarias de pérdida. En cada etapa se construye un predictor débil y se “mejora” mediante la construcción de un nuevo clasificador ajustando los residuales para cualquier función de pérdida.

Mantenimiento de modelos

El ciclo de vida de un modelo no termina una vez que un modelo ha sido entrenado. Cuando se pone en producción, el modelo estará sujeto a pudese descalibrarse o sufrir reducciones en su poder predictivo. Lo anterior se monitorea y controla mediante una serie de reportes que nos proporcionan la información necesaria para determinar que nuestro modelo continúa siendo útil. Listemos dichos reportes:

- Estabilidad de la población: proporciona la distribución de la población vista con nuestro modelo dentro de un periodo de tiempo. A través de pruebas estadísticas de bondad de ajuste se determina cuando la población que alimenta al modelo ha cambiado.
- Desempeño: Contrasta la predicción del modelo con la realidad, dependiendo del modelo la métrica será distinta.
- Potencia de características: Mide variable por variable la potencia estadística que presenten. Cuando una variable deja de ser relevante en la predicción, el modelo deberá recalibrarse.