

Aprendizaje Automático TC3020

Primer examen parcial

Aspectos Teóricos

- 1) Explica con tus propias palabras en qué consiste de forma general la metodología CRISP-DM. Además, explica cada fase implicada de forma detallada.

La metodología CRISP-DM es un modelo de procesos para llevar a cabo proyectos relacionados con ciencia de datos o minería de datos, este tiene su propio ciclo de vida con diferentes fases para obtener un proyecto satisfactorio. Son 6 fases las que tiene el ciclo de vida, el primero es el de 'Business Understanding' en donde se deben hacer todos los requisitos y entender que es lo que se necesita para que el proyecto/negocio salga bien.

El segundo paso es el de 'Data Understanding', aquí es donde se analizan todos los datos que se tienen y se ocupan y a partir de estos se pueden ir generando hipótesis sobre lo que saldrá del proyecto.

El tercer paso es el 'Data Preparation' en donde como lo dice su nombre se preparan los datos para después ser trabajados, se debe hacer limpieza de estos, seleccionar cuales serán útiles y formatear en caso de que se deban (ejemplo: Tengo datos categóricos, cambiarlos para poder trabajar con ellos de forma matemática).

El cuarto paso es el de 'Modeling' aquí es donde se aplican diferentes técnicas y/o algoritmos de modelado de datos para obtener los valores deseados, en esta fase puede llegar el momento en que los datos se tengan que volver a preparar una vez modelados, por lo tanto, se puede regresar a la fase anterior.

El quinto paso es el de 'Evaluation' donde se analizan y evalúan los modelos creados en la fase anterior y poder determinar cuales dan mejores resultados, así como guardar información de estos mismo para proyectos futuros que se parezcan.

El sexto y último paso es el de 'Deployment' donde el modelo final que se escogió en la fase anterior es puesto en producción en el negocio/proyecto para el cual fue hecho, también se analizan y revisan como fueron se trabajó en todas la fases para que en próximas iteraciones se pueda tener una mejora

- 2) Explica con tus propias palabras y detalladamente en qué consiste el aprendizaje supervisado, aprendizaje no supervisado, y el aprendizaje por refuerzo. Haz un análisis comparativo entre estos paradigmas de aprendizaje explicando sus ventajas, desventajas y casos en los que pueden ser ocupados.

Aprendizaje supervisado: En este aprendizaje los algoritmos generan los modelos a partir de una base de datos en donde los datos ya vienen clasificados por una persona y por lo tanto los algoritmos ya saben como realizar los ajustes necesarios para que se obtengan los resultados esperados. En este aprendizaje se dan los algoritmos de clasificación y regresión, para saber cuáles usar con los datos es necesario saber si estos pueden ser clasificados en grupos o si se pueden solamente usar con números reales. La desventaja de esto mismo es que un humano tiene que clasificar bien los datos para que puedan ser usados. Los algoritmos de este aprendizaje pueden ser usados para predecir cosas a futuro como la cantidad de robos que se esperan en un banco o el predecir el costo monetario que tendrán ciertos tipos de accidentes.

Aprendizaje no supervisado: Este tipo de aprendizaje a diferencia del supervisado es que un humano no le debe dar los datos de entrada ya clasificados o modificados, por lo tanto, la salida/respuesta que dan los algoritmos pueden variar a cómo los interpreta la máquina, aquí entra el clustering de datos para obtener los resultados y lo que hace el algoritmo es que divide los datos que tienen parentesco y así poder buscar patrones para dar resultados. La desventaja de estos algoritmos es que se deben modificar de acuerdo al problema a resolverse. Los algoritmos de aprendizaje no supervisado se pueden usar para poder hacer segmentación de mercado y así las empresas saber a qué tipo de clientes mandar cierta información.

Aprendizaje por refuerzo: Este tipo de aprendizaje es usado sobre todo en robots pues lo que hacen es aprender a través de un sistema de "recompensas" en donde se sacan varias formas de lograr un objetivo y si un camino no fue el correcto, el algoritmo aprende de esta equivocación y a la siguiente iteración sabrá que no debe tomar este mismo. Estos algoritmos como ya mencioné se pueden usar en entrenar robots como a carros, para que sepan si van por una calle o si se están yendo en línea recta, este aprendizaje ayuda bastante al manejo automatizado.

- 3) Explica con tus propias palabras cómo funciona el método Batch Gradient Descent. Da todos los detalles posibles.

El primer paso del algoritmo es establecer un learning rate entre los valores de 0 a 1, por lo general entre más bajo sea la predicción final será más exacta. Después de inicializar el vector theta el cual tendrá el mismo tamaño que el vector X a usarse. Lo siguiente es conseguir la convergencia entre los puntos de los vectores con ayuda de la función sigmoide la cual requiere hacer operaciones con el vector theta y x. La convergencia se obtiene hasta que la resta entre Thehta0 y theta1 sea menor al threshold que está establecido por el valor 0.0001. Una vez que se tiene la convergencia el vector de theta tendrá guardados los valores del coeficiente de intercepción que se obtiene durante la iteración de valores.

- 4) Dado $\hat{p}(\sim\beta) = \frac{1}{1+e^{-\beta^T \sim x}}$ $y J(\sim\beta) = -\sum_{i=1}^n [y_i \ln(\hat{p}_i(\sim\beta)) + (1 - y_i) \ln(1 - \hat{p}_i(\sim\beta))]$, calcular $\frac{\partial J(\sim\beta)}{\partial \beta_j}$. Recuerda que $\sim\beta = (\beta_0, \beta_1, \dots, \beta_j, \dots, \beta_{m+1})$, $\sim x = (x_0, x_1, \dots, x_j, \dots, x_{m+1})$ y que $\sim\beta^T \sim x = \beta_0 x_0 + \beta_1 x_1 + \dots + \beta_{m+1} x_{m+1} = \sum_{i=0}^{m+1} \beta_i x_i$.

R-. No supe como hacerlo

- 5) Explica con tus propias palabras cómo funciona el algoritmo ID3 para árboles de decisión.

El algoritmo crea un árbol a partir de una tabla con información de la cual se obtendrán los mejores o mejor atributo dependiendo de las probabilidades que haya que una probabilidad ocurra. Este se va construyendo de arriba a abajo siendo el nodo inicial un aspecto fundamental para que se de la probabilidad a buscar, en las ramas se irán segregando en las probabilidades que hay y de estas mismas se llegan a otros estados, los cuales también pueden/tienen probabilidades de que ocurran a consecuencias de otras probabilidades diferentes a las que tenía el nodo padre. Los árboles quedan equilibrados y no usan backtraking pues no tienen que almacenar información de los nodos anteriores para tomar decisiones.

- 6) Resuelve el siguiente ejercicio empleando la técnica k-NN. Dado el dataset de la tabla 1, clasificar los puntos $\sim a = (3.9, 0.1)$ y $\sim b = (10.1, 5.5)$ usando 3-NN con la métrica de Minkowski (L_p). El parámetro p se da

asignado por el profesor. Mostrar todo el procedimiento en forma manuscrita. Recuerda que $L_p(\sim x, \sim y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p}$.

R-. No supe como hacerlo

Aspectos prácticos

Precisiones finales:

```
Glass Identification Database
Precision Regresion Lineal Scikit-Learn: 0.6046511627906976
Precision Arboles de Decision Scikit-Learn 0.6046511627906976
Precision 50-NearestNeighbors Scikit-Learn: 0.46511627906976744
Johns Hopkins University Ionosphere database
Precision Regresion Lineal Scikit-Learn: 0.8873239436619719
Precision Arboles de Decision Scikit-Learn 0.9295774647887324
Precision 50-NearestNeighbors Scikit-Learn: 0.6901408450704225
```