**Los 10 Algoritmos esenciales en Machine Learning**

Los **algoritmos de aprendizaje automátic**o se pueden dividir en tres grandes categorías: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje de refuerzo.

El aprendizaje supervisado es útil en los casos en que una propiedad (etiqueta) está disponible para un determinado conjunto de datos (conjunto de formación), pero debe predecirse para otras instancias.

El aprendizaje sin supervisión es útil en los casos en que el desafío consiste en descubrir relaciones implícitas en un conjunto de datos no etiquetado (los elementos no están asignados previamente).

El aprendizaje de refuerzo cae entre estos dos extremos: hay alguna forma de retroalimentación disponible para cada paso o acción predictiva, pero no hay etiqueta precisa o mensaje de error.

**APRENDIZAJE SUPERVISADO**

1. **Árboles de decisión:** Un árbol de decisiones es una herramienta de apoyo a la decisión que utiliza un gráfico o un modelo similar a un árbol de decisiones y sus posibles consecuencias, incluidos los resultados de eventos fortuitos, los costos de recursos y la utilidad. Presentan una apariencia como esta:

Diagram

Description automatically generated

Desde el punto de vista de la toma de decisiones empresariales, un árbol de decisiones es el número mínimo de preguntas sí / no que uno tiene que hacer, para evaluar la probabilidad de tomar una decisión correcta, la mayoría del tiempo. Este método le permite abordar el problema de una manera estructurada y sistemática para llegar a una conclusión lógica.

2. **Naïve Bayes Clasification:** Los claseificadores Naïve Bayes son una familia de simples clasificadores probabilísticos basado en la aplicación de Bayes ‘teorema con fuertes (Naïve) supuestos de independencia entre las características’. La imagen destacada es la ecuación – con P (A | B) es probabilidad posterior, P (B | A) es probabilidad, P (A) es probabilidad previa de clase, y P (B) predictor probabilidad previa.

Graphical user interface

Description automatically generated with medium confidence

3.**Ordinary Least Squares Regression:**Si has estado en contacto con la estadística, probablemente hayas oído hablar de regresión lineal antes. Ordinary Least Squares Regression es un método para realizar la regresión lineal. Se puede pensar en la regresión lineal como la tarea de ajustar una línea recta a través de un conjunto de puntos. Hay varias estrategias posibles para hacer esto, y la estrategia de «mínimos cuadrados ordinarios» va así: puede dibujar una línea y luego, para cada uno de los puntos de datos, medir la distancia vertical entre el punto y la línea y sumarlos; La línea ajustada sería aquella en la que esta suma de distancias sea lo más pequeña posible.

Chart, scatter chart

Description automatically generated

Linear se refiere al tipo de modelo que está utilizando para ajustar los datos, mientras que los mínimos cuadrados se refieren al tipo de métrica de error que está minimizando.

4.**Logistic Regression:**La regresión logística es una poderosa manera estadística de modelar un resultado binomial con una o más variables explicativas. Mide la relación entre la variable dependiente categórica y una o más variables independientes estimando las probabilidades utilizando una función logística, que es la distribución logística acumulativa.

Chart, surface chart

Description automatically generated

5. **Support Vector Machines:**SVM es un algoritmo de clasificación binario. Dado un conjunto de puntos de 2 tipos en el lugar N dimensional, SVM genera un hiperlano (N – 1) dimensional para separar esos puntos en 2 grupos. Digamos que usted tiene algunos puntos de 2 tipos en un papel que son linealmente separables. SVM encontrará una línea recta que separa esos puntos en 2 tipos y situados lo más lejos posible de todos esos puntos.

Chart, scatter chart

Description automatically generated

En términos de escala, algunos de los mayores problemas que se han resuelto utilizando SVMs (con implementaciones adecuadamente modificadas) son publicidad en pantalla, reconocimiento de sitios de empalme humanos, detección de género basada en imágenes, clasificación de imágenes a gran escala.

6.**Métodos Ensemble:**Los métodos Ensemble son algoritmos de aprendizaje que construyen un conjunto de clasificadores y luego clasifican nuevos puntos de datos tomando un voto ponderado de sus predicciones. El método de conjunto original es Bayesian promediando, pero los algoritmos más recientes incluyen error de corrección de salida de codificación.

**UNSUPERVISED LEARNING**

7.**Algoritmos Clustering:**Clustering es la tarea de agrupar un conjunto de objetos tales que los objetos en el mismo grupo (cluster) son más similares entre sí que a los de otros grupos.

8. **Análisis de Componentes Principales**: PCA es un procedimiento estadístico que usa una transformación ortogonal para convertir un conjunto de observaciones de variables posiblemente correlacionadas en un conjunto de valores de variables linealmente no correlacionadas llamadas componentes principales.

Chart, diagram

Description automatically generated

Análisis de componentes principales

Algunas de las aplicaciones de PCA incluyen compresión, simplificación de datos para un aprendizaje más fácil, visualización. Tenga en cuenta que el conocimiento del dominio es muy importante al elegir si seguir adelante con PCA o no. No es adecuado en los casos en que los datos son ruidosos (todos los componentes de PCA tienen una variación bastante alta).

9.**Singular Value Decomposition:**En el álgebra lineal, SVD es una factorización de una matriz compleja real. Para una matriz M \* n dada, existe una descomposición tal que M = UΣV, donde U y V son matrices unitarias y Σ es una matriz diagonal.

Diagram

Description automatically generated

10.**Análisis de Componentes Independientes:**ICA es una técnica estadística para revelar los factores ocultos que subyacen a conjuntos de variables, mediciones o señales aleatorias. ICA define un modelo generativo para los datos multivariad

os observados, que se suele dar como una gran base de datos de muestras. En el modelo, se supone que las variables de datos son mezclas lineales de algunas variables latentes desconocidas, y el sistema de mezcla también es desconocido. Las variables latentes se asumen no gaussianas y mutuamente independientes, y se les llama componentes independientes de los datos observados.

Graphical user interface, chart

Description automatically generated

ICA está relacionado con PCA, pero es una técnica mucho más poderosa que es capaz de encontrar los factores subyacentes de fuentes cuando estos métodos clásicos fallan por completo. Sus aplicaciones incluyen imágenes digitales, bases de datos de documentos, indicadores económicos y mediciones psicométricas.