

Qué es aprendizaje de máquina Tipos de aprendizajes Aprendizaje supervisado Aprendizaje no supervisado

2 Aprendizaje automático práctico

Origen de los datos Términos técnicos Términos básicos

3 Casos de uso actual

Qué es aprendizaje de máquina

Tipos de aprendizajes Aprendizaje supervisado Aprendizaje no supervisado

2 Aprendizaje automático práctico Origen de los datos Términos técnicos Términos básicos

3 Casos de uso actual



Aprendizaje

En general por aprendizaje puede entenderse un proceso por medio del cual se adquieren o modifican conocimientos, habilidades, destrezas, conductas o valores como resultado del estudio, la experiencia, la instrucción, el razonamiento y la observación.



Aspectos claves en un proceso de aprendizaje:

- 1 adquisición de conocimiento;
- 2 inferir nuevo conocimiento a partir del ya adquirido;
- 3 adquisición y perfeccionamiento de habilidades, destrezas, conductas y valores haciendo uso del conocimiento adquirido.



Aprendizaje automático

Definición

El aprendizaje de máquina es un campo de estudio que tiene como objetivo desarrollar técnicas que brinden a las computadoras la capacidad de aprender. Más precisamente, se trata de explorar la construcción y el estudio de algoritmos que en lugar de seguir la instrucciones estáticas de un programa, puedan por medio de la experiencia desarrollar habilidades como generalizar comportamientos, reconocer patrones entre otros.



Un concepto más moderno

Tom Mitchell

Un programa de computador se dice que aprende de la experiencia E con respecto a alguna clase de tareas T y medida de rendimiento P, si su desempeño en tareas en T, medida por P, mejora con la experiencia E.

Ejemplo

Un programa para jugar ajedrez puede aumentar su capacidad para ganar partidas. En este caso:

- Tarea T: jugar al ajedrez;
- Medida del rendimiento P: probabilidad que el programa gane la siguente partida;
- Experiencia de entrenamiento E: jugar partidas de ajedrez.



Más ejemplos

Reconocimiento de manuscritos

- Tarea T: reconocer y clasificar palabras manuscritas;
- Medida del rendimiento P: probabilidad de que el programa reconozca y clasifique bien palabras siguientes;
- Experiencia de entrenamiento E: una base de datos de palabras manuscritas correctamente clasificadas.

Predicción del valor de un inmueble

- Tarea T: estimar el valor de un inmueble;
- Medida del rendimiento P: probabilidad de que el programa prediga bien el valor de inmuebles a futuro;
- Experiencia de entrenamiento E: una base de datos de valores de inmuebles respecto a diversos atributos.



Qué es aprendizaje de máquina Tipos de aprendizajes Aprendizaje supervisado Aprendizaje no supervisado

- 2 Aprendizaje automático práctico Origen de los datos Términos técnicos Términos básicos
- 3 Casos de uso actual



Tipos de aprendizajes

En aprendizaje de máquina, la distinción en el tipo de variable de salida, y del abordaje del tratamiento de los ejemplos, ha llevado a una convención de agrupar los diferentes algoritmos en:

Aprendizaje supervisado

La base de conocimientos del sistema está formada por ejemplos de los cuales se conoce la salida correcta (es decir, ejemplos de los que sabemos su clasificación correcta). Esto genera una función que establece una correspondencia entre las entradas y las salidas deseadas del sistema, con base en estos ejemplos etiquetados a priori.

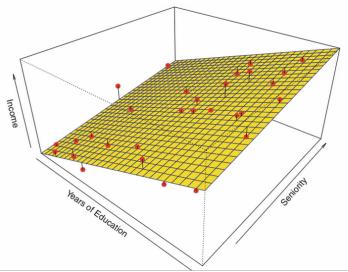


Aprendizaje no supervisado

se distingue del aprendizaje supervisado por el hecho de que no hay un conocimiento a priori. Esto es, la base de conocimientos del sistema está formada por ejemplos de los cuales se conoce sólo las entradas y se desconoce su salida correcta. Así, el aprendizaje no supervisado típicamente trata los objetos de entrada como un conjunto de variables aleatorias, y se busca que el sistema sea capaz de reconocer patrones para poder etiquetar las nuevas entradas.

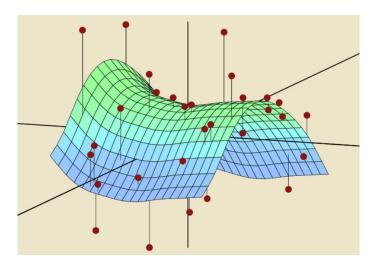


Aproximación de funciones



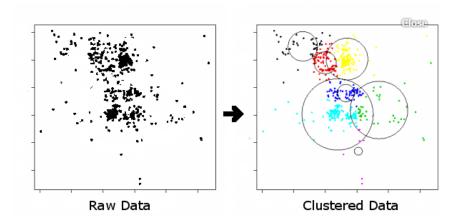


Aproximación de funciones





No supervisado





Qué es aprendizaje de máquina Tipos de aprendizajes

Aprendizaje supervisado

Aprendizaje no supervisado

- Aprendizaje automático práctico Origen de los datos Términos técnicos Términos básicos
- 3 Casos de uso actual



Problemas supervisados

En problemas supervisados la distinción en el tipo de variable de salida ha llevado a una convención de nombre en las tareas.

Regresión

se intenta predecir resultados de valor continuo, en este caso se tiene un tipo de output cuantitativo.

Ejemplos

- 1 predecir el valor de la bolsa a partir del comportamiento de ésta;
- predecir la nota de un alumno en el examen final con base en las notas obtenidas a lo largo curso;
- 3 predecir el valor de un inmueble teniendo en cuenta los datos sobre el tamaño de las casas en el mercado de bienes raíces.



Problemas supervisados

Clasificación

se intenta predecir resultados de valor discreto, en este caso se tiene un tipo de output cualitativo. i.e., se intentan predecir la clasificación de objetos sobre un conjunto de clases prefijadas. Si solo se permiten 2 posibles clases, entonces se llama clasificación binaria; si se permiten más de 2 clases, se dice clasificación multiclase.

Ejemplos

- 1 predecir si un correo electrónico es o no un spam;
- 2 predecir si una transacción en línea es fraude o no lo es.



Qué es aprendizaje de máquina Tipos de aprendizajes Aprendizaje supervisado

Aprendizaje no supervisado

- Aprendizaje automático práctico Origen de los datos Términos técnicos Términos básicos
- 3 Casos de uso actual



Problemas y técnicas no supervisados

El aprendizaje no supervisado es útil tanto para la compresión de datos, como el agrupamiento de los mismos con base a sus características.

Técnicas

- Clustering: agrupan objetos en regiones donde la similitud mutua es elevada;
- Visualización: permiten observar el espacio de instancias en un espacio de menor dimensión;
- 3 Extracción de características: construyen nuevos atributos (pocos) a partir de los atributos originales (muchos).



Qué es aprendizaje de máquina Tipos de aprendizajes Aprendizaje supervisado Aprendizaje no supervisado

2 Aprendizaje automático práctico Origen de los datos Términos técnicos

Términos técnicos

Términos básicos

3 Casos de uso actual



Origen de los datos

Obtención y limiepza del contenido de los datos

- GPS
- Web
- Empresas
- Imagenes
- •

Lectura de los datos

- Excel
- XML
- JSON
- MySQL
- HDF5
- •

La idea es realizar sobre estos datos un aprendizaje automático.



Qué es aprendizaje de máquina Tipos de aprendizajes Aprendizaje supervisado Aprendizaje no supervisado

2 Aprendizaje automático práctico
Origen de los datos

Términos técnicos

Términos básicos

3 Casos de uso actual



Detalles en aplicaciones prácticas

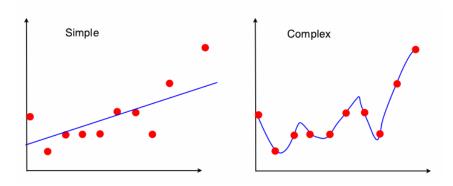
Términos técnicos

Lo más importante es tener clara la pregunta de investigación.

- Detalles matemáticos
- Complejidad computacional
- Cross validation
- Predicción



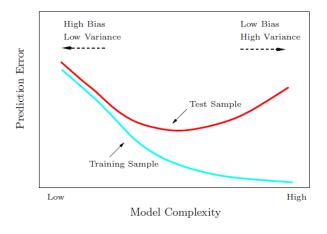
Overfiting Términos técnicos





Bias variance trade off

Términos técnicos





Qué es aprendizaje de máquina Tipos de aprendizajes Aprendizaje supervisado Aprendizaje no supervisado

2 Aprendizaje automático práctico Origen de los datos

Términos técnicos

Términos básicos

3 Casos de uso actual



Términos básicos

En general, POSITIVO:= identificado y NEGATIVO:= rechazado. Entonces

- Verdaderos positivos = identificados correctamente.
- Falsos positivos = identificados incorrectamente.
- Verdadero negativo = rechazado correctamente.
- Falso negativo = rechazado incorrectamente.



Casos actuales

 Ciudades inteligentes: Chicago. 400 servidores guardan datos desde estadísticas de crimenes hasta permisos de construcción. La ciudad usa software libre para recolectar datos no estructurados, como registros en tiempo real de llamados a la línea 911.



Casos actuales

 General electric: La empresa mantiene un enorme 'hub' de datos que incluye la posibilidad de analizar variables para revelar información como cuánta energía gastan sus electrodomésticos y cuánto le costará al comprador utilizarlos.



Casos actuales

 Wal—Mart: Wal-Mart desarrolló Polaris, un motor de búsqueda semántico que estudia algoritmos para entender lo que desea la gente y, de esa manera, impulsar las ventas.
 Sobre ese motor está montada la tecnología que denominamos Social Genome Product (con aplicaciones varias como Shopycat, para analizar los comentarios en Facebook), que repasa miles de tuits, mensajes en Facebook, posts en blogs y videos en YouTube de la marca para detectar las intenciones de compra e impulsar el e-commerce.



Derechos

Las imágenes tomadas son de google imagenes. Y las citas de google.com

