

TOULOUSE LAUTREC

APRENDIZAJE AUTOMATICO CON PYTHON

Tipos de Aprendizaje



Ing. Alexander Valdez

Curso 2290, Clases Lunes y Miercoles 20:00-22:30pm

Primera Clase

1 Aprendizaje Supervisado

Las tareas del aprendizaje supervisado están bien definidas y se pueden aplicar a una gran variedad de situaciones, como identificar spam o predecir precipitaciones.

2 Conceptos basicos de Aprendizaje Supervisado

El aprendizaje automático supervisado se basa en los siguientes conceptos principales:

- Datos
- Modelo
- Entrenamiento
- Evaluación
- Inferencia

Datos

Los datos son la fuerza que impulsa el AA. Los datos se presentan en forma de palabras y números almacenados en tablas, o como valores de píxeles y formas de onda capturados en imágenes y archivos de audio. Almacenamos datos relacionados en conjuntos de datos. Por ejemplo, podríamos tener un conjunto de datos de lo siguiente:

Imágenes de gatos Precios de vivienda Información del clima Los conjuntos de datos se componen de ejemplos individuales que contienen características y una etiqueta. Podrías pensar en un ejemplo como análogo a una sola fila de una hoja de cálculo. Los atributos son los valores que usa un modelo supervisado para predecir la etiqueta. La etiqueta es la “respuesta” o el valor que queremos que el modelo prediga. En un modelo meteorológico en el que se predice la lluvia, los atributos pueden ser latitud, longitud, temperatura, humedad, cobertura de nubes, dirección del viento y presión atmosférica. La etiqueta sería cantidad de lluvia.

Los ejemplos que contienen atributos y una etiqueta se denominan ejemplos etiquetados.

Dos ejemplos etiquetados

Features								Label
date	lat	long	temp	humidity	cloud_coverage	wind_direction	atmp_pressure	rainfall
2021-09-09	49.71N	82.16W	74	20	3	N	18.6	.01
2021-09-09	32.71N	117.16W	82	42	6	SW	29.94	.23

Example

Por el contrario, los ejemplos sin etiqueta contienen atributos, pero no etiquetas. Después de crear un modelo, este predice la etiqueta a partir de los atributos.

Dos ejemplos sin etiqueta

Features							
date	lat	long	temp	humidity	cloud_coverage	wind_direction	atmp_pressure
2021-09-09	49.71N	82.16W	74	20	3	N	18.6
2021-09-09	32.71N	117.16W	82	42	6	SW	29.94

Example

Características del conjunto de datos

Un conjunto de datos se caracteriza por su tamaño y diversidad. El tamaño indica la cantidad de ejemplos. La diversidad indica el rango que abarcan esos ejemplos. Los buenos conjuntos de datos son grandes y muy diversos.

Algunos conjuntos de datos son grandes y diversos. Sin embargo, algunos conjuntos de datos son grandes, pero tienen una baja diversidad, y otros son pequeños, pero muy diversos. En otras palabras, un conjunto de datos grande no garantiza una diversidad suficiente, y un conjunto de datos muy diverso no garantiza ejemplos suficientes.

Por ejemplo, un conjunto de datos puede contener 100 años de datos, pero solo para el mes de julio. Si se usa este conjunto de datos para predecir las lluvias en enero, se producirán predicciones deficientes. Por el contrario, un conjunto de datos puede abarcar solo algunos años, pero contenerlo todos los meses. Este conjunto de datos puede producir predicciones deficientes porque no contiene suficientes años para considerar la variabilidad.

Un conjunto de datos también se puede caracterizar por la cantidad de sus atributos. Por ejemplo, algunos conjuntos de datos meteorológicos pueden contener cientos de características, que van desde imágenes satelitales hasta valores de cobertura de nubes. Otros conjuntos de datos pueden contener solo tres o cuatro características, como la humedad, la presión atmosférica y la temperatura. Los conjuntos de datos con más atributos pueden ayudar a un modelo a descubrir patrones adicionales y realizar mejores predicciones. Sin embargo, los conjuntos de datos con más atributos no siempre producen modelos que realizan mejores predicciones, ya que es posible que algunos atributos no tengan una relación causal con la etiqueta.

MODELO

En el aprendizaje supervisado, un modelo es la compleja colección de números que definen la relación matemática entre patrones de atributos de entrada específicos y valores de etiquetas de salida específicos. El modelo descubre estos patrones a través del entrenamiento.

ENTRENAMIENTO

Antes de que un modelo supervisado pueda hacer predicciones, se debe entrenar. Para entrenar un modelo, le damos un conjunto de datos con ejemplos etiquetados. El objetivo del modelo es encontrar la mejor solución para predecir las etiquetas a partir de los atributos. El modelo encuentra la mejor solución mediante la comparación del valor predicho con el valor real de la etiqueta. En función de la diferencia entre los valores predichos y reales (definidos como la pérdida), el modelo actualiza su solución de forma gradual. En otras palabras, el modelo aprende la relación matemática entre los atributos y la etiqueta para poder hacer las mejores predicciones con los datos no vistos.

Por ejemplo, si el modelo predijo 1.15 inches de lluvia, pero el valor real era .75 inches, el modelo modifica su solución para que su predicción se acerque a .75 inches. Después de que el modelo analizó cada ejemplo en el conjunto de datos (en algunos casos, varias veces), llega a una solución que realiza, en promedio, las mejores predicciones para cada uno de los ejemplos.

A continuación, se muestra cómo entrenar un modelo:

1. El modelo toma un solo ejemplo etiquetado y proporciona una predicción.

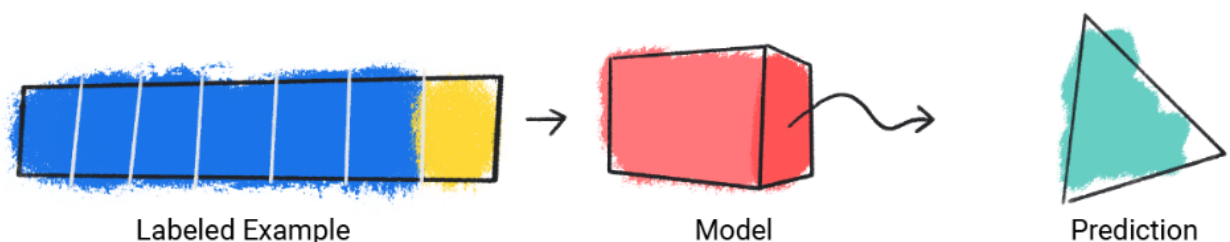


Figura 1: Un modelo de AA que realiza una predicción a partir de un ejemplo etiquetado.

1. El modelo compara el valor predicho con el real y actualiza su solución.

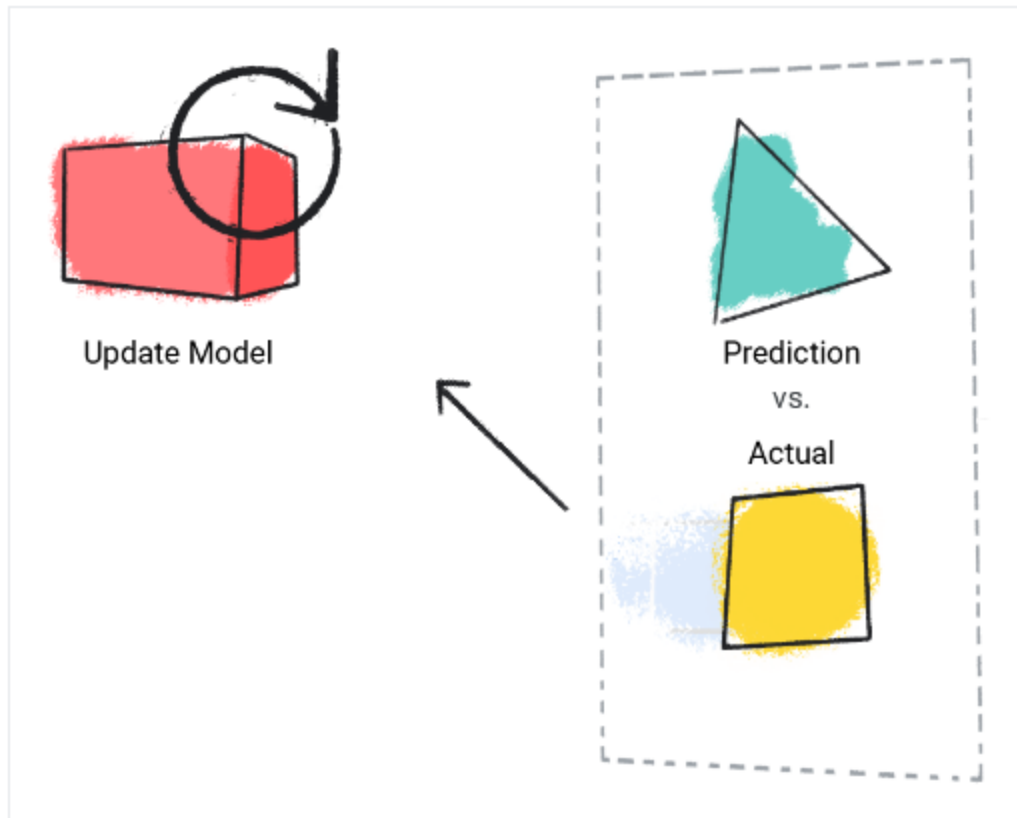


Figura 2. Un modelo de AA que actualiza su valor predicho.

1. El modelo repite este proceso para cada ejemplo etiquetado en el conjunto de datos.

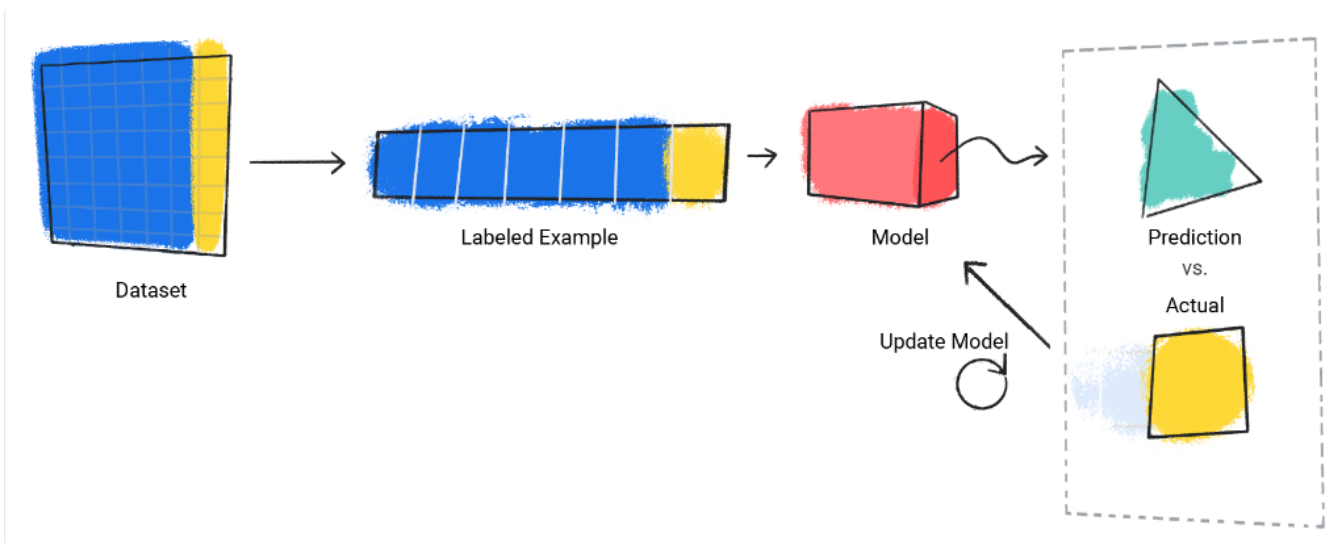


Figura 3. Un modelo de AA que actualiza sus predicciones para cada ejemplo etiquetado en el conjunto de datos de entrenamiento.

De esta manera, el modelo aprende gradualmente la relación correcta entre los atributos y la etiqueta. Esta comprensión gradual también es la razón por la que conjuntos de datos grandes y diversos producen un

mejor modelo. El modelo vio más datos con un rango de valores más amplio y definió mejor su comprensión de la relación entre los atributos y la etiqueta.

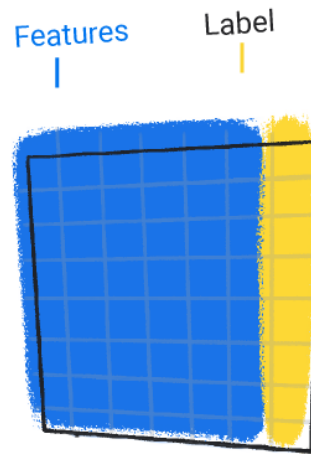
Durante el entrenamiento, los profesionales de AA pueden realizar ajustes sutiles en la configuración y las características que usa el modelo para hacer predicciones. Por ejemplo, ciertos atributos tienen más poder predictivo que otros. Por lo tanto, los profesionales de AA pueden seleccionar qué atributos usa el modelo durante el entrenamiento. Por ejemplo, supongamos que un conjunto de datos meteorológicos contiene `time_of_day` como atributo. En este caso, un profesional de AA puede agregar o quitar `time_of_day` durante el entrenamiento para ver si el modelo hace mejores predicciones con o sin él.

Evaluacion

Evaluamos un modelo entrenado para determinar qué tan bien aprendió. Cuando evaluamos un modelo, usamos un conjunto de datos etiquetado, pero solo le damos al modelo los atributos del conjunto de datos. Luego, comparamos las predicciones del modelo con los valores verdaderos de la etiqueta.

1

Use a **trained ML Model** to predict a value from a given dataset with labeled examples.



2

Compare **Predicted Values** with **Actual Values**.

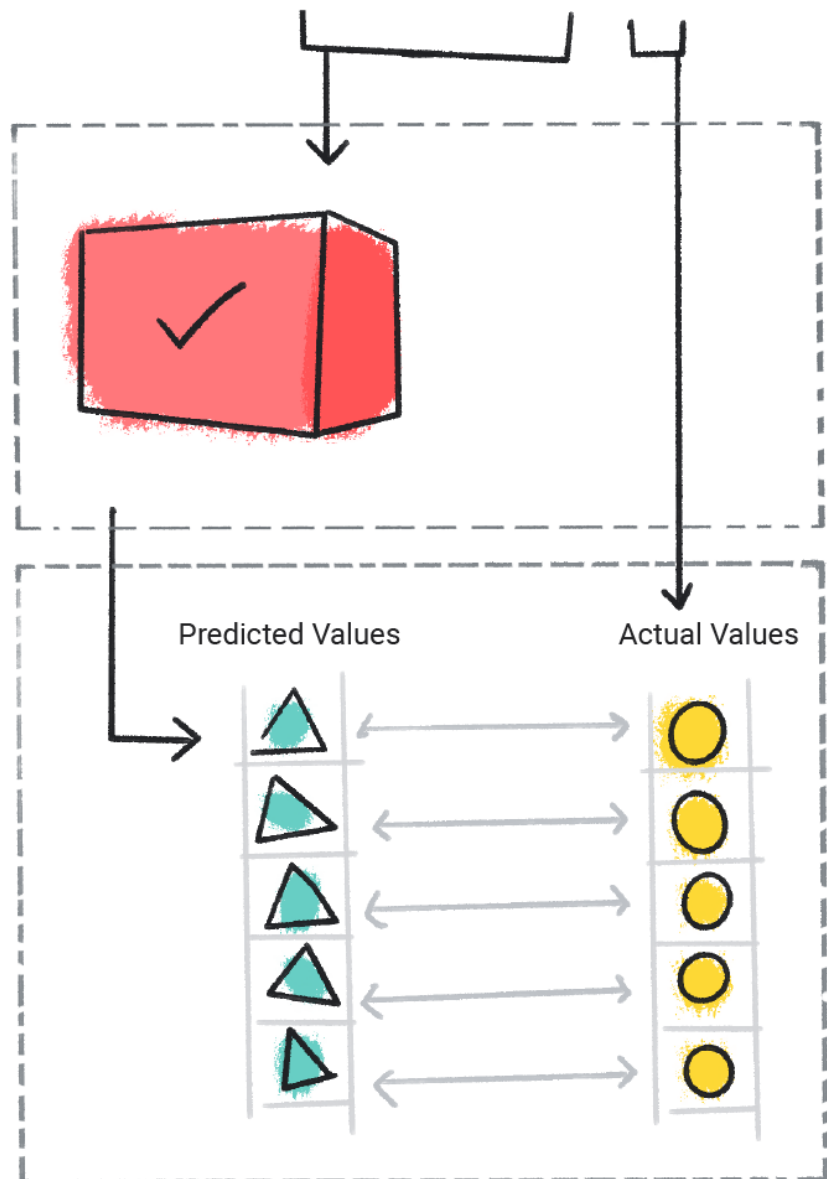


Figura 4. Evaluar un modelo de AA mediante la comparación de sus predicciones con los valores reales

Según las predicciones del modelo, es posible que realicemos más entrenamiento y evaluaciones antes de implementarlo en una aplicación real.

INFERENCIAS

Una vez que estemos satisfechos con los resultados de la evaluación del modelo, podemos usarlo para hacer predicciones, llamadas inferencias, en ejemplos sin etiqueta. En el ejemplo de la app meteorológica, le daríamos al modelo las condiciones climáticas actuales, como la temperatura, la presión atmosférica y la humedad relativa, y predeciría la cantidad de lluvias

In []: