Aprendizaje automático

El aprendizaje profundo pertenece históricamente al campo del aprendizaje automático estadístico, ya que se refiere fundamentalmente a métodos capaces de aprender representaciones a partir de datos. Las técnicas implicadas proceden originalmente de las redes neuronales artificiales, y el calificativo "profundo" subraya que los modelos son largas composiciones de mapeos, ahora conocidas para lograr un mayor rendimiento.

La modularidad, versatilidad y escalabilidad de los modelos profundos han dado lugar a una plétora de métodos matemáticos específicos y herramientas de desarrollo de software, estableciendo el aprendizaje profundo como un campo técnico distinto.

1.1 Aprender de los datos

El caso de uso más sencillo para un modelo entrenado a partir de datos es cuando se tiene acceso a una señal x, por ejemplo, la imagen de una matrícula, a partir de la cual se quiere predecir una cantidad y, como la cadena de caracteres escritos en la matrícula.

En muchas situaciones del mundo real, en las que x es una señal de alta dimensionalidad captada en un entorno no controlado, es demasiado complicado desarrollar una receta analítica o fórmula matemática que relacione x e y de manera precisa.

Lo que sí se puede hacer es recopilar un gran conjunto de entrenamiento 𝒟 de pares (xn,yn), e idear un modelo paramétrico f. Se trata de un fragmento de código informático que incorpora parámetros entrenables w que modulan su comportamiento y que, con los valores adecuados w∗ sea un buen predictor. "Bueno" significa aquí que si se da una x a este trozo de código, el valor yˆ= f(x;w∗) que calcula es una buena estimación del valor y que habría estado asociado con x en el conjunto de entrenamiento si hubiera estado allí.

Esta noción de bondad suele formalizarse con una pérdida ℒ(w) que es pequeña cuando f(.;w) es buena en 𝒟. Entonces, el entrenamiento del modelo consiste en calcular un valor w∗ que minimice ℒ(w∗).

La mayor parte del contenido de este libro trata sobre la definición de f, que, en escenarios realistas, es una combinación compleja de submódulos predefinidos.

Los parámetros entrenables que componen w suelen denominarse pesos, por analogía con los pesos sinápticos de las redes neuronales biológicas. Además de estos parámetros, los modelos suelen depender de los meta-parámetros, que se establecen de acuerdo con el conocimiento previo del dominio previo del dominio, las mejores prácticas o las limitaciones de recursos. También pueden optimizarsepero con técnicas distintas de las utilizadas para optimizar w.

1.2 Regresión de la función de base

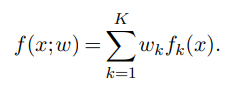
Podemos ilustrar el entrenamiento de un modelo en un caso sencillo en el que xn e yn son dos números reales, la pérdida es el error cuadrático medio:

Imagen que contiene objeto, reloj

Descripción generada automáticamente

*N es el número total de muestras.*

f(-;w) es una combinación lineal de una base predefinida de funciones f1,...,fK, con w = (w1,...,wK):



Como f(xn;w) es lineal con respecto a wks y ℒ(w) es cuadrática con respecto a f(xn;w), la pérdida ℒ(w) es cuadrática con respecto a las wks, y encontrar w∗ que la minimice se reduce a resolver un sistema lineal.

1.3 Under y overfitting

Un elemento clave es la interacción entre la capacidad del modelo, es decir, su flexibilidad y habilidad para ajustarse a datos diversos, y la cantidad y calidad de los datos de entrenamiento. Cuando la capacidad es insuficiente, el modelo no puede ajustarse a los datos, lo que se traduce en un alto error durante el entrenamiento. Esto se denomina underfitting.

Por el contrario, cuando la cantidad de datos es insuficiente, el modelo suele aprender características específicas de los ejemplos de entrenamiento, lo que se traduce en un excelente rendimiento durante el entrenamiento, a costa de un peor ajuste a la estructura global de los datos y un rendimiento deficiente a las nuevas entradas. Esto se denomina overfitting.

Por lo tanto, una gran parte del arte del aprendizaje automático consiste en diseñar modelos que no sean demasiado flexibles pero capaces de ajustarse a los datos. Esto se consigue creando el sesgo inductivo correcto en un modelo, lo que significa que su estructura se corresponde a la estructura subyacente de los datos en cuestión.

Aunque esta perspectiva clásica es relevante para los modelos profundos de tamaño razonable, las cosas se vuelven confusas con aquellos que son muy grandes y tienen un gran número de parámetros entrenables y una capacidad extrema, pero aun así rinden bien en la predicción.

1.4 Categorías de modelos

Podemos organizar el uso de modelos de aprendizaje en tres grandes categorías:

- Regresión: consiste en predecir un vector de valor continuo y ∈ R^K, por ejemplo, la posición geométrica de un objeto, dada una señal de entrada X. Se trata de una generalización multidimensional de la configuración que vimos en § 1.2. El conjunto de entrenamiento se compone de pares de una señal de entrada y un valor de verdad fundamental.

- Clasificación: El objetivo es predecir un valor de un conjunto finito {1,...,C}, por ejemplo, la etiqueta Y de una imagen X . Al igual que en la regresión, el conjunto de entrenamiento se compone de pares de señales de entrada y la cantidad fundamental de verdad, en este caso una etiqueta de ese conjunto.

- Modelado de densidad: tiene como objetivo modelar la función de densidad de probabilidad de los datos µX, por ejemplo, imágenes. En ese caso, el conjunto de entrenamiento se compone de valores xn sin cantidades asociadas para predecir, y el modelo entrenado debería permitir la evaluación de la función de densidad de probabilidad, o el muestreo de la distribución, o ambos.

Tanto la regresión como la clasificación se refieren generalmente al aprendizaje supervisado, ya que el valor a predecir, que se requiere como objetivo durante el entrenamiento, debe ser proporcionado, por ejemplo, por expertos humanos.

Por el contrario, la modelización de densidad generalmente se considera aprendizaje no supervisado, ya que basta con tomar datos existentes sin necesidad de producir una verdad de referencia asociada.

Estas tres categorías no son independientes; por ejemplo, la clasificación puede considerarse como una regresión de puntuaciones de clase, o la modelización de densidad de secuencia discreta como clasificación iterada.

Además, no cubren todos los casos. Uno puede querer predecir cantidades compuestas, o múltiples clases, o modelar una densidad condicional en función de una señal.

­