

Enfoques del aprendizaje

Machine Learning

- Predicción precisa de problemas a gran escala (la generalización es importante)
- La eficiencia del algoritmo es un problema
- Dependiente de los avances en técnicas de optimización y regularización.
- Contraste: el sobreajuste es siempre una posibilidad

Statistical Learning

- Se hacen inferencias utilizando distribuciones de probabilidad
- Buenos resultados solo bajo la hipótesis asumida
- Funciona mal con problemas a gran escala

Minería de datos (estadística y ciencias de la computación)

- Se extraen dependencias entre variables en grandes bases de datos. Inferencia a gran escala
- Comparte muchas herramientas con ML
- Los algoritmos y hardware con alto nivel de escalabilidad son importantes

Bayesian Learning (probabilístico)

- Un enfoque probabilístico completo basado en distribuciones a priori como conocimiento previo
- El sobreajuste no es un problema en general
- Mucho más complejo matemática y computacionalmente
- Muy poca atención a problemas de algoritmos y computacionales

Aprendizaje vs Diseño

El **diseño** se basa en recopilar información sobre el problema que luego pueda ser usada en el aprendizaje. Ej: construir un modelo físico, con el que construimos una distribución de probabilidad que luego usamos para clasificar

En el **aprendizaje** el algoritmo de aprendizaje busca una hipótesis que clasifique bien los datos, para clasificar un nuevo elemento usamos dicha hipótesis.

Definiciones de Machine Learning

- Arthur Samuel: “el campo de estudio que brinda a las computadoras la capacidad de aprender sin estar programadas explícitamente”. Ésta es una definición informal más antigua.

- Tom Mitchell (más formal): “Se dice que un programa de computadora aprende de la experiencia **E** con respecto a alguna clase de tareas **T** y medida de desempeño **P**, si su desempeño en las tareas de **T**, medido por **P**, mejora con la experiencia **E**.”
 - Ejemplo: jugar a las damas
 - E = la experiencia de jugar muchos juegos de damas.
 - T = la tarea de jugar a las damas.
 - P = la probabilidad de que el programa gane el próximo juego.

Paradigmas de Machine Learning

Aprendizaje supervisado

Hay unos datos de muestra y unas etiquetas que clasifican correctamente a cada uno de los datos (aprendizaje estático). Tipos:

- **Regresión:** la salida es un número real (variable continua). P.ej predecir la temperatura a partir de los registros
- **Clasificación:** la salida es una etiqueta (variable discreta). P.ej detectar si un correo electrónico es spam
- **Clasificación probabilística:** La salida es un vector de probabilidad sobre las etiquetas P.ej identificación de objetos en imágenes

Aprendizaje no supervisado

Sólo tenemos los datos de muestra, que se modelan para descubrir relaciones entre ellos. Algunos métodos son:

- Estructura geométrica: **agrupamiento (clustering)**
- Descubrir dependencias: **patrones**
- **Reducción de dimensionalidad:** eliminar características irrelevantes

Aprendizaje reforzado

Hay unos datos de muestra y unas recompensas asociadas a ciertas acciones/soluciones. (aprendizaje dinámico)

Enfoque formal

1. Información disponible

- Datos:

$$\mathcal{P}(\mathcal{D})$$

- Características a utilizar:

$$\mathcal{X} \subset \mathcal{P}(\mathcal{D})$$

- Condición de muestreo: datos de forma idénticamente independiente distribuidos

2. Tarea de predicción:

$$f : X \rightarrow Y$$

(De características a etiquetas)

3. Configuración del modelo (representación)

- Elegir la clase de funciones a utilizar

$$\mathcal{H}$$

- Caracterizar cada elemento

$$h \in \mathcal{H}$$

según los parámetros w

4. Elegir la mejor función candidata

$$h \in \mathcal{H}$$

- Usamos algoritmo **A** para obtenerla
- Optimizamos la función de error para obtener el mínimo error posible y garantizar el aprendizaje. Funciones posibles:
 - ERM: Empirical Risk Minimization
 - SRM: Structural Risk Minimization
 - MDL: Minimum Description Length. Principio de la navaja de Ockham, elegir la explicación más sencilla que explique el conjunto.

Elementos principales de la tarea de aprendizaje

- **Entrada:** vector de características
- **Salida:** clase o etiqueta
- **Función de destino:** desconocida
- **Muestra de datos:**

$$x_i$$

- **Muestra de entrenamiento:** datos etiquetados

Tarea de aprendizaje

- Partimos de \mathbf{X} , \mathbf{Y} , y \mathbf{D} , que vienen dados por la tarea de aprendizaje
- Elegimos una clase de funciones

$$\mathcal{H}$$

(cjo de hipótesis candidatas) que puedan representar a \mathbf{f} , que es la función teórica ideal

- A través de nuestro **algoritmo de aprendizaje**, seleccionamos una función

$$g \in \mathcal{H}$$

, donde esperamos que

$$g \approx f$$

, y la usamos para las nuevas muestras.