Clase 02: Modelo Formal de Aprendizaje

Responsable: Manuel García Garduño

EST-25134, Primavera 2021
Dr. Alfredo Garbuno Iñigo
Enero 19, 2021

1. Marco Formal del Aprendizaje Estadístico

1. Las entradas
   - Conjunto de dominio $\mathcal{X} \subseteq \mathbb{R}^d, d < \infty$.
   - Conjunto de etiquetas $\mathcal{Y}$, por ejemplo, los conjuntos $\{0, 1\}, \{-1, 1\}$.
   - Conjunto de entrenamiento: $\mathcal{S} = \{(x_i, y_i), i = 1, ..., m\}$; $m < \infty$, en donde $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$.
2. La regla de predicción: $h : \mathcal{X} \mapsto \mathcal{Y}$.
3. Un algoritmo de aprendizaje A, en donde $A(\mathcal{S})$ es la hipótesis que el algoritmo de aprendizaje genera al observar el conjunto de entrenamiento.
4. Un modelo que genera los datos
   i) Asumimos que $\mathcal{X}$ tiene una medida de probabilidad $\mathcal{D}$ (distribución) que se desconoce.
   ii) Asumimos que existe una función que etiqueta correctamente los datos, es decir, $\exists\ f : \mathcal{X} \mapsto \mathcal{Y}$ tal que $f(x_i) = y_i$
5. Una métrica de éxito

   Definition 1.1 (Error del clasificador)**.** El error de un clasificador h, es la probabilidad de etiquetar incorrectamente una instancia generada por $\mathcal{D}$.

   - El error de h también puede expresarse como:

   $$\mathcal{L}_{(\mathcal{D},f)}(h) = P\{h(x) \neq f(x))\} = \mathcal{D}(\{x : h(x) \neq f(x)\})$$

   - Conocemos $\mathcal{S}$, pero desconocemos f y $\mathcal{D}$

2. Minimización del Riesgo Empírico

A pesar de no poder calcular el verdadero error de clasificación puesto que ignoramos a la función que etiqueta correctamente a los elementos de $\mathcal{X}$ y tampoco conocemos su distribución, sí podemos construir una medida del error que es calculable con los datos que tenemos.

Definition 2.1 (Riesgo Empírico)**.** Para un subconjunto de m elementos de $\mathcal{X}$, definimos el riesgo empírico de nuestra regla de predicción h como

$$L_f(h) = \frac{|\{i = 1, ..., m : h(x_i) \neq f(x_i)\}|}{m}$$

2.1. ¿Qué podría salir mal?

Supongamos que a partir de nuestro conjunto de entrenamiento $\mathcal{S}$ decidimos definir la siguiente regla de predicción:

$$h(x) = \begin{cases} y_i \text{ si existe} i \text{ tal que} x_i = x \\ 0 \text{ en otro caso} \end{cases}$$

Realmente lo que la regla de prediccón está haciendo es asignar a cada valor de x el valor de y que se observa en el conjunto de entrenamiento, y si el valor de x no se encontraba en el conjunto de entrenamiento original le asigna el valor 0. Observemos que el error empírico de h calculado sobre los datos de entrenamiento es cero, puesto que la forma en que h fue definida nos garantiza que siempre va a etiquetar correctamente a los datos de entrenamiento. Pero, ¿acaso h etiqueta correctamente a las x's que no estaban en el conjunto de entrenamiento? Posiblemente no, porque h siempre les asignará la etiqueta O . Más aún, el verdadero error del clasificador (que es el error que verdaderamente nos importa) será muy grande. Moraleja: minimizar el riesgo empírico no minimiza el error real del clasificador.

2.2. ERM con sesgo inductivo

Para solucionar este problema se escoge, antes de ver los datos, una familia (clase) $\mathcal{H}$ de posibles candidatos para h. De esta forma, quizá no se minimice el error empírico, pero el modelo tendrá una mayor capacidad de generalización que reduzca el error real del clasificador.

Agradecimientos

Este *t*emplate se ha adaptado y traducido del provisto en la clase ACM 204 (Otoño 2017) por el profesor Joel Tropp.