Clase 11: Aprendizaje Estadístico

Responsable: José Pablo Sánchez

EST-25134, Primavera 2021
Dr. Alfredo Garbuno Iñigo
Febrero 23, 2021

1. Clase 11: Aprendizaje no uniforme (NUL)

Vimos que en el modelo PAC establecíamos una relacion entre el tamaño de muestra (m) y los parámetros (ε,δ). Estos parámetros son uniformes con respecto a f y D.

$\Rightarrow$ las clases son limitadas($VCdim(\mathcal{H}) < \infty$

- Ahora buscamos cómo relajar la noción de aprendizaje
- NUL $\longrightarrow$ incorpora una hipótesis ($h \in \mathcal{H}$) contra la que estamos comparando. Esto relaja PAC agnóstico.
- Caracterizar: una unión numerable de posibles clases donde cada elemento es uniforme.
- Esto da lugar al paradigma de minimización de riesgo estructural (SRM)

1.1. Capacidad de aprendizaje no uniforme (NUL)

Definición: Una Hipótesis (h) es (ε, δ)-competitiva con respecto a h' si con probabilidad $\geq 1 - \delta$ se cumple:

$$L_D(h) \leq L_D(h') + \varepsilon$$

Definición: Una clase $\mathcal{H}$ es aprendible no uniformemente (NUL) si existe un algoritmo de aprendizaje, A, y una función $M_H^{NUL} : (0,1)^2 \times H \longrightarrow \mathbb{N}$ tal que $(\varepsilon, \delta) \in (0,1)^2$ y $h \in \mathcal{H}$.
Si $m \geq M_H^{NUL}(\varepsilon, \delta, h)$ entonces $\forall D$ con probabilidad $\geq 1 - \delta$ bajo $S \sim D^m$ tenemos que

$$L_D(A(S)) \leq L_D(h) + \varepsilon$$

1.2. Caracterización de NUL

Teorema 1.1. *Una clase $\mathcal{H}$ de clasificadores binarios es NUL sí y sólo sí es una unión numerable de clases PAC agnósticas.*

Teorema 1.2. *Sea $\mathcal{H}$ una clase tal que $\mathcal{H} = \bigcup_{n \in \mathbb{N}} \mathcal{H}_n$ donde cada $\mathcal{H}_n$ es uniforme. Entonces $\mathcal{H}$ es NUL.*

Ejemplo
Sean $\mathcal{H}_n$ = {clasificadores polinomios de grando = n}, es decir $h \in \mathcal{H}_n$, entonces $h(x) = signo(P_n(x))$
Sea $\mathcal{H} = \bigcup_{n=1} \mathcal{H}_n$ = {todos los polinomios posibles $\in \mathbb{R}$}, luego es fácil ver que $VCdim(\mathcal{H}) = \infty$ y que $VCdim(\mathcal{H}_n) \leq n + 1$

$\therefore \mathcal{H}$ no será PAC agnóstico, pero, por los teoremas anteriores, $\mathcal{H}$ es NUL

1.3. Minimización de Riesgo Estructural (SRM)

Si representamos nuestro espacio de posibles hipótesis $\mathcal{H} = \bigcup_{n\in\mathbb{N}} \mathcal{H}_n$ tendremos que asignar un W_n (peso) para cada $\mathcal{H}_n$

Definición: Sea $\mathcal{H} = \bigcup_{n\in\mathbb{N}} \mathcal{H}_n$ con cada $\mathcal{H}_n$ uniforme con $m_H^{UC}(\varepsilon, \delta)$ y definimos $\varepsilon_n : \mathbb{N} \times (0,1) \longrightarrow (0,1)$ como:

$$\varepsilon_n(m, \delta) = min.\{\varepsilon \in (0,1) : m_H^{UC}(\varepsilon, \delta) \leq m\}$$

Nota: nos fijamos en una cota mínima posible usando n obervaciones.
Si utilizamos la definición de convergencia uniforme y la definición de ε_n tenemos que $\forall(\varepsilon, \delta)$ con probabilidad $\geq 1 - \delta$ bajo $S \sim D^m$ se satisface que

$$\forall h \in \mathcal{H}_n,\ |L_D(h) - L_S(h)| \leq \varepsilon_n(m, \delta)$$

Tomemos ahora W_n tal que $\sum W_n \leq 1$
Si tenemos N posibles candidatos $\mathcal{H}_n$ podríamos considerar cada familia con el mismo peso $W_n = \frac{1}{N}$, más esto no es posible en el caso infinito

Teorema 1.3. *Sea $W : \mathbb{N} \longrightarrow [0,1]$ tal que $\sum_{n=1}^{\infty} W(n) \leq 1$.*
Sea $\mathcal{H} = \bigcup_{n=1}^{\infty} \mathcal{H}_n$ con cada $\mathcal{H}_n$ uniforme con $m_{H_n}^{UC}$.
Sea ε_n como arriba, entonces $\forall\delta \in (0,1)$ y D con probabilidad $\geq 1 - \delta$ sobre $S \sim D^m$ se satisface de manera simultanea, es decir $\forall n \in \mathbb{N}$ y $h \in H_n$, la desigualdad

$$|L_D(h) - L_S(h)| \leq \varepsilon_n(m, W_n\delta)$$

$$\therefore \forall\delta \in (0,1) \text{ y } D \text{ con probabilidad} \geq 1 - \delta \text{ se cumplirá } \forall h \in \mathcal{H}$$

$$L_D(h) \leq L_S(h) + \min \varepsilon_n(m, W_n\delta)$$

Notemos que $n = n(h) = \min\{n : h \in \mathcal{H}_n\}$

Teorema 1.4. *Sea $\mathcal{H} = \bigcup_{n=1}^{\infty} \mathcal{H}_n$ con cada $\mathcal{H}_n$ uniforme con $m_{H_n}^{UC}$.*
Sea $W : \mathbb{N} \longrightarrow [0,1]$ tal que $W(n) = \frac{6}{n^2\pi^2}$.
Entonces $\mathcal{H}$ es NUL usando el SRM con

$$m_H^{NUL}(\varepsilon, \delta, h) \leq m_H^{UC}(\frac{\varepsilon}{2}, W(n)\delta).$$

Resumen:
- Nuestra cota en el error de generalización se basa en evidencia empírica (error de entrenamiento)
- No podemos establecer un tamaño de muestra suficiente, y dependerá del mejor candidato $h \in \mathcal{H} \implies$ la calidad de nuestra respuesta depende de nuestras preferencias.
- Nos Ayudará a seleccionar modelos cuando nuestra información previa es incompleta