
Clase 05: Convergencia Uniforme

Responsable: Emanuel Santiago Payró Costilla

EST-25134, Primavera 2021

Dr. Alfredo Garbuno Iñigo

Enero 26, 2021

- Recordando lo visto anteriormente, dada una \mathcal{H} finita, en el ERM:
 1. Recibimos una muestra $S \sim \mathcal{D}^m$.
 2. Evaluamos el L_S en cada $h \in \mathcal{H}$
 3. $h_S \in \min_{h \in \mathcal{H}} L_S(h)$
- Esperamos que $L_S(h_S)$ sea cercana a $L_{\mathcal{D}}(h_S)$
- Necesitamos que todos los riesgos bajo \mathcal{H} sean buenas aproximaciones.

Definición 0.1. S es una muestra ε -representativa con respecto a Z , \mathcal{D} , l y \mathcal{H} si:

$$\forall h \in \mathcal{H}, \quad |L_S(h) - L_{\mathcal{D}}(h)| \leq \varepsilon$$

Lemma 0.2. Supongamos que S es un conjunto $\frac{\varepsilon}{2}$ -representativo (c.r.a. Z , \mathcal{D} , l y \mathcal{H}). Entonces $h_S \in \min L_S(h)$ satisface:

$$L_{\mathcal{D}}(h_S) \leq \min_{h \in \mathcal{H}} L_{\mathcal{D}}(h) + \varepsilon$$

Definición 0.3. \mathcal{H} tiene la propiedad de convergencia uniforme con respecto a Z y l si:

$\exists m_H^{uc} : (0, 1)^2 \rightarrow \mathbb{N}$ tal que $\forall \varepsilon, \delta \in (0, 1)$ y toda distribución \mathcal{D} sobre Z tenemos que S es una muestra de tamaño m , donde $m \geq m_H^{uc}(\varepsilon, \delta)$ entonces con probabilidad mayor o igual a $1 - \delta$, S es ε -representativo.

Corolario 0.4. Si \mathcal{H} tiene convergencia uniforme con m_H^{uc} , entonces la clase puede aprender en el sentido PAC agnóstico, con una complejidad muestral:

$$m_{\mathcal{H}}(\varepsilon, \delta) \leq m_H^{uc}\left(\frac{\varepsilon}{2}, \delta\right)$$

Ahora nos gustaría establecer condiciones para definir cuando una clase de hipótesis es PAC aprendible de manera agnóstica. Para ello, vamos a:

1. Acotar la probabilidad de hacer un error generalizable por medio de uniones.
2. Utilizar un principio de concentración de medida para garantizar la desigualdad.

Para el primer inciso, dados ε, δ necesitamos encontrar un tamaño de muestra m tal que $\forall \mathcal{D}$ con probabilidad $\geq 1 - \delta$, la elección de la muestra $S \sim \mathcal{D}^m$ garantiza que:

$$\forall h \in \mathcal{H} \quad |L_S(h) - L_{\mathcal{D}}(h)| \leq \varepsilon$$

Entonces si tenemos que: $\mathcal{D}^m(\{S : \forall h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| \leq \varepsilon\}) \geq 1 - \delta$

ó bien $\mathcal{D}^m(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}) < \delta$

Ahora como: $\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\} = \bigcup_{h \in \mathcal{H}} \{S : |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}$
 $\Rightarrow \mathcal{D}^m(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}) \leq \sum_{h \in \mathcal{H}} \mathcal{D}^m(\{S : |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\})$

Lemma 0.5. (Desigualdad de Hoeffding) Si $\theta_1, \dots, \theta_m$ son independientes e idénticamente distribuidas, tal que $\forall i \ \mathbb{E}(\theta_i) = \mu$ y $\mathbb{P}(a \leq \theta \leq b) = 1$, entonces:

$$\mathbb{P}(|\bar{\theta}_m - \mu| > \varepsilon) \leq 2 \exp\left(-\frac{2m\varepsilon^2}{(b-a)^2}\right) \text{ donde } \bar{\theta}_m = \frac{1}{m} \sum_{i=1}^m \theta_i$$

$$\theta_i = l(h, z_i); z_i \stackrel{iid}{\sim} \mathcal{D}; \theta_i \text{ iid.}$$

$$\bar{\theta}_m = L_S(\theta) \quad \mu = L_{\mathcal{D}}(h)$$

$$l(h, \cdot) \in [0, 1] \quad \Rightarrow \quad \theta_i \in [0, 1]$$

$$\mathcal{D}^m(\{S : |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}) = \mathbb{P}(|\bar{\theta}_m - \mu| > \varepsilon) \leq 2 \exp(-2m\varepsilon^2)$$

$$\Rightarrow \mathcal{D}^m(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}) \leq \sum_{h \in \mathcal{H}} 2 \exp(-2m\varepsilon^2)$$

$$= 2|\mathcal{H}| \exp(-2m\varepsilon^2)$$

$$\Rightarrow m \geq \frac{\log(\frac{2|\mathcal{H}|}{\delta})}{2\varepsilon^2}$$

Corolario 0.6. Sea \mathcal{H} una clase finita de hipótesis, \mathcal{Z} el dominio y $l: \mathcal{H} \times \mathcal{Z} \rightarrow [0, 1]$ una función de pérdida. Entonces \mathcal{H} posee la propiedad de convergencia uniforme con complejidad muestral:

$$m_{\mathcal{H}}^{uc}(\varepsilon, \delta) \leq \left\lceil \frac{\log(\frac{2|\mathcal{H}|}{\delta})}{2\varepsilon^2} \right\rceil$$

Además, el algoritmo ERM utilizando una complejidad muestral:

$$m_{\mathcal{H}}(\varepsilon, \delta) \leq m_{\mathcal{H}}^{uc}(\frac{\varepsilon}{2}, \delta)$$

Nos garantiza que \mathcal{H} es aprendible como PAC agnóstico.

Resumen: Si tenemos convergencia uniforme para \mathcal{H} entonces L_S estará cercano a $L_{\mathcal{D}}$

Si CU + ERM \Rightarrow PAC agnóstico

Agradecimientos

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