

Aprendizaje Estadístico III: Consistencia, Minimización del Riesgo Empírico y Cotas

Agosto de 2017

Contenido

- 1 Dimensión VC
- 2 Límites al aprendizaje de máquinas

Cotas: Dimensión VC

- Estimar $\Pi_H(2m)$ es usualmente difícil.
- VC descubrieron lo siguiente: decimos que τ_m aplasta \mathbb{F}_0 si $\Pi_H(m) = 2^m$.
- Definimos la dimensión VC de \mathbb{F}_0 como el mayor m tal que existe una muestra τ_m que la aplasta $\Pi_H(m)$. Si el m máximo no existe, decimos que la dimensión VC es infinita.
- Varios investigadores demostraron que, si la dimensión VC es finita d :

$$\Pi_H(m) \leq \sum_{i=0}^d \binom{m}{i}$$

Cotas: Dimensión VC

- En particular, si $m \geq d$: $\Pi_H(m) \leq \frac{em}{d}$, donde e es el número de Euler. Para ver esto obsérvese que:

$$\begin{aligned} \left(\frac{d}{m}\right)^d \sum_{i=0}^d \binom{m}{i} &\leq \sum_{i=0}^d \left(\frac{d}{m}\right)^d \binom{m}{i} \\ &\leq \sum_{i=0}^m \left(\frac{d}{m}\right)^d \binom{m}{i} = \left(1 + \frac{d}{m}\right)^m \leq e^d \end{aligned}$$

Cotas: Dimensión VC

- en Resumen el anterior resultado muestra que si la dimensión VC es finita, la función de crecimiento crece polinomialmente y por lo tanto la minimización del riesgo es consistente.
- Si es infinita, para todo n existe una muestra que la aplasta es espacio de hipótesis. Y por el teorema de Mendelson el riesgo empírico no es consistente.
- Concluyendo: la máquina de aprendizaje descrita en el teorema de Stone es consistente (universalmente) y la minimización del riesgo empírico es consistente si y solo si, la dimensión VC es finita.

Contenido

- 1 Dimensión VC
- 2 Límites al aprendizaje de máquinas

Límites al aprendizaje de máquinas

Theorem (Arbitrariamente cerca al clasificador aleatorio)

Sea M una máquina de aprendizaje sobre muestras finitas de tamaño n . Entonces existe una distribución de probabilidad P (generador de datos) tal que:

- 1 *El riesgo del clasificador de Bayes es cero bajo P .*
 - 2 *El riesgo esperado de la máquina M es arbitrariamente cerca a 0,5.*
- No hay forma de aprender sin hipótesis sobre la estructura del fenómeno que se quiere estudiar (una teoría!).

Límites al aprendizaje de máquinas

Theorem (Arbitrariamente cerca al clasificador aleatorio)

Sea M una máquina de aprendizaje sobre muestras finitas de tamaño n . Entonces existe una distribución de probabilidad P (generador de datos) tal que:

- 1 *El riesgo del clasificador de Bayes es cero bajo P .*
 - 2 *El riesgo esperado de la máquina M es arbitrariamente cerca a 0,5.*
- No hay forma de aprender sin hipótesis sobre la estructura del fenómeno que se quiere estudiar (una teoría!).