Machine Learning

Training vs testing
Teoría de la generalización

Angel Vázquez-Patiño
angel.vazquezp@ucuenca.edu.ec

Departamento de Ciencias de la Computación
Universidad de Cuenca

4 de noviembre de 2017

Objetivos

1. Entender qué es el número efectivo de hipótesis de un hypothesis set
2. Entender qué es la función de crecimiento
3. Entender la necesidad de acotar la función de crecimiento
4. Entender el significado de la definición de VC dimension
5. Conocer qué es la cota de la generalización VC

Contenido

Training vs testing

- Training set - ejemplo de examen
- La intención es ayudar al estudiante a que le vaya bien en el examen “real”
- ¿Por qué no dar el examen “real”?
- El objetivo no es una buena calificación, sino aprender la materia
- Si fuera el caso, no se podría ver qué tan bien se ha aprendido
- Lo mismo en el enfoque training y test set

Teoría de la generalización

Teoría de la generalización

- Error out-of-sample E_{out} mide qué tan bien el entrenamiento en D ha generalizado los datos que no han sido visto antes
- E_{out} se basa en el rendimiento sobre todo el input space X
- E_{out} se estima con una muestra de datos “frescos” que no hayan sido usados en el entrenamiento (test set)

Teoría de la generalización

- E_{in} se basa en los data points usados para el entrenamiento
- Se tiene el beneficio de conocer la salida (y) de cada x y se ajusta de acuerdo a eso
- Puede no reflejarse el mismo rendimiento en el test set

Teoría de la generalización

Error de generalización

- Diferencia entre E_{in} y E_{out}
- La desigualdad de Hoeffding brinda una forma de caracterizar el error de generalización con una delimitación probabilística

$$\mathbb{P}[|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 2Me^{-2\epsilon^2 N}, \quad \forall \epsilon > 0$$

Teoría de la generalización

Error de generalización

$$\mathbb{P}[|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 2Me^{-2\epsilon^2 N}, \quad \forall \epsilon > 0$$

- Nivel de tolerancia δ, e.g. $\delta = 0.05$
- Denotando el miembro derecho por δ,

$$\delta = 2Me^{-2\epsilon^2 N}$$

, se puede decir con una confianza de 1-δ que

$$|E_{\text{in}}(h) - E_{\text{out}}(h)| \leq \epsilon$$

Teoría de la generalización

Error de generalización

$$|E_{\text{in}}(g) - E_{\text{out}}(g)| \leq \epsilon$$

$$\delta = 2Me^{-2\epsilon^2 N}$$

$$\frac{\delta}{2M} = \frac{1}{e^{2\epsilon^2 N}} \implies e^{2\epsilon^2 N} = \frac{2M}{\delta}$$

$$\epsilon = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

Teoría de la generalización

Error de generalización

$$|E_{\text{in}}(g) - E_{\text{out}}(g)| \leq \epsilon$$

- Error bound

$$\epsilon = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

- Generalization bound (cota de la generalización)

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

Teoría de la generalización

Error de generalización

- El error bound

$$\epsilon = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

depende del tamaño de H (M)

- Casi todos los modelos de aprendizaje tienen un $\mathcal{H}$ infinito (e.g. perceptrón)
- Cómo estudiar la generalización en esos modelos

Teoría de la generalización

Error de generalización

- Lo deseable es reemplazar M con un valor finito para que el límite tenga sentido
- Recordando la forma en que se obtuvo M ...

Probabilidad al rescate

$$\text{“}|E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\text{”} \implies \text{“} \begin{array}{l} \quad |E_{\text{in}}(h_1) - E_{\text{out}}(h_1)| > \epsilon \\ \textbf{or } |E_{\text{in}}(h_2) - E_{\text{out}}(h_2)| > \epsilon \\ \cdots \\ \textbf{or } |E_{\text{in}}(h_M) - E_{\text{out}}(h_M)| > \epsilon \text{”} \end{array}$$

Propiedad deseada:
las hipótesis h_m's son fijas

Probabilidad al rescate

Regla de probabilidad

$$\text{if } \mathcal{B}_1 \Longrightarrow \mathcal{B}_2, \text{ then } \mathbb{P}[\mathcal{B}_1] \leq \mathbb{P}[\mathcal{B}_2]$$

Union bound

$$\mathbb{P}[\mathcal{B}_1 \textbf{ or } \mathcal{B}_2 \textbf{ or } \cdots \textbf{ or } \mathcal{B}_M] \leq \mathbb{P}[\mathcal{B}_1] + \mathbb{P}[\mathcal{B}_2] + \cdots + \mathbb{P}[\mathcal{B}_M]$$

- Usando las dos reglas se tiene que

Probabilidad al rescate

$$\begin{aligned} \mathbb{P}[\,|E_{\text{in}}(g) - E_{\text{out}}(g)| > \epsilon\,] &\leq \mathbb{P}[\quad |E_{\text{in}}(h_1) - E_{\text{out}}(h_1)| > \epsilon \\ &\qquad \textbf{or } |E_{\text{in}}(h_2) - E_{\text{out}}(h_2)| > \epsilon \\ &\qquad \cdots \\ &\qquad \textbf{or } |E_{\text{in}}(h_M) - E_{\text{out}}(h_M)| > \epsilon\,] \\ &\leq \sum_{m=1}^{M} \mathbb{P}[|E_{\text{in}}(h_m) \quad E_{\text{out}}(h_m)| > \epsilon]\,. \end{aligned}$$

$$\mathbb{P}[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq 2Me^{-2\epsilon^2 N}, \quad \forall \epsilon > 0$$

Teoría de la generalización

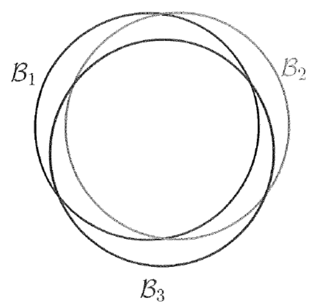
Error de generalización

- Si los eventos $\mathcal{B}_1$, $\mathcal{B}_2$, ..., $\mathcal{B}_M$ están muy traslapados

$$\mathbb{P}[\mathcal{B}_1 \textbf{ or } \mathcal{B}_2 \textbf{ or } \cdots \textbf{ or } \mathcal{B}_M] \leq \mathbb{P}[\mathcal{B}_1] + \mathbb{P}[\mathcal{B}_2] + \cdots + \mathbb{P}[\mathcal{B}_M]$$

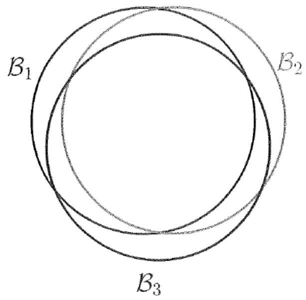
- La probabilidad es altamente sobre estimada

Teoría de la generalización



- El área total de B1, … o BM es más pequeña que la suma de B1, …, BM individuales
- Cierto pero sobre estimado
- En un modelo de aprendizaje común muchas h's son parecidas

Teoría de la generalización



- E.g. en el modelo del perceptrón un cambio pequeño de los pesos w da como resultado infinitas hipótesis que difieren muy poco la una de la otra
- La teoría matemática de la generalización depende de esta observación
- Una vez que se pueda analizar los traslapes de las hipótesis se podría reemplazar M por un número efectivo y establecer una condición más útil bajo la cuál E_{out} está cerca a E_{in}

Número efectivo de hipótesis

Número efectivo de hipótesis

Función de crecimiento

- Cantidad que formaliza el número efectivo de hipótesis
- Reemplaza a M en la acotación de la generalización

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

Número efectivo de hipótesis

Función de crecimiento

- Cantidad combinatoria que captura qué tan diferentes son las hipótesis en $\mathcal{H}$
- Cuánto solapamiento hay en los diferentes eventos de

$$\begin{aligned} \text{``} \quad & |E_{\text{in}}(h_1) - E_{\text{out}}(h_1)| > \epsilon \\ \textbf{or } & |E_{\text{in}}(h_2) - E_{\text{out}}(h_2)| > \epsilon \\ & \dots \\ \textbf{or } & |E_{\text{in}}(h_M) - E_{\text{out}}(h_M)| > \epsilon \text{ ''} \end{aligned}$$

Número efectivo de hipótesis

Definir la función de crecimiento y estudiar sus propiedades básicas

↓

Cómo acotar el valor de la función de crecimiento

↓

Reemplazar M en la acotación de la generalización con la función de crecimiento

Número efectivo de hipótesis

Función de crecimiento

- Función objetivo f binaria
- Cada h $\in \mathcal{H}$ mapea X hacia $\{-1, +1\}$
- La definición de función de crecimiento se basa en el número de diferentes hipótesis que $\mathcal{H}$ puede implementar pero sólo sobre un número finito de muestras y no del $\mathcal{X}$ entero

Número efectivo de hipótesis

Dicotomía

- Si se aplica h $\in \mathcal{H}$ a una muestra finita x1, …, xN $\in \mathcal{X}$, se obtiene una N-tupla h(x1), …, h(xN) de ± 1's
- La N-tupla se llama dicotomía ya que divide x1, ...xN en dos grupos: h(xn) = -1 y h(xn) = +1
- Cada h $\in \mathcal{H}$ genera una dicotomía en x1, ...xN pero dos h's diferentes generan la misma dicotomía si dan el mismo patrón de ± 1's en esa muestra particular

Número efectivo de hipótesis

Dicotomía

Definición

- Sea $\mathbf{x}_1$, ..., $\mathbf{x}_N \in \mathcal{X}$. Las dicotomías generadas por $\mathcal{H}$ en estos puntos son definidos por

$$\mathcal{H}(\mathbf{x}_1, \cdots, \mathbf{x}_N) = \{(h(\mathbf{x}_1), \cdots, h(\mathbf{x}_N)) \,|h \in \mathcal{H}\}$$

- Un $\mathcal{H}(\mathbf{x}_1, ..., \mathbf{x}_N)$ grande significa un $\mathcal{H}$ más diverso que genera más dicotomías en $\mathbf{x}_1$, ..., $\mathbf{x}_N$
- La función de crecimiento está basada en el número de dicotomías

Número efectivo de hipótesis

Función de crecimiento

Definición

- La función de crecimiento para un hypothesis set $\mathcal{H}$ está definida por

$$m_{\mathcal{H}}(N) = \max_{\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathcal{X}} |\mathcal{H}(\mathbf{x}_1, \cdots, \mathbf{x}_N)|$$

- donde |●| denota la cardinalidad de un conjunto

Número efectivo de hipótesis

Función de crecimiento

Definición

$$m_{\mathcal{H}}(N) = \max_{\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathcal{X}} |\mathcal{H}(\mathbf{x}_1, \cdots, \mathbf{x}_N)|$$

- En palabras, $m_{\mathcal{H}}(\mathrm{N})$ es el número máximo de dicotomías que pueden ser generadas por $\mathcal{H}$ en **cualquier** muestra de N puntos
- Para calcular $m_{\mathcal{H}}(\mathrm{N})$ se considera todas las posibles elecciones de N puntos $\mathbf{x}_1$, ..., $\mathbf{x}_N$ de $\mathcal{X}$ y se toma la que da el mayor número de dicotomías

Número efectivo de hipótesis

Función de crecimiento

- Como M, $m_{\mathcal{H}}(\mathrm{N})$ es una medida del número de hipótesis en $\mathcal{H}$, excepto que una hipótesis es ahora considerada en N puntos y no en el $\mathcal{X}$ completo
- Ya que $\mathcal{H}(\mathbf{x}_1, ..., \mathbf{x}_N) \subseteq \{-1,+1\}^{\mathrm{N}}$, el valor de $m_{\mathcal{H}}(\mathrm{N})$ es, a lo sumo, $|\{-1,+1\}^{\mathrm{N}}|$ por lo tanto

$$m_{\mathcal{H}}(N) \leq 2^N$$

Número efectivo de hipótesis

Función de crecimiento

- Si $\mathcal{H}$ es capaz de generar todas las posibles dicotomías en x1,…,xN, entonces $\mathcal{H}$(x1,…, xN)={-1, +1}$^{\mathrm{N}}$ y se dice que $\mathcal{H}$ puede shatter (romper/destrozar/estallar) x1,…,xN
- Significa que $\mathcal{H}$ es tan diverso como puede ser posible en la muestra $\mathbf{x}_1$, ..., $\mathbf{x}_N$ particular

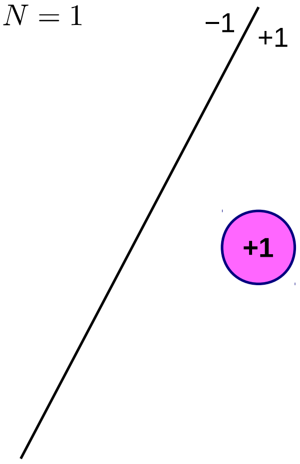
Número efectivo de hipótesis

Función de crecimiento

Ejemplo

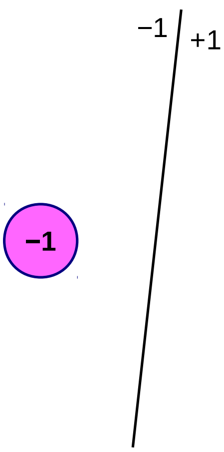
- $\mathcal{X}$, un plano Euclideano
- $\mathcal{H}$, un perceptrón de dos dimensiones

$N = 1$

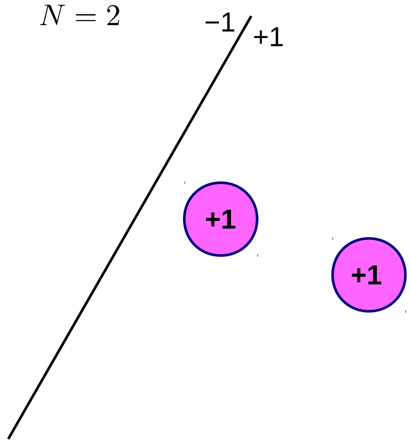


$N = 1$

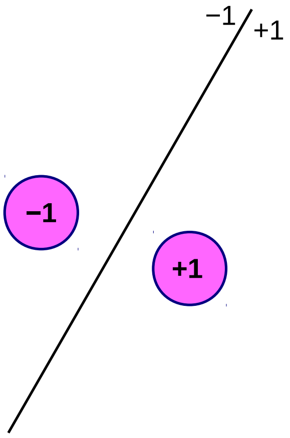
$m_{\mathcal{H}}(N) = 2$



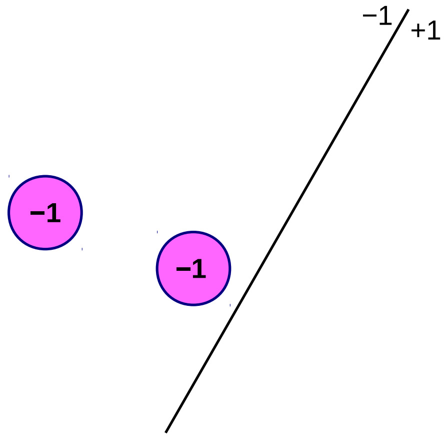
$N = 2$

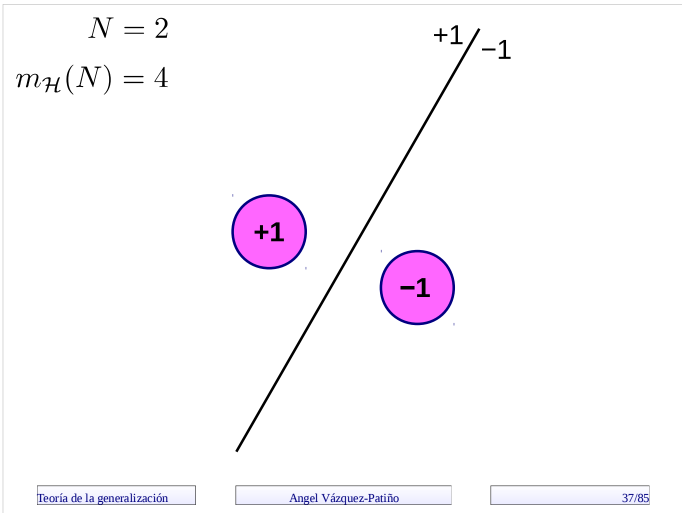


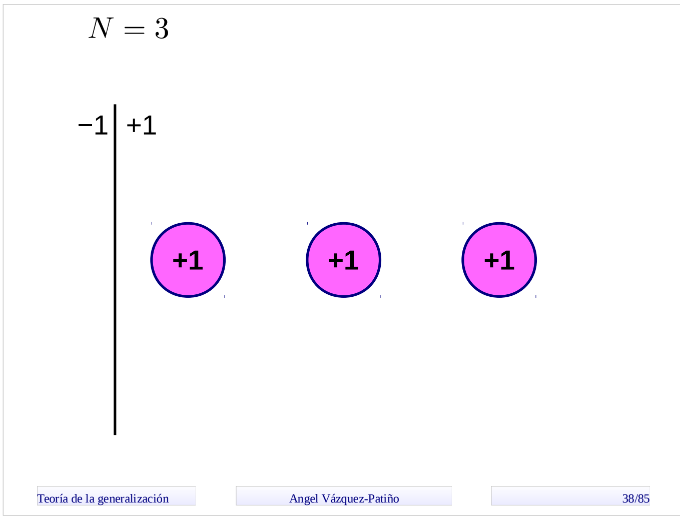
$N = 2$

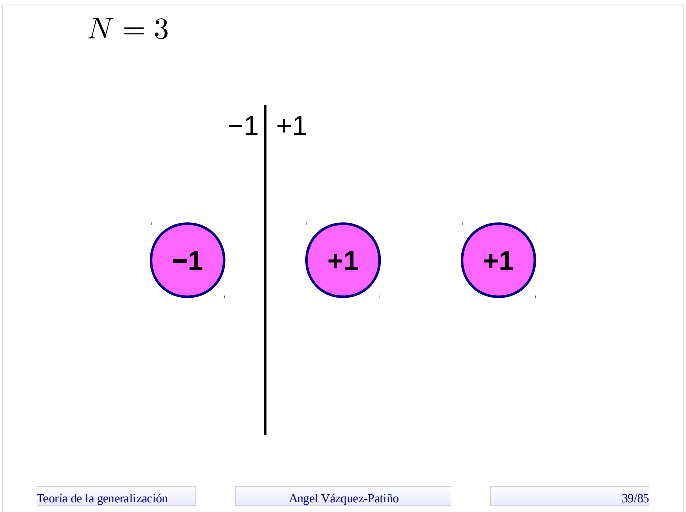


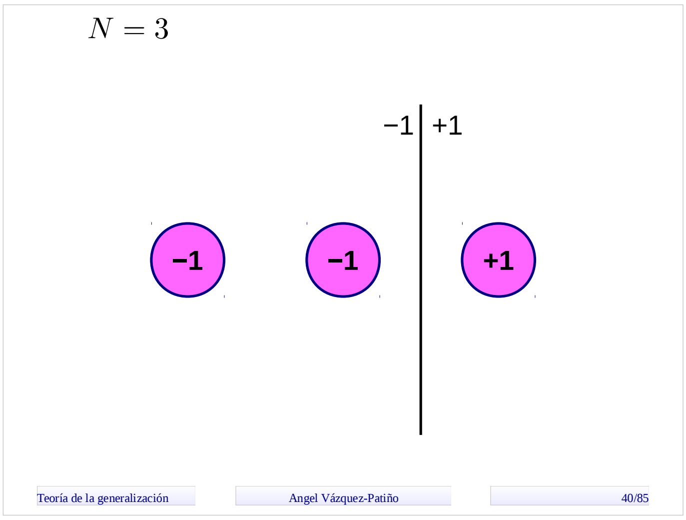
$N = 2$

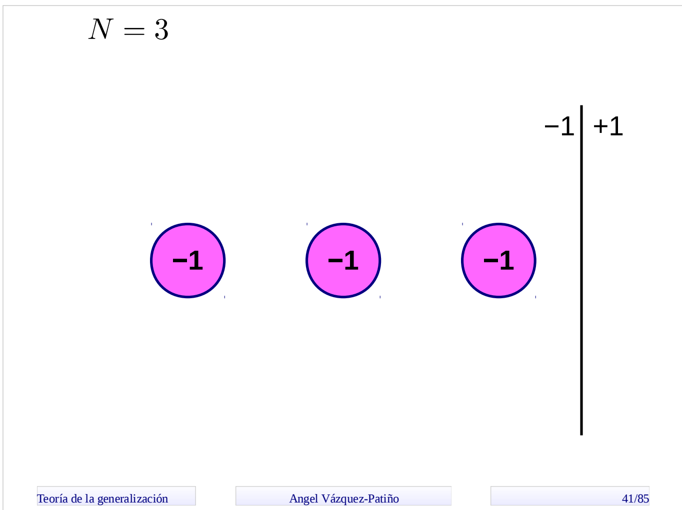


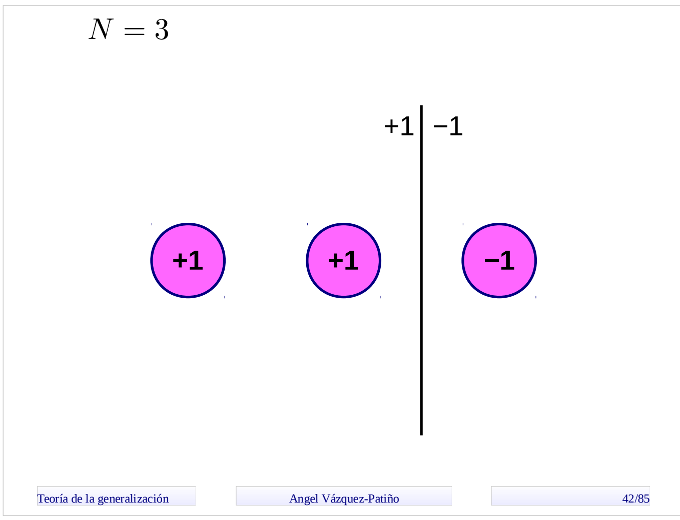

$N = 2$
$m_{\mathcal{H}}(N) = 4$
+1
−1
+1
−1
Teoría de la generalización
Angel Vázquez-Patiño
37/85


$N = 3$
−1
+1
+1
+1
+1
Teoría de la generalización
Angel Vázquez-Patiño
38/85

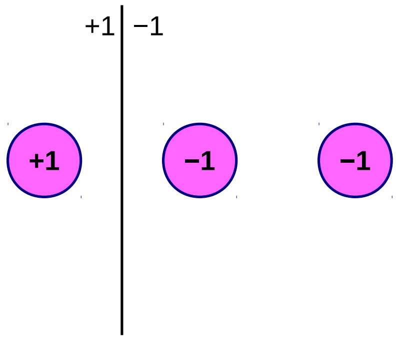

$N = 3$
−1
+1
−1
+1
+1
Teoría de la generalización
Angel Vázquez-Patiño
39/85


$N = 3$
−1
+1
−1
−1
+1
Teoría de la generalización
Angel Vázquez-Patiño
40/85

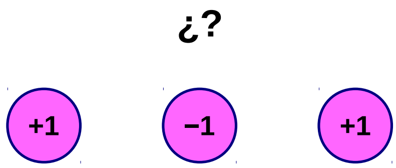

$N = 3$
−1
+1
−1
−1
−1
Teoría de la generalización
Angel Vázquez-Patiño
41/85


$N = 3$
+1
−1
+1
+1
−1
Teoría de la generalización
Angel Vázquez-Patiño
42/85

$N = 3$

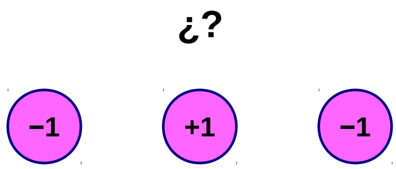


$N = 3$

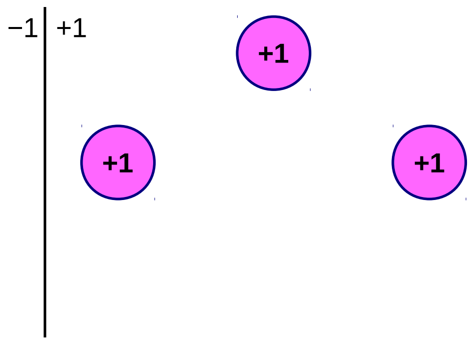


$N = 3$

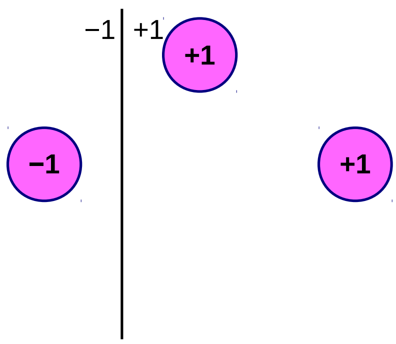
$\mathcal{H}\left(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\right) = 6$



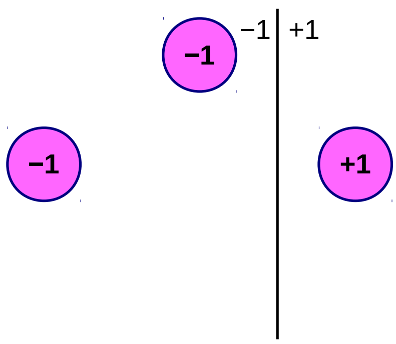
$N = 3$



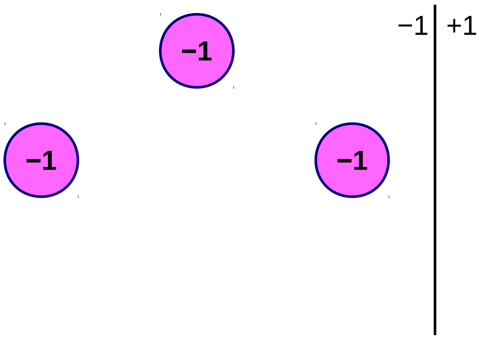
$N = 3$



$N = 3$

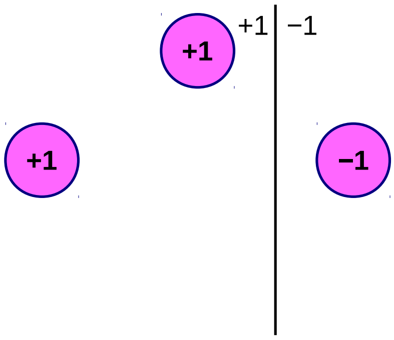


$N = 3$



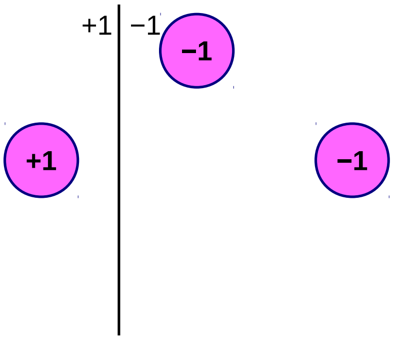
Teoría de la generalización Angel Vázquez-Patiño 49/85

$N = 3$



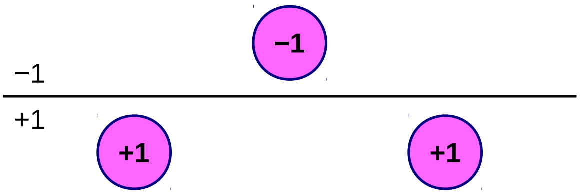
Teoría de la generalización Angel Vázquez-Patiño 50/85

$N = 3$



Teoría de la generalización Angel Vázquez-Patiño 51/85

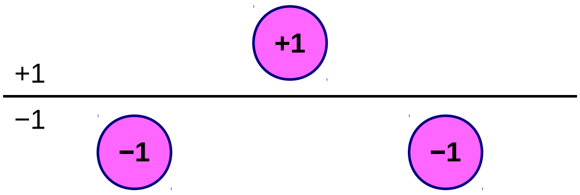
$N = 3$



Teoría de la generalización Angel Vázquez-Patiño 52/85

$N = 3$

$\mathcal{H}\left(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\right) = 8$



Teoría de la generalización Angel Vázquez-Patiño 53/85

$$\mathcal{H}\left(\mathbf{x}_1, \cdots, \mathbf{x}_N\right) = \left\{\left(h(\mathbf{x}_1), \cdots, h(\mathbf{x}_N)\right) | h \in \mathcal{H}\right\}$$

$$m_{\mathcal{H}}(N) = \max_{\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathcal{X}} \left|\mathcal{H}\left(\mathbf{x}_1, \cdots, \mathbf{x}_N\right)\right|$$

$$m_{\mathcal{H}}(3) = 8$$

Teoría de la generalización Angel Vázquez-Patiño 54/85

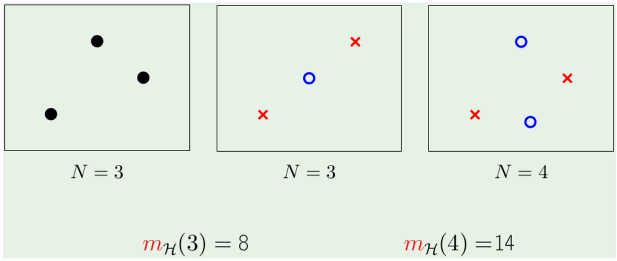
Número efectivo de hipótesis

Función de crecimiento

Ejemplo

- $\mathcal{X}$, un plano Euclideano
- $\mathcal{H}$, un perceptrón de dos dimensiones
- ¿$m_{\mathcal{H}}(3)$ y $m_{\mathcal{H}}(4)$?

Función de crecimiento



$$\mathcal{H}(\mathbf{x}_1, \cdots, \mathbf{x}_N) = \{(h(\mathbf{x}_1), \cdots, h(\mathbf{x}_N)) \,|\, h \in \mathcal{H}\}$$

$$m_{\mathcal{H}}(N) = \max_{\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathcal{X}} |\mathcal{H}(\mathbf{x}_1, \cdots, \mathbf{x}_N)|$$

Número efectivo de hipótesis

Función de crecimiento

Ejemplos

- Encontrar una fórmula para $m_{\mathcal{H}}(N)$
- Rayos positivos
- Intervalos positivos

Número efectivo de hipótesis

Definir la función de crecimiento y estudiar sus propiedades básicas

↓

Cómo acotar el valor de la función de crecimiento

↓

Reemplazar M en la acotación de la generalización con la función de crecimiento

Acotar la función de crecimiento

Acotar la función de crecimiento

- El hecho más importante acerca de las funciones de crecimiento es que si la condición

$$m_{\mathcal{H}}(N) = 2^N$$

se rompe en algún punto, se puede delimitar

$$m_{\mathcal{H}}(N)$$

para todos los valores de N por un polinomio simple basado en este punto de quiebre

Delimitar la función de crecimiento

- El hecho de que el límite es polinomial es crucial
- Si no hay un punto de quiebre

$$m_{\mathcal{H}}(N)$$

- será

$$m_{\mathcal{H}}(N) = 2^N$$

- para todo valor de N

Delimitar la función de crecimiento

- Si se reemplaza M en el límite de la generalización

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

- por $m_{\mathcal{H}}(N)$ en el límite $\sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$
- el error de la generalización no podría llegar a ser cero sin importar cuántos ejemplos de entrenamiento se tengan

Delimitar la función de crecimiento

- Sin embargo, si $m_{\mathcal{H}}(N)$ puede ser limitado por un polinomio, el error de la generalización se acercará a cero según $N \to \infty$
- Va a haber una buena generalización dado un suficiente número de ejemplos

Delimitar la función de crecimiento

Teorema

- Si $m_{\mathcal{H}}(k) < 2^k$ para algún valor de k, entonces

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i} \quad \text{para todo } N$$

- Polinomial en N de grado k-1
- Si $\mathcal{H}$ tiene un punto de quiebre, se tiene lo que se necesita para asegurar una buena generalización, un límite polinomial en $m_{\mathcal{H}}(N)$

Delimitar la función de crecimiento

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

$$m_{\mathcal{H}}(N)$$

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$$

VC dimension

VC dimension

- El teorema delimita la función de crecimiento completa $m_{\mathcal{H}}$ en términos de cualquier punto de quiebre

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln \frac{2m_{\mathcal{H}}(N)}{\delta}}$$

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$$

VC dimension

- Se tiene así una definición de un solo parámetro que caracteriza la función de crecimiento

La Vapnik-Chervonenkis dimension

- La VC dimension de un hypothesis set $\mathcal{H}$, denotado por $d_{\text{vc}}(\mathcal{H})$ o simplemente d_{vc}, es el valor más grande de N para el cual $m_{\mathcal{H}}(N)=2^N$. Si $m_{\mathcal{H}}(N)=2^N$ para todo N, entonces $d_{\text{vc}}(\mathcal{H})=\infty$

VC dimension

- Si d_{vc} es la VC dimension de $\mathcal{H}$, entonces $k = d_{\text{vc}} + 1$ es un punto de quiebre para $m_{\mathcal{H}}$ ya que $m_{\mathcal{H}}(N)$ no puede ser igual a 2^N para cualquier $N > d_{\text{vc}}$ por definición
- No existe un punto de quiebre más pequeño ya que $\mathcal{H}$ puede romper d_{vc} puntos, así puede romper cualquier subgrupo de estos puntos

VC dimension

- Ya que $k = d_{\text{vc}} + 1$ es un punto de quiebre para $m_{\mathcal{H}}$, el teorema puede ser reescrito en términos de la VC dimension

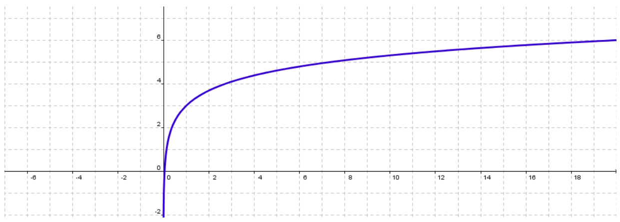
$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{\text{vc}}} \binom{N}{i}$$

- Así, la VC dimension es el orden de la cota polinomial en $m_{\mathcal{H}}(N)$
- Una mayor simplificación

$$m_{\mathcal{H}}(N) \leq N^{d_{\text{vc}}} + 1$$

VC dimension

$$E_{\text{out}}(g) \overset{?}{\leq} E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln \frac{2m_{\mathcal{H}}(N)}{\delta}}$$



VC dimension

Dos tipos de modelos

- Modelos buenos
  - d_{vc} finito
  - N suficientemente grande, E_{in} cerca a E_{out}
  - Rendimiento in-sample generaliza out-sample
- Modelos malos
  - d_{vc} infinito
  - No importa que tan grande sea N, no se puede hacer conclusiones de generalización de E_{in} a E_{out}

VC dimension

El concepto de VC dimension es de suma importancia en Machine Learning

VC dimension

- Para entender mejor el concepto se puede calcular para un modelo de aprendizaje
- Se puede calcular d_{vc} exactamente para el perceptrón

Dos pasos

1) Se muestra que d_{vc} es al menos cierto valor
2) Se muestra que a lo mucho el mismo valor

- Diferencia lógica en argumentar 1) y 2)

VC dimension

$$d_{\text{vc}} \geq N \iff \text{there } \textbf{exists } \mathcal{D} \text{ of size } N \text{ such that } \mathcal{H} \text{ shatters } \mathcal{D}$$

Conclusiones

1) Hay un grupo de N puntos que pueden ser rotos por $\mathcal{H}$: $d_{vc} \geq N$
2) Cualquier grupo de N puntos puede ser roto por $\mathcal{H}$: más que suficiente información para concluir $d_{vc} \geq N$
3) Hay un grupo de N puntos que no pueden ser rotos por $\mathcal{H}$. No se puede concluir nada acerca del valor d_{vc}
4) Ningún grupo de N puntos pueden ser rotos por $\mathcal{H}$. $d_{vc} < N$

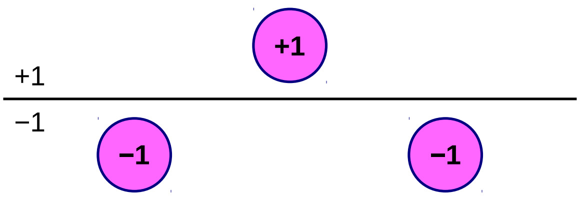
VC dimension

En el perceptrón

- d dimensiones: $d_{vc} = d + 1$

$$N = 3$$

$$\mathcal{H}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = 8$$



VC dimension

En el perceptrón

- Buen caso para intuir qué es d_{vc} ya que $\mathrm{d}+1$ es además el número de parámetros del modelo
- Se podría ver a d_{vc} como una medida de número efectivo de parámetros
- Mientras más parámetros tiene el modelo, más diverso es $\mathcal{H}$, lo cual es reflejado en un valor más grande de la función de crecimiento $m_{\mathcal{H}}(N)$
- En los perceptrones los parámetros efectivos corresponden a w_0, ..., w_d

VC dimension

En el perceptrón

- No en todo modelo es tan obvio los parámetros efectivos (implícitos)
- d_{vc} mide estos parámetros eficaces (grados de libertad) que permiten al modelo expresar un grupo diverso de hipótesis
- La diversidad no necesariamente es buena en el contexto de la generalización
- E.g. el conjunto de todas las posibles hipótesis son tan diversas como se quiera, así $m_{\mathcal{H}}(N)=2^N$ para todo N y $d_{vc}(N) = \infty$. No generalización

Número efectivo de hipótesis

Definir la función de crecimiento y estudiar sus propiedades básicas

↓

Cómo acotar el valor de la función de crecimiento

↓

Reemplazar M en la acotación de la generalización con la función de crecimiento

Cota de la generalización VC

Cota de la generalización VC

$$E_{\text{out}}(g) \overset{?}{\leq} E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln \frac{2m_{\mathcal{H}}(N)}{\delta}}$$

- Para cualquier tolerancia $\delta > 0$

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$$

- La cota de la generalización VC es el resultado matemático más importante en la teoría del aprendizaje
- Establece la viabilidad de aprender con $\mathcal{H}$'s infinitos

Conceptos y términos importantes

Conceptos y términos importantes

Cada detalle

Referencia

- Abu-Mostafa, Y.S., Magdon-Ismail, M., Lin, H.-T., 2012. Learning from data: a short course. AMLbook.com, USA.

Preguntas