

Bonus 1

Luis Suárez Lloréns

1 Factibilidad del aprendizaje

Cuestión. a) ¿Puede S producir una hipótesis que garantice mejor comportamiento que la aleatoria sobre cualquier punto fuera de la muestra?
b) Asumir desde ahora que todos los ejemplos en \mathcal{D} tienen $y_n = +1$. ¿Es posible que la hipótesis que produce C sea mejor que la hipótesis que produce S ?

Respuesta. a) No. Es más, tampoco podrías asegurarlo sobre la muestra, salvo que toda la muestra tuviera la misma evaluación. Veámoslo en más detalle.

Supongamos que la estrategia S elige clasificar como $+1$. Entonces, si hay un punto $x \in \mathbb{R}$ tal que $f(x) = -1$, nuestra estimación fallaría, mientras que elegir al azar podría acertar. Luego para ese punto, nuestra estimación es peor que la aleatoria.

Por muchos datos que tengamos para aprender, nunca podríamos asegurar que no existe dicho punto x , como mucho podríamos decir que es más o menos probable que exista.

b) Sí. Volvemos a estar en una condición parecida al apartado anterior. Por muchas muestras que tengamos, no podemos decir nada con seguridad sobre lo que pasa fuera de la muestra.

Lo máximo que podemos decir, dado estos ejemplos, es que es más probable que la hipótesis producida por S sea mejor que la producida por C , pero existe la posibilidad que la mejor de las dos sea C .

2 Error y ruido

Cuestión. Tenemos una hipótesis h de error μ para una función determinística f . Si usamos la misma h para una versión ruidosa:

- a) ¿Cuál es la probabilidad de error que comete h al aproximar y ?
- b) Para que valor de λ será h independiente de μ .

Respuesta. a) El enunciado nos dice que la posibilidad de error de nuestra aproximación h de la función f es μ . Sabemos también que la versión ruidosa que vamos a aproximar es:

$$P(y|x) = \begin{cases} \lambda & y = f(x) \\ 1 - \lambda & y \neq f(x) \end{cases}$$

Entonces, vamos a estudiar dos casos.

Por un lado, si nuestra función h consigue predecir el verdadero valor de f — la probabilidad de este suceso es $1 - \mu$ —, existen dos opciones, que la versión con ruido modifique el valor o no. Si modifica el valor —probabilidad $1 - \lambda$ —, nuestra h falla. Si no lo modifica acierta.

Por otro lado, si nuestra función h falla —probabilidad μ — y la versión con ruido no modifica el valor —probabilidad λ —, falla. Si modifica el valor, acierta.

Del análisis anterior, podemos obtener el error del nuevo sistema, sumando las posibilidades de error analizadas. Es decir:

$$(1 - \mu)(1 - \lambda) + \mu\lambda$$

- b) Si expandimos la expresión anterior obtenemos:

$$1 - \lambda - \mu + 2\mu\lambda = error$$

Queremos que μ desaparezca de la ecuación, es decir:

$$-\mu + 2\mu\lambda = 0; \quad \mu = 2\lambda\mu$$

De donde finalmente obtenemos:

$$1 = 2\lambda; \quad \lambda = \frac{1}{2}$$

3 Error de generalización

Cuestión. Para funciones objetivo binarias, mostrar que $P[h(x) \neq f(x)]$ puede escribirse como un valor esperado de una medida de error cuadrático medio en los siguientes casos:

- a) Cuando usamos $\{0,1\}$ como los valores de la función binaria.
- b) Cuando usamos $\{-1,+1\}$ como los valores de la función binaria.

Respuesta. Vamos a estudiar primero el caso en el que el dominio X sea discreto, por ser más intuitivo, y después obtendremos la forma general.

En el caso discreto, sabemos que la probabilidad de error, $P[h(x) \neq f(x)]$, es:

$$P[h(x) \neq f(x)] = \frac{n^{\circ errores}}{n^{\circ total}}$$

Ahora miremos al error cuadrático, $(f(x) - h(x))^2$. En el primer apartado del ejercicio, codificando como $[0,1]$, obtenemos que en caso de acierto el error es 0, y en caso de fallo el error es 1.

Entonces, al hacer el error cuadrático medio, obtenemos:

$$\frac{\sum_1^n (f(x) - h(x))^2}{n} = \frac{n^{\circ errores}}{n}$$

Es decir, la probabilidad de error, justo lo que buscábamos. En el caso de clasificar con $[-1,1]$, al calcular el error cuadrático de un fallo, obtenemos $4 - (1 - (-1))^2 = 4$. Por tanto, para obtener el resultado, tenemos que dividir por 4.

Para pasar al caso infinito, no podemos usar la sumatoria. Lo que tenemos que realizar es la esperanza matemática. Por tanto, las dos opciones quedan así:

a)

$$[0, 1] \longrightarrow E_X ((f(x) - h(x))^2)$$

b)

$$[-1, 1] \longrightarrow \frac{1}{4} E_X ((f(x) - h(x))^2)$$

4 Función de crecimiento y punto de ruptura

Cuestión. Calcular la función de crecimiento del modelo de aprendizaje construido por dos círculos concéntricos en \mathbb{R}^2 .

Respuesta. El enunciado completo del ejercicio, nos indica que la familia de funciones \mathcal{H} clasifican como +1 a los puntos $a \leq x_1^2 + x_2^2 \leq b$, con $a, b \in \mathbb{R}$, y -1 cualquier otro.

En realidad, la fórmula $x_1^2 + x_2^2$ es el cuadrado de la distancia al origen, que llamaremos d^2 . Lo más importante, es una función $d^2 : \mathbb{R}^2 \rightarrow \mathbb{R}$. Por tanto, cuando utilizamos la familia \mathcal{H} , lo primero que se hace es llevar los puntos en 2 dimensiones a una recta.

Sabiendo esto, es claro que este modelo de aprendizaje tiene la misma función de crecimiento que el modelo de los intervalos en \mathbb{R} visto en clase. El único problema que se podría encontrar a esto es que el modelo de los dos círculos concéntricos no nos va a dar nunca un valor negativo al calcular la distancia, y el modelo de los intervalos sí acepta valores negativos. Esto no es un problema, porque para conocer la función de crecimiento del modelo de los intervalos solo utilizamos que los N puntos sean distintos, luego podemos añadir si queremos que sean también positivos.

Por tanto, la función de crecimiento, como vimos en teoría, es:

$$m_{\mathcal{H}} = \binom{N+1}{2} + 1$$

De todas maneras, vamos a calcular este número. Las elecciones que tenemos en el modelo son a y b , y tenemos $N+1$ intervalos —entre los puntos— donde,elijamos el punto que elijamos, la clasificación va a ser la misma. Luego el número de elecciones que podemos hacer, sin repetición, es $\binom{N+1}{2}$. El +1 se añade para el caso donde todo se clasifique a -1, que se da cuando tomamos a y b en el mismo intervalo.