Tarea 6

Eduardo Acuña Yeomans

7 de abril de 2015

1. Error de generalización

Si imponemos un error máximo de generalización $\epsilon = 0{,}05$ y queremos que la cota superior de probabilidad que la generalización sea peor que ϵ sea a lo más $0{,}03$, ¿Cuál es el número de ejemplos N que se requieren en el caso que

1. $M = 1$

   a) 840 ejemplos

2. $M = 10$

   a) 1301 ejemplos

3. $M = 100$

   a) 1761 ejemplos

2. Punto de quiebre

Para un clasificador binario lineal en tres dimensiones, ¿Cuál es el (menor) punto de quiebre?

$$h(x) = sign(w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3)$$

La respuesta a la pregunta es 5, la manera en la que argumento dicha respuesta carece de rigor matemático y se basa mas que nada en ideas e intuiciones sobre los clasificadores lineales.

Para determinar que el punto de quiebre es 5 primero demuestro que hay un conjunto de 4 puntos en $\mathbb{R}^3$que pueden ser clasificados con un perceptrón para cualquiera de las posibles 2^4 asignaciones. Pudiera dar un ejemplo en concreto e incluso mostrar las diferentes separaciones lineales para cada asignación de ese ejemplo, sin embargo voy a construír un argumento que considero lo suficientemente bueno para este ejercicio.

Primero, consideramos el hecho que en dos dimensiones, tres puntos diferentes no colineales son linealmente separables para cualquiera de las 2^3asignaciones. Después consideramos el caso de tres puntos diferentes no colineales en tres dimensiones: $p_1 = (1, 0, 0)$, $p_2 = (0, 1, 0)$, y $p_3 = (0, 0, 1)$, estos tres puntos yacen en un plano P en el cual podemos trazar rectas que separan cualquier asignación dada. Ahora consideramos un cuarto punto p_4 fuera de P. Para cada una de las 2^3asignaciones de p_1, p_2 y p_3 hay dos posibles asignaciones para p_4 y al menos una línea recta R_i sobre P que separa a los primeros tres puntos para la i-ésima asignación. Para separar la asignación i considerando p_4 tenemos que elegir otro plano P'_i tal que $P \cap P'_i = R_i$ que clasifique a p_4 de manera correcta, ya que p_4 es diferente al resto de los puntos y no es colineal con ni uno de ellos existe un plano $P{*}_i$ que contiene a R_i y a p_4. Para conseguir el P'_i deseado, solo tenemos que rotar $P{*}_i$ con R_i como un eje fijo, de tal manera que p_4 quede de un lado o de otro de la clasificación.

Para determinar que no se pueden colocar 5 puntos en $\mathbb{R}^3$ tal que 2^5 asignaciones pueden ser separables por planos (es decir: el punto de quiebre para h es 5), realizamos un procedimiento similar al primer paso pero ahora considerando la colocación de p_4 y p_5 fuera de P. Se realizan todas las posibles asignaciones para los primeros tres puntos y se aborda para cada asignación de los nuevos dos puntos un plano P_i'' tal que $P \cap P_i'' = R_i$ que clasifique a p_4 y p_5 de manera correcta. Podemos determinar que no existe P_i'' para cada $i \in \{1, \ldots, 2^5\}$ ya que la rotación del plano se realiza considerando los primeros tres datos ya clasificados, los puntos de una partición del espacio serán de una clase k y la otra partición serán de la otra clase k^c; esto establece una dependencia entre la clasificación de p_4 y la clasificación de p_5, de tal manera que para ni un acomodo de los puntos podremos realizar las 2^5 asignaciones.

Una manera mas formal de abordar este punto sería demostrar el teorema de Radon: "Si se toman $d+2$ puntos en $\mathbb{R}^d$, se pueden repartir en dos conjuntos disjuntos A y B tales que las envolventes convexas de A y B se intersectan". En este caso, los puntos en conjunto A serían los clasificados con la clase $+1$ y los puntos en el conjunto B serían los clasificados con la clase -1. Ya que las envolventes convexas se componen de rectas entre los puntos de un conjunto, hablar de que podemos encontrar una manera de repartir los puntos en A y B tal que sus envolventes se intersectan es análogo a decir que hay una asignación de clases de los $d+1$ puntos que no pueden ser linealmente separables por el perceptrón.

3. Función de crecimiento

Se plantea un algoritmo de parendizaje llamado 2-intervalos, el cual es un conjunto de hipótesis dela forma $h : \mathbb{R} \to \{-1, +1\}$ de manera que $h(x) = +1$ si el valor de x se encuentra dentro de alguno de dos intervalos establecidos a partir de cuatro parámetros que definen a h. En caso contrario $h(x = -1$.

1. ¿Cuál es el (menor) punto de quiebre del conjunto de hipótesis?

   a) Para 5 puntos existe una asignación que no puede encontrar el algoritmo: 1,0,1,0,1. Ya que es la menor cantidad de puntos en donde es necesario tener 3 intervalos .

2. ¿Cuál sería la función de crecimiento de $m_{\mathcal{H}}(N)$ para $\mathcal{H}$ el conjunto de todas las hipótesis posibles del método de 2-intervalos?

   a) El algorítmo de 2-intervalos es mas poderoso que el 1-intervalo, además va a poder clasificar otras asignaciones. La cantidad de nuevas asignaciones admitidas es C_4^{N+1}, por lo tanto $m_{\{\mathcal{H}\}}(N) = C_4^{N+1} + C_2^{N+1} + 1 = \frac{N^4}{24} - \frac{N^3}{12} + \frac{11N^2}{24} + \frac{7N}{12} + 1$ (para $N \geq 3$), en otro caso es

3. ¿Cuál sería el (menor) punto de quiebre k para el método de M-intervalos?

   a) $2 \times N + 1$

4. El método de círculos concéntricos

Se considera un método de aprendizaje donde $\mathcal{X} = [-1, 1] \times [-1, 1]$, de manera que $x = (x_1, x_2)$ es un punto en el plano, donde el valor $(0, 0)$ es el centro. Sea $\mathcal{H}$ el conjunto de todas las funciones tales que $h(x) = 1$ cuando $a^2 \leq x_1^2 + x_2^2 \leq b^2$ y $h(x) = -1$ en otro caso. Cada hipótesis h es caracterizada por dos parámetros a y b. Calcule la función de crecimiento $\mathcal{H}$.

Este ejercicio es en donde tengo mas dudas, la solución propuesta se basa en la idea tomar una función $f : \mathbb{R}^2 \to \mathbb{R}$ que transformará cada punto en un valor real, donde $f(x) = \sqrt{x_1^2 + x_2^2}$, considerando esta transformación podemos replantear $\mathcal{H}$ como el conjunto de todas las funciones tales que $h(x) = 1$ cuando $a \leq f(x) \leq b$ para $0 \leq a \leq b$ y $h(x) = -1$ en otro caso. Este conjunto de hipótesis es muy parecido al de intervalo positivo, por lo tanto estoy casi seguro que la respuesta es que $m_{\mathcal{H}}(N) = C_2^{N+1} + 1$.