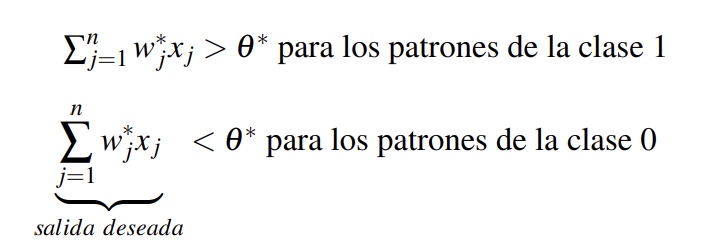
Demostración del teorema de convergencia del perceptrón:

Teorema: Si hay un vector de pesos w\* tal que f (x(p)∙w\*)=t(p) para todos los p, entonces para cualquier vector de inicio w, la regla de aprendizaje del perceptrón convergerá en un número finito de pasos a un vector de pesos (no necesariamente único y no necesariamente w\*) que proporciona la respuesta correcta para todos los vectores de entrenamiento.

Supón patrones linealmente separables, por ende existen unos valores w1,...,wn,θ ∗ tal que:



Sabiendo que la regla de aprendizaje del perceptrón es: wi(nuevo)=wi(anterior) + atxi y que solo se aplica cuando y!=t.

Suponemos una interacción k donde se debe aplicar la regla al ser y(k)!=t(k)

w(k+1) = w(k)+atxi, tenemos por tanto que la nueva distancia entre el vector objetivo y el vector será:

∑(wj(k +1)−wj\* )^2 = ∑(wj(k)+at(k)xj(k) −wj\* )^2

Para simplificar utilizamos las propiedades del binomio simple:

(w+atx-b)^2 = (w+atx-b) \* (w+atx-b) = w^2+watx-wb+atxw+(atx)^2-atxb-bw-atxb+b^2

Agrupamos por términos semejantes:

w^2+watx+watx-wb-wb+(atx)^2-atxb-atxb+b^2=

w^2+2watx-2wb+(atx)^2-2atxb+b^2=

(w^2-2wb+b^2) + 2atxw-2atxb+(atx)^2=

(w^2-2wb+b^2) + 2atx(w-b) + (atx)^2

(w-b)^2 + 2atx(w-b) + (atx)^2

Dado que ∑(ai+bi) = ∑ai + ∑bi y que ∑a\*bi = a∑bi, separamos los sumatorios y sacamos las constantes:

∑(wj(k)-wj\*)^2 + 2at∑xj(k)(wj(k)-wj\*) + a^2t^2∑xj(k)^2

Estudiamos el sumatorio del medio, primero hacemos el producto

2at∑xj(k)(wj(k)-wj\*) = 2at(t∑xj(k)wj(k)- xj(k)wj\*) =

2at (∑xj(k)\*wj(k) - ∑xj(k)wj\*) =

2at∑xj(k)\*wj(k) - 2at ∑xj(k)wj\*

El primer término siempre será negativo pues si ∑xj(k)\*wj(k)>0 quiere decir que t<0 ya que en caso contrario no se produce el aprendizaje, es decir t siempre tendrá el signo contrario a ∑xj(k)\*wj(k) y por ende siempre dará como resultado un valor negativo.

Por otro lado el segundo término siempre será positivo ya que el signo de t debe coincidir con el de ∑xj(k)wj\*

Por ende llegamos a:

-2a |∑xj(k)\*wj(k)| -2a|∑xj(k)wj\*|

-2a |∑xj(k)\*wj(k)| -2a|∑xj(k)wj\*| <= -2a|∑xj(k)wj\*|

Sabiendo esto, podemos decir:

∑(wj(k +1)−wj\* )^2 <= ∑(wj(k)-wj\*)^2 + a^2∑x^2 -2a|∑xj(k)wj\*|

Sea D(k+1) = ∑(wj(k +1)−wj\* )^2 y D(K) = ∑(wj(k)−wj\* )^2

Siendo L=max(∑xj(k)^2) y T=min(|∑xj(k)wj\*|)

D(k+1) <= D(k) + a^2L - 2aT

D(k+1) <= D(k) + a(aL -2T)

Si tomamos aL-2T<0 es decir a<2T/L, tendríamos que D(k) > D(k+1), es decir, con una 0<a<2T/L, conseguimos que, como mínimo nos acerquemos en la cantidad de aL-2T al vector solución, por lo que llegamos en un número finito de pasos a la solución pues en caso contrario D(K) tomaría un valor negativo lo cual no es posible.

Demostración del teorema del aproximador universal:

Una red neuronal con una sola capa oculta y una función de activación no lineal puede aproximar cualquier función continua en un espacio compacto con cualquier nivel de precisión deseado.

La idea es la siguiente, comprimes mucho el gráfico con un peso muy grande.



Con una capa oculta y dos neuronas ya seríamos capaces de conseguir



De esta forma, añadiendo más y más neuronas en la capa oculta podríamos hacer una función torre tan parecida a la función original como se desee

