CURSO ESPECIALIZACIÓN INTELIGENCIA ARTIFICIAL Y BIG DATA

Tema 5: Perceptrón

Profesor: Sebastián Rubio Valero

Diciembre 2025

BORRADOR:NOVIEMBRE

1 Entrenamiento del Perceptrón

Por su sencillez estudiaremos primero el ajuste de los peso en los perceptrones, empezando además por el caso más elemental de un único elemento. La función local es la suma ponderada y el umbral es abrupto de forma que la respuesta es binaria

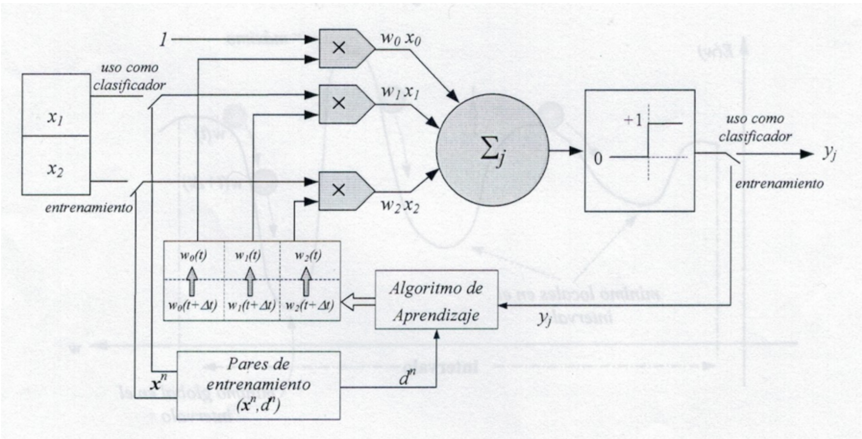


Figure 1: Enter Caption

$$y = \begin{cases} 1, & \text{si } w^T \cdot x > 0, \\ 0, & \text{si } w^T \cdot x \leq 0 \end{cases} \tag{1}$$

Supongamos ahora que conocemos las respuesta deseada d^n, para un conjunto P de patrones $\mathbf{x^n}$, $n = 1, 2, ..., p$. Al ir aplicando de forma sucesiva los distintos pares ($\mathbf{x^n}$, d^n calculamos el error y lo usamos para modificar el valor de los pesos. Si la respuesta es correcta no introducimos ninguna modificación y si es incorrecta (clasifica como "0" lo que debía ser clasificado como "1" o viceversa) introducimos un incremento en los pesos tal que

$$\Delta w^n_{ji} = \mu(d^n_i - y^n_i).x^n_j$$

donde μ es la tasa de aprendizaje, de la misma forma que lo vimos en la regresión lineal.

Los pasos a seguir en este método iterativo son los siguientes:

1. Se especifica el conjunto de datos de entrenamiento ($\mathbf{x^n}$, d^n)

2. Se elige un conjunto inicial de valores para la matriz de pesos $\mathbf{w_{ij}}$

BORRADOR: NOV...RE

3. Se elige un valor para la tasa de aprendizaje μ que podremos modificar de forma empírica para ajustar el ritmo de aprendizaje y garantizar la convergencia

4. Se presenta el vector de entradas correspondientes al primer para de entrenamiento ($\mathbf{x^1}$, d^1) y se calcula la respuesta real (y^1), de acuerdo con la expresión

$$y = \begin{cases} 1, & \text{si } w^T \cdot x > 0, \\ 0, & \text{si } w^T \cdot x \leq 0 \end{cases} \tag{2}$$

5. Se calcula el error ($d^1 - y^1$) y el incremento correspondiente en los pesos

6. Se sustituye $\mathbf{x^n}$ por $\mathbf{x^{n+1}}$ y se regresa al paso 4 has que agotemos todos los datos de entrenamiento.

En la siguiente tabla tenemos los cálculos de los pesos, observa que al ser discreta (binaria) la salida, la actualización de los pesos en cada iteración, se incrementa o decrementa de forma proporcional a la tasa de aprendizaje.

Table 1: Cálculo de pesos

d^n	y^n	$d^n - y^n$	$\Delta w_{ji}^n(t)$
0	0	0	0
0	1	-1	$-\mu.x_i^1$
1	0	1	$\mu.x_i^1$
1	1	0	0

El proceso de entrenamiento se para cuando conseguimos que el error sea cero para todos los pares de entrenamiento, de que los clasifica adecuamente a todos ellos o cuando se ha alcanzado un cierto número de iteraciones, aunque no hayamos conseguido clasificarlas todas correctamente.

La salida puede ser una función Heaviside (2), o puede ser una función de signo

$$y = \begin{cases} -1, & \text{si } w^T \cdot x < 0 \\ 0, & \text{si } w^T \cdot x = 0 \\ 1, & \text{si } w^T \cdot x > 0 \end{cases} \tag{3}$$

A esta neurona también se le conoce con el nombre de ULU (Unidad Lógica de Umbral). Una sola ULU puede actuar como un clasificador binario. Calcula la función lineal de sus entradas, y si el resultado supera un umbral, da como salida la clase positiva. En caso contrario, genera la clase negativa.

BORRADOR.NOVIEMBRE

Un perceptrón de una o mas ULU organizadas en una sola capa, donde todas las ULU se conectan a todas las entradas.En un perceptron, el límite de decisión de cada neurona de salida es lineal, así que los perceptrones son incapaces de aprender patrones complejos. Sin embargo, si las instancias de entrenamiento son separables linealmente, este modelo converge a una solución En un perceptron, el límite de de-

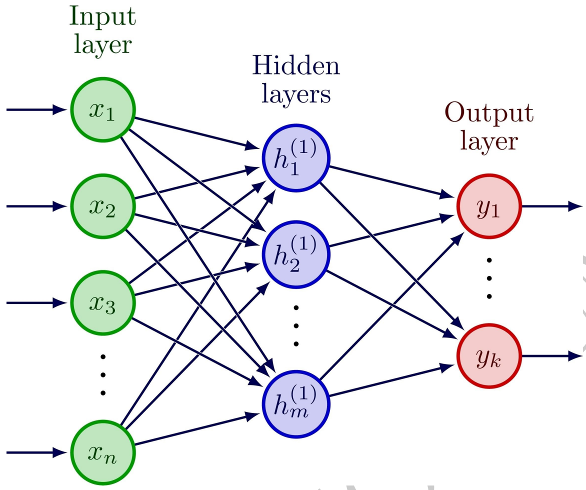


Figure 2: Enter Caption

cisión de cada neurona de salida es lineal, así que los perceptrones son incapaces de aprender patrones complejos. Sin embargo, si las instancias de entrenamiento son separables linealmente, este modelo converge a una solución.

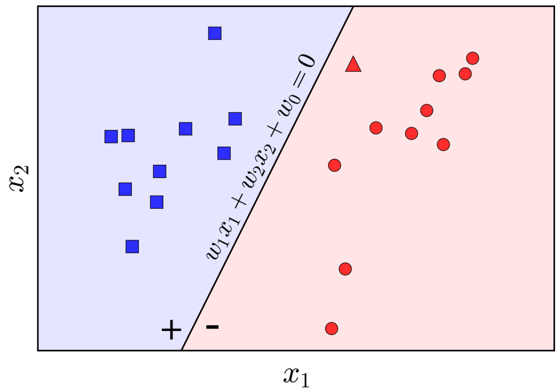


Figure 3: Enter Caption

Table 2: Cálculo de pesos

d^n	y^n	$d^n - y^n$	$\Delta w_{ji}^n(t)$
0	0	0	0
0	1	-1	$-\mu.x_i^1$
1	0	1	$\mu.x_i^1$
1	1	0	0

La Sensibilidad mide la capacidad de un modelo para identificar correctamente todos los casos positivos reales. En otras palabras, responde a la pregunta: De todos los casos que son realmente positivos, ¿cuántos clasificó el modelo correctamente como positivos?

Es especialmente importante cuando el costo de un Falso Negativo (FN) —un positivo real que se clasifica erróneamente como negativo— es alto. Por ejemplo, en el diagnóstico médico, es crucial que un modelo de detección de enfermedades tenga alta sensibilidad para no pasar por alto a pacientes enfermos.

La Especificidad mide la capacidad de un modelo para identificar correctamente todos los casos negativos reales.En otras palabras, responde a la pregunta: De todos los casos que son realmente negativos, ¿cuántos clasificó el modelo correctamente como negativos? Es muy importante cuando el costo de un Falso Positivo (FP) es alto. Un Falso Positivo es un caso negativo real que se clasifica erróneamente como positivo. Por ejemplo:

Diagnóstico Médico: Si un paciente sano es diagnosticado erróneamente con una enfermedad grave (Falso Positivo), esto puede llevar a tratamientos innecesarios y costosos, estrés, y procedimientos invasivos. Una alta especificidad minimiza este riesgo.

Filtros de Spam: Un correo electrónico legítimo que se clasifica como spam (Falso Positivo) resulta en que el usuario pierda información importante.

La Precisión (Precision) es una métrica fundamental que a menudo se estudia junto con la Sensibilidad y la Especificidad.

La Precisión (o Valor Predictivo Positivo - VPP) mide la fiabilidad de las predicciones positivas de un modelo. En otras palabras, responde a la pregunta:

De todos los casos que el modelo clasificó como positivos, ¿cuántos eran realmente positivos?

La Precisión se centra en la calidad del resultado positivo del modelo y es vital cuando la prioridad es minimizar los Falsos Positivos (FP).

Ejemplo de Importancia Detección de Fraude Bancario: Si un modelo marca una transacción legítima como fraudulenta (Falso Positivo), esto puede causar inconve-

BORRADOR: NOVIEMBRE

nientes al cliente (bloqueo de tarjeta, llamada de verificación) y generar altos costos operativos para el banco. En este caso, se necesita una alta precisión para asegurar que la mayoría de las alertas son casos de fraude real.

Ejercicio

Consideremos un perceptrón con cuatro argumentos, pesos $w_0 = 0,7$ $w_1 = -0,4$ $w_2 = 0,3$, $w_3 = 0,1$, $w_4 = -1$ y que usa la función umbral como función de activación.

Se pide calcular la matriz de confusión, y derivar a partir de ella la tasa de aciertos,la sensibilidad, la especificidad y la precisión, del perceptrón sobre conjunto de ejemplos.

x1	x2	x3	x4	y
-0.5	0.5	-1	-1	0
1	1	0.5	-1	1
-1	1	1	1	1
0	-1	-1	1	0
-1	0.5	0.5	1	1
-0.5	-0.5	0	-0.5	0
-0.5	0	0.5	-0.5	1
1	0.5	0	0	0
0.5	0.5	1	1	1
0.5	0.5	0.5	0.5	0

En este primer ejemplo no vamos hacer entrenamiento

Sabemos por otros temas estudiados que el intercepto está multiplicado por "1". Sabiendo esto vamos a calcular cada suma ponderada por los pesos de la siguiente forma:

$$z = 1.w_0 + x_1.w_1 + x_2.w_2 + x_3.w_3 + x_4 w_4$$

Una vez calculada la suma le aplicamos la función de activación $f(z)$ que en este ejemplo es la función umbral. Para la primera instancia del dataset de entrenamiento quedaría de la siguiente forma

$$z_1 = 1.0,7 + (-0,5)(-0,4) + (0,5)(0,3) + (-1)(0,1) + (-1)(-1) = 1,95$$
$$f(1,95) = 1$$

Siguiendo estos cálculos obtenemos que

$$z_2 = 1,65; f(z_2) = 1$$
$$z_3 = 0,5; f(z_3) = 1$$
$$z_4 = -0,7; f(z_4) = 0$$
$$z_5 = 0,3 f(z_5) = 1$$
$$z_6 = 1,25 f(z_6) = 1$$
$$z_7 = 1,45; f(z_7) = 1$$
$$z_8 = 0,45, f(z_8) = 1$$
$$z_9 = -0,25; f(z_9) = 0$$
$$z_{10} = 0,2; f(z_{10} = 1$$

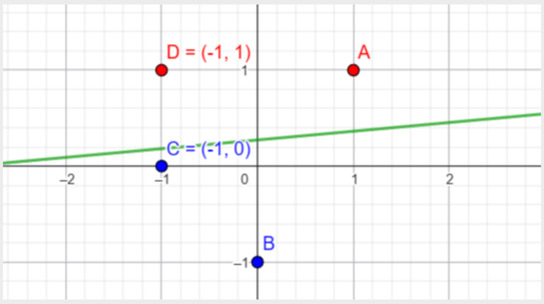
Repite el ejercicio con la función de activación signo

Ejercicio 2

Consideremos el conjunto de entradas siguiente:

x_1	x_2	y
1	1	1
0	-1	0
-1	0	0
-1	1	1

El conjunto de puntos es separable por lo tanto podemos aplicar el perceptrón.



Vamos a entrenar el modelo para una única iteración, sabiendo que $w_0 = 0,1$, $w_1 = w_2 = -0,1$ y una tasa de aprendizaje $\mu = 0,1$

Calculamos la suma ponderada de la primera entrada, acuerdate de que el bias debe ser multiplicado por "1"

$$z = 1.(0,1) + 1.(-0,1) + 1.(-0,1) = -0,1$$

El siguiente paso le aplicamos la función de activación, que en nuestro caso es la función umbral

$$f(z) = 0$$

Observamos que la salida esperada es "1", y sin embargo, la obtenida es "0", por lo tanto hay un error. Así que, debemos de reajustar los pesos de la siguiente forma

$$\begin{gathered} w_0(t+1) = w_0(t) + \mu(y-0) = 0,1 + 0,1(1-0) = 0,2 \\ w_1(t+1) = w_1(t) + \mu(y-0).1 = (-0,1) + 0,1(1-0).1 = 0 \\ w_2(t+1) = w_2(t) + \mu(y-0).1 = (-0,1) + 0,1(1)1 = 0 \end{gathered}$$

Entonces los nuevos pesos son: $w_0 = 0,2$, $w_1 = w_2 = 0$

Cogemos la siguiente entrada y calculamos la suma ponderada con los nuevos pesos

$$z = 1.(0,2) + 0.(0) + (-1).(0) = 0,2 \; f(z) = 1$$

La respuesta esperada era "0" y hemos obtenido "1", se ha cometido un error, y tenemos que volver a calcular los pesos de la misma forma que hemos realizado anteriormente. Los nuevos pesos son

$$w_0 = 0,1, \; w_1 = 0, w_2 = 0,1$$

A continuación cogemos la siguiente entrada y calculamos la suma ponderada con los nuevos pesos y la función de activación, asi sucesivamente hasta que finalicemos con todas las entradas (Se deja como ejercicio que se finalicen los cálculos).

En un caso real, el entrenamiento no acabaría aquí. Si el criterio de fin de entrenamiento es el número de épocas (p.e 4), se debe realizar todo el proceso anterior otras tres veces. Otra forma de indicarle que finalice es si ya no se encuentras errores.

Ejercicio

En este ejercicio debes realizar un entrenamiento de un perceptrón con las siguiente características:

- $w_0 = 0,2$,$w_1 = 0,3$,$w_2 = w_3 = 0,1$
- $\mu = 0,2$
- la función de activación es la función de signo
- Una sólo época

El conjunto de entrenamiento

x_1	x_2	x_3	y
0,5	-1	-1	-1
-1	-0,5	0	1
1	-1	0,5	1
0,5	1	0	1
-1	-1	-0,5	-1

Una vez realizado el entrenamiento responde a las siguientes cuestiones:

1. ¿El conjunto de datos es separable?
2. Obtén la matriz de confusión
3. Calcula la tasa de aciertos, sensibilidad,especificidad y precisión
4. Según todos los resultados obtenidos, saca una conclusión

BORRA