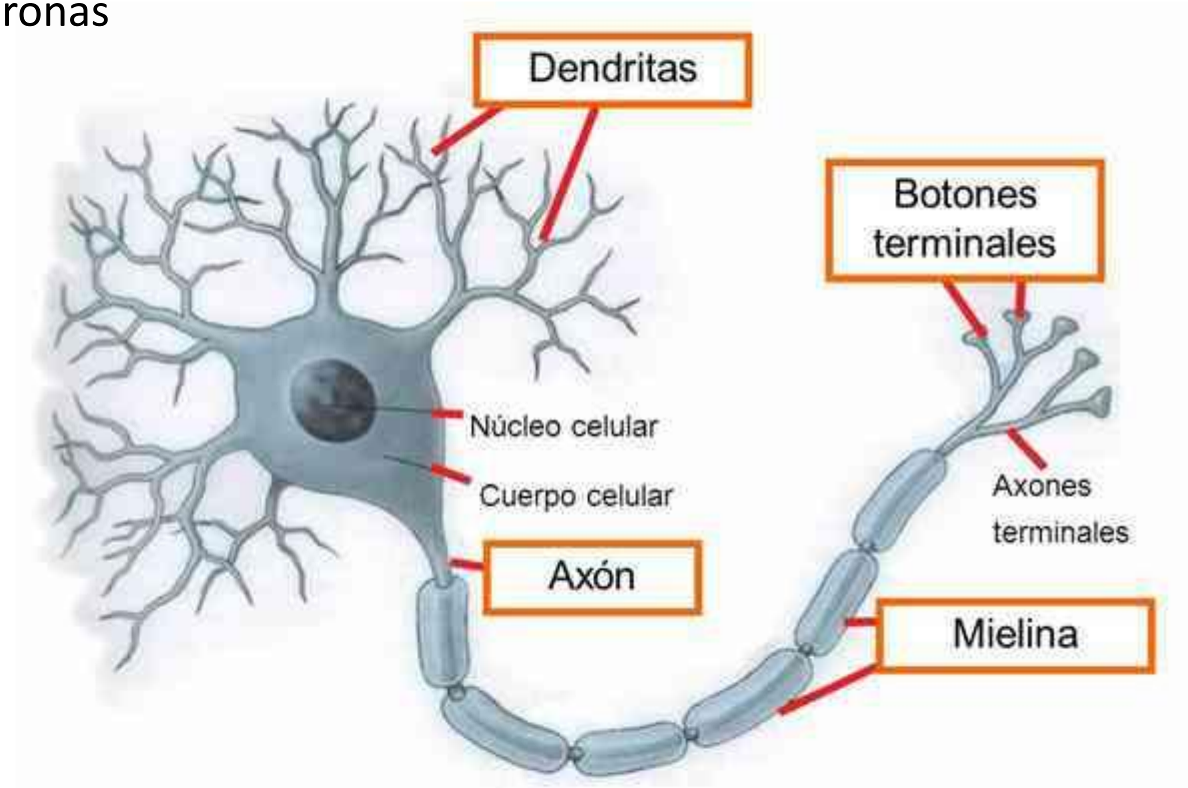
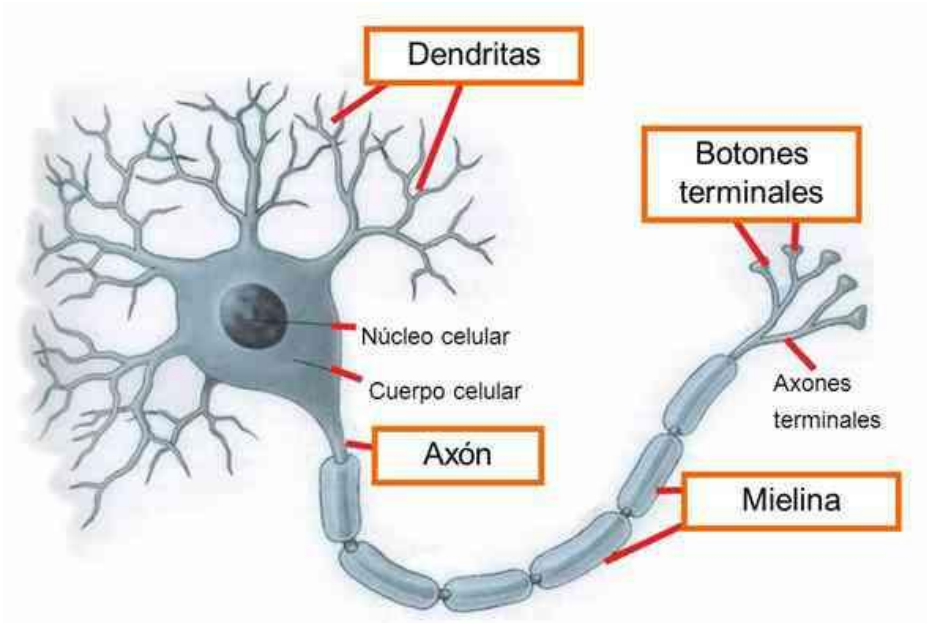
Redes Neuronales

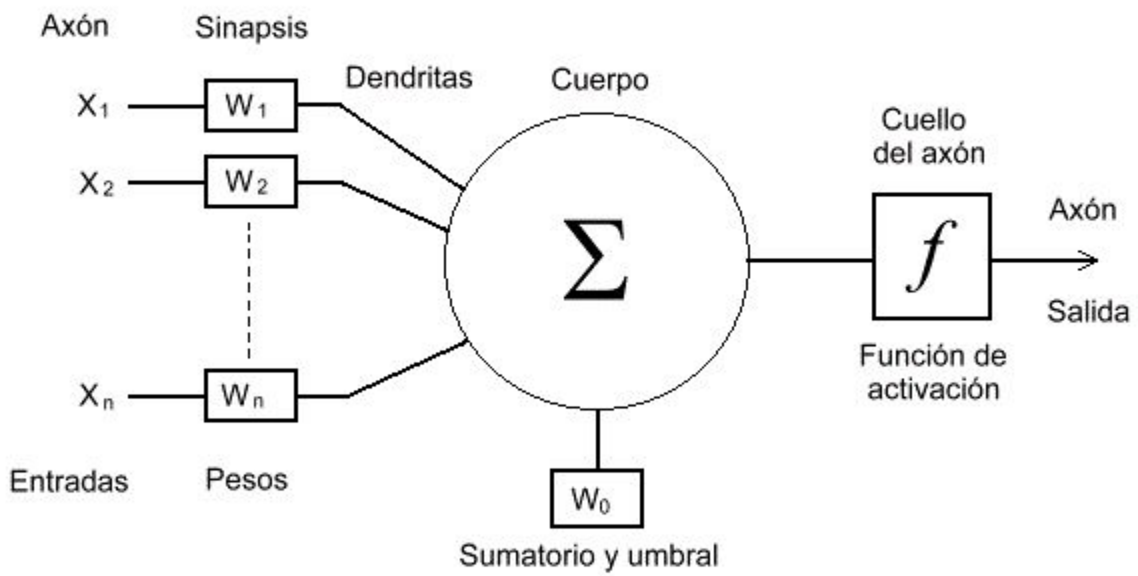
Una introducción…

Frank Rosenblatt, en la década de los 50, se inspira en las neuronas biológicas…



…para definir lo que se conoce como neurona artificial…





- **Núcleo celular**: Zona de procesamiento.
- **Dendritas**: Zonas receptoras.
- **Axón**: Líneas de transmisión.
- **Axones terminales (sinapsis)**: Conexiones excitadoras o inhibidoras.

x1 = 23, x2 = 68, x3 = 170
w1 = 1, w2 = 2, w3 = 3

23(1) + 68 (2) + 170 (3) = 669

El Perceptrón

- Neurona artificial o unidad básica de inferencia en forma de discriminador lineal.

$$y = f(\vec{w} \cdot \vec{x}) = f\left(\sum_j w_j x_j\right)$$

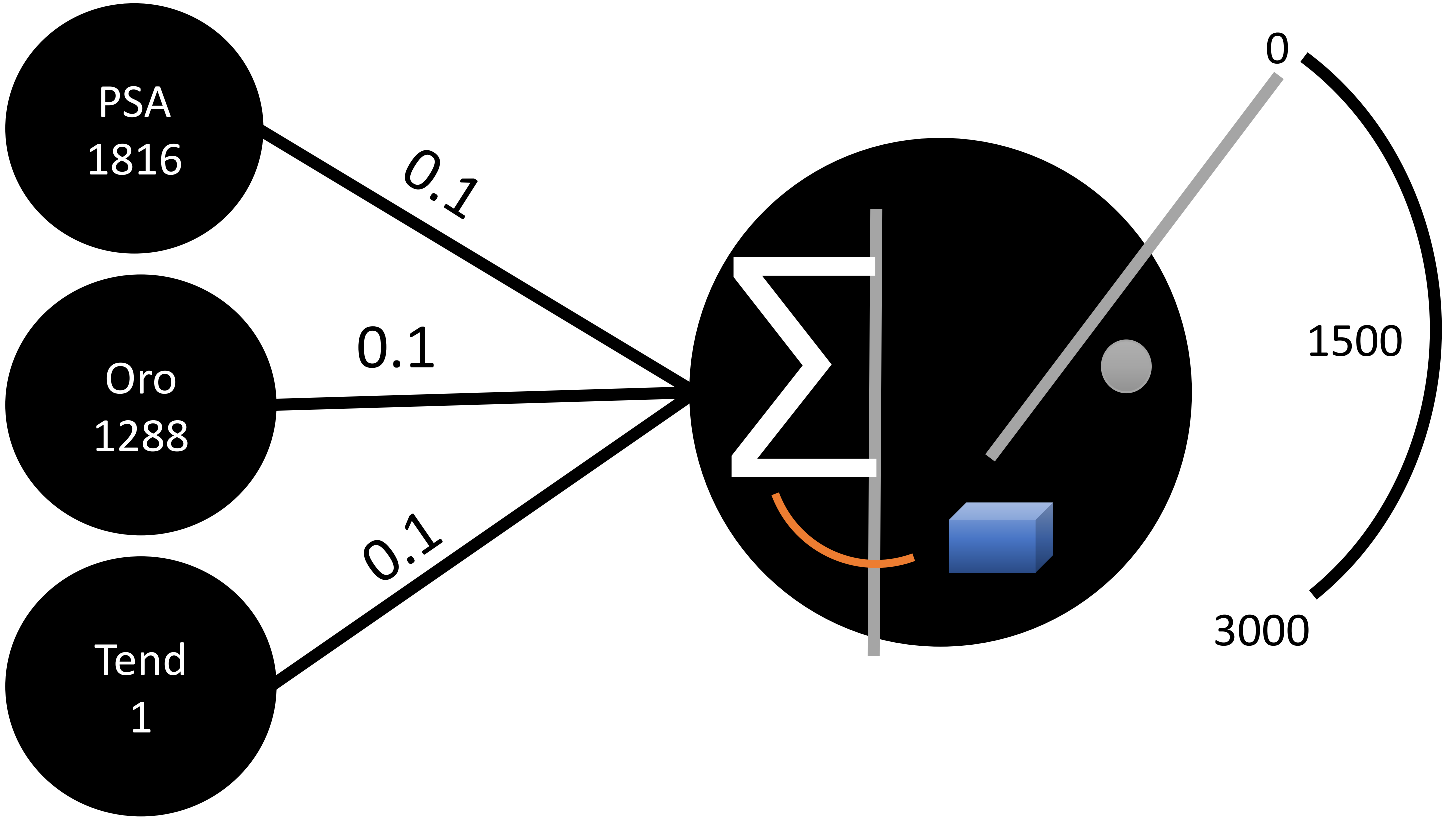
Donde:

- x_j : característica j de la muestra.
- w_j : peso sináptico
- f : función de activación

Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

PERCEPTRÓN

Presentación



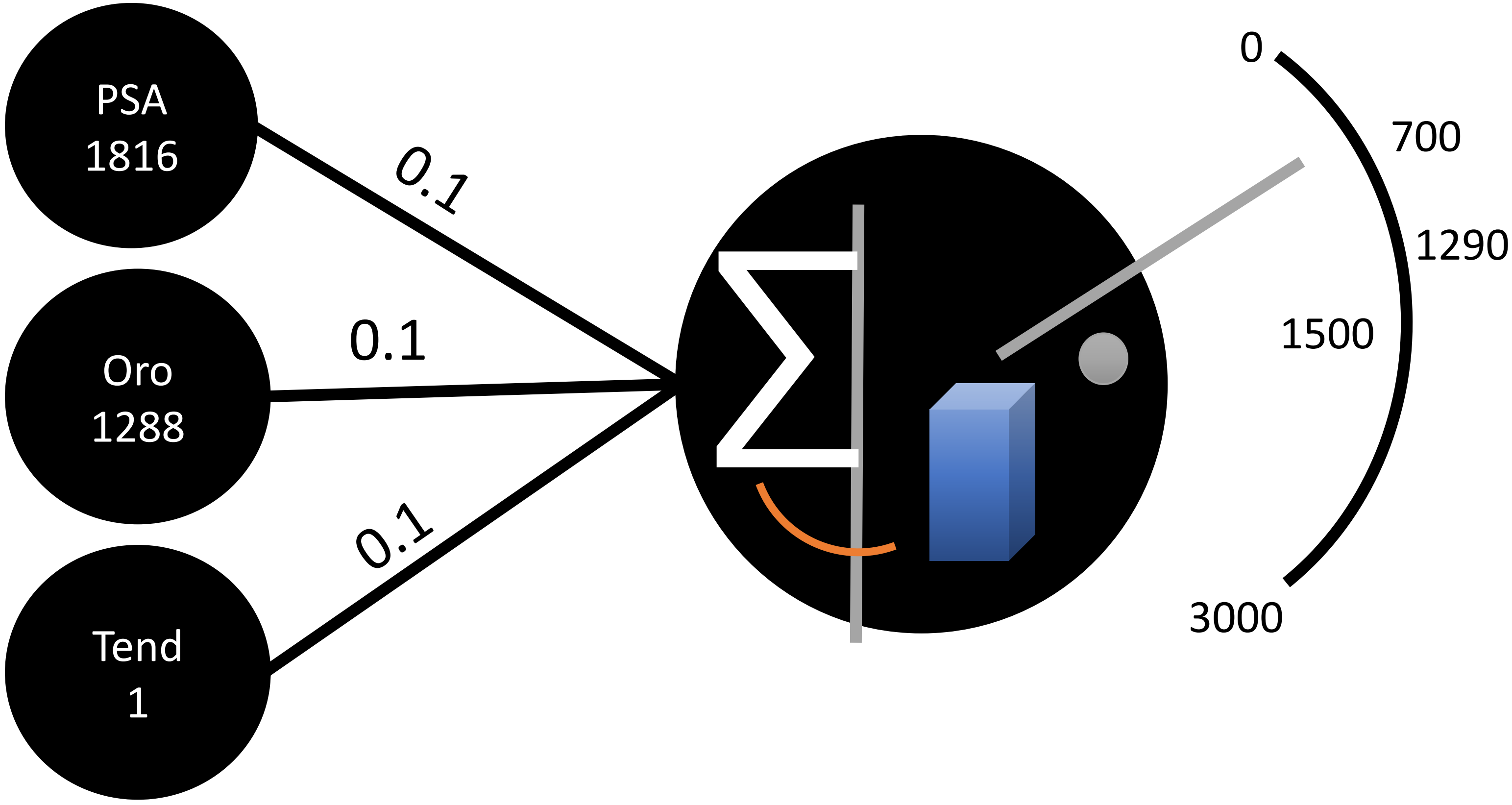
Del Prof. Pablo Sanz

Salida esperada: 1290

Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

PERCEPTRÓN

Primera lectura

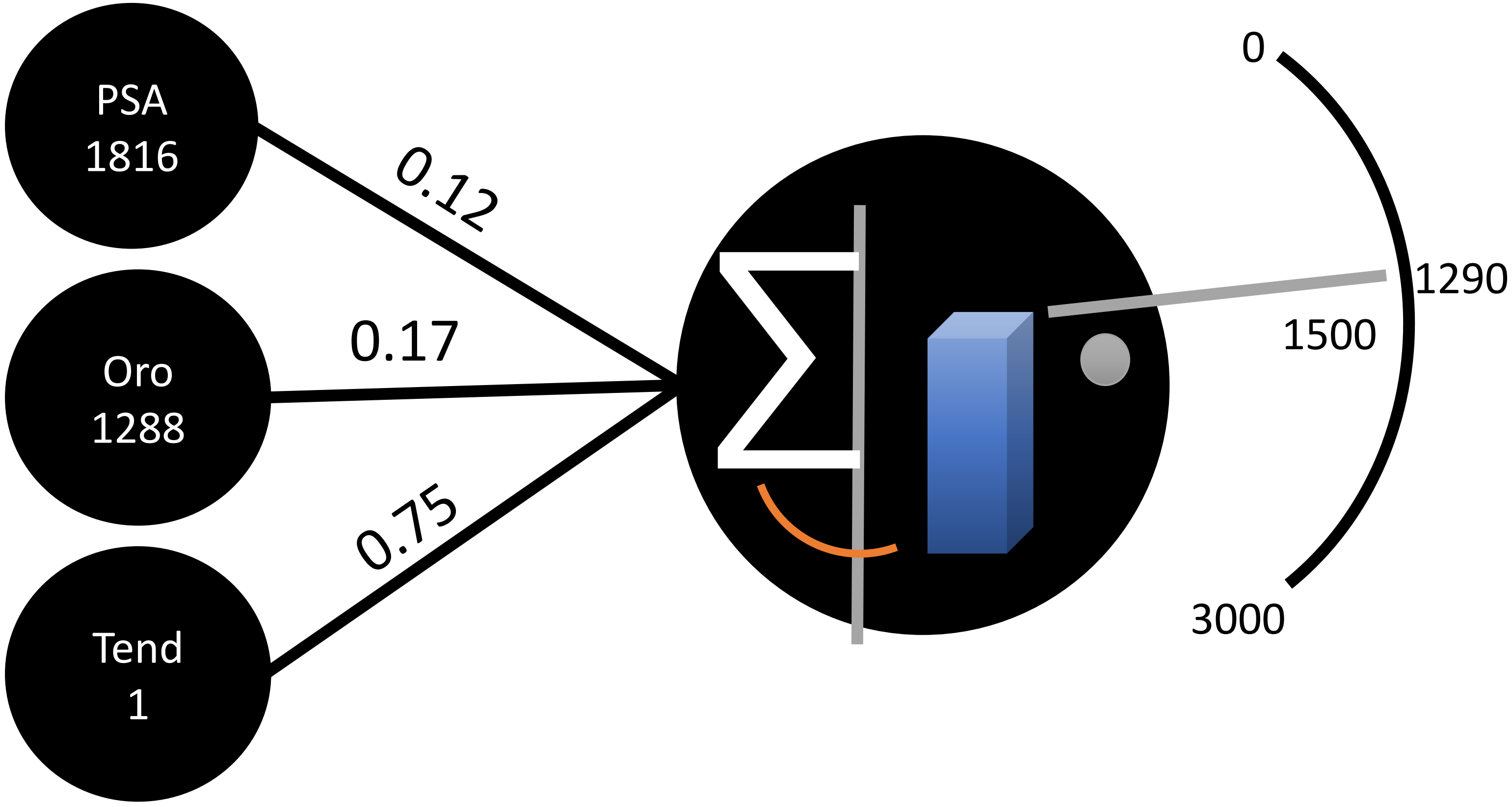


Error: **590**

Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

PERCEPTRÓN

Modificación de pesos

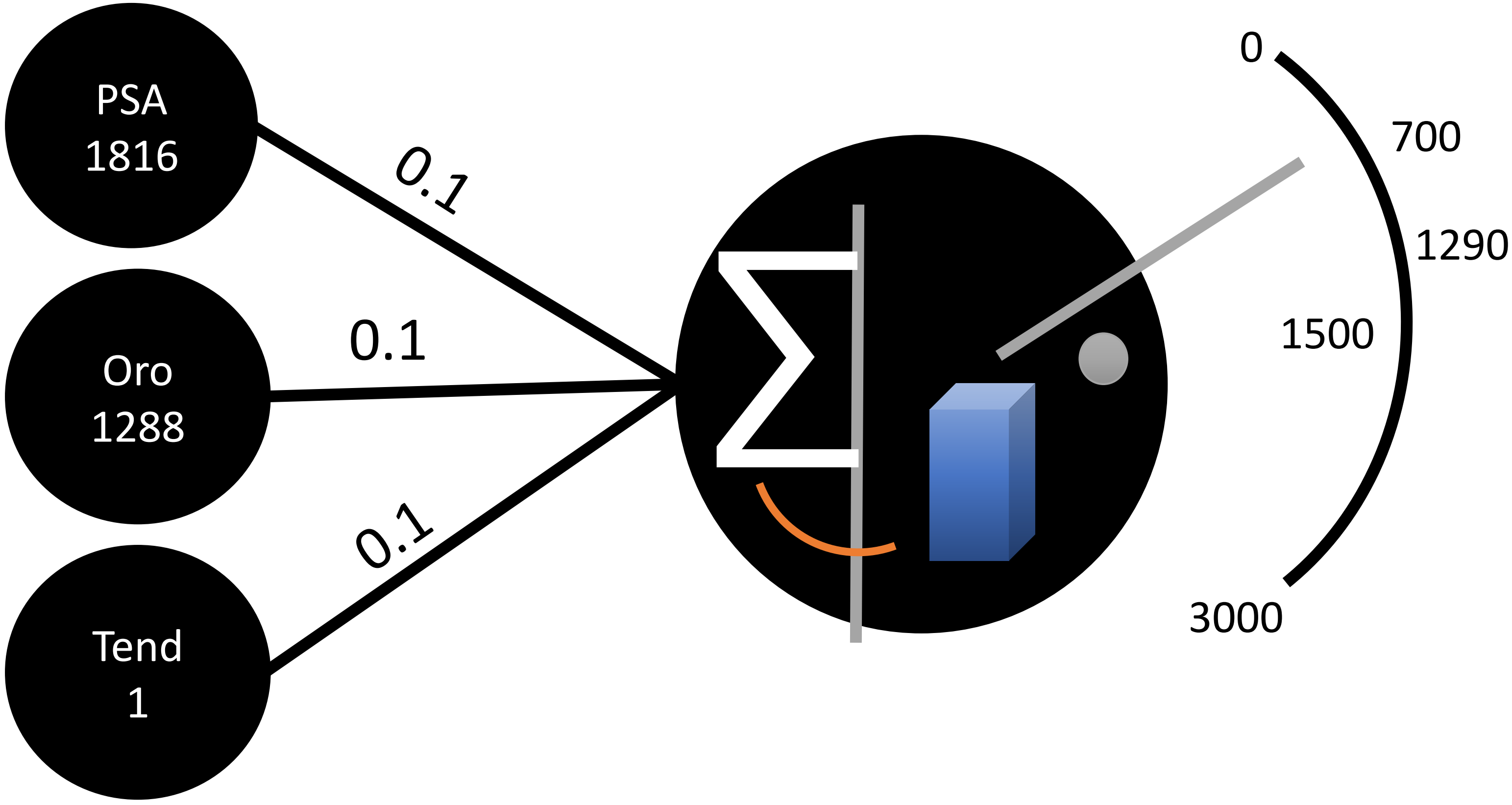


Error: 0

Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

PERCEPTRÓN

Entrenamiento (FP)

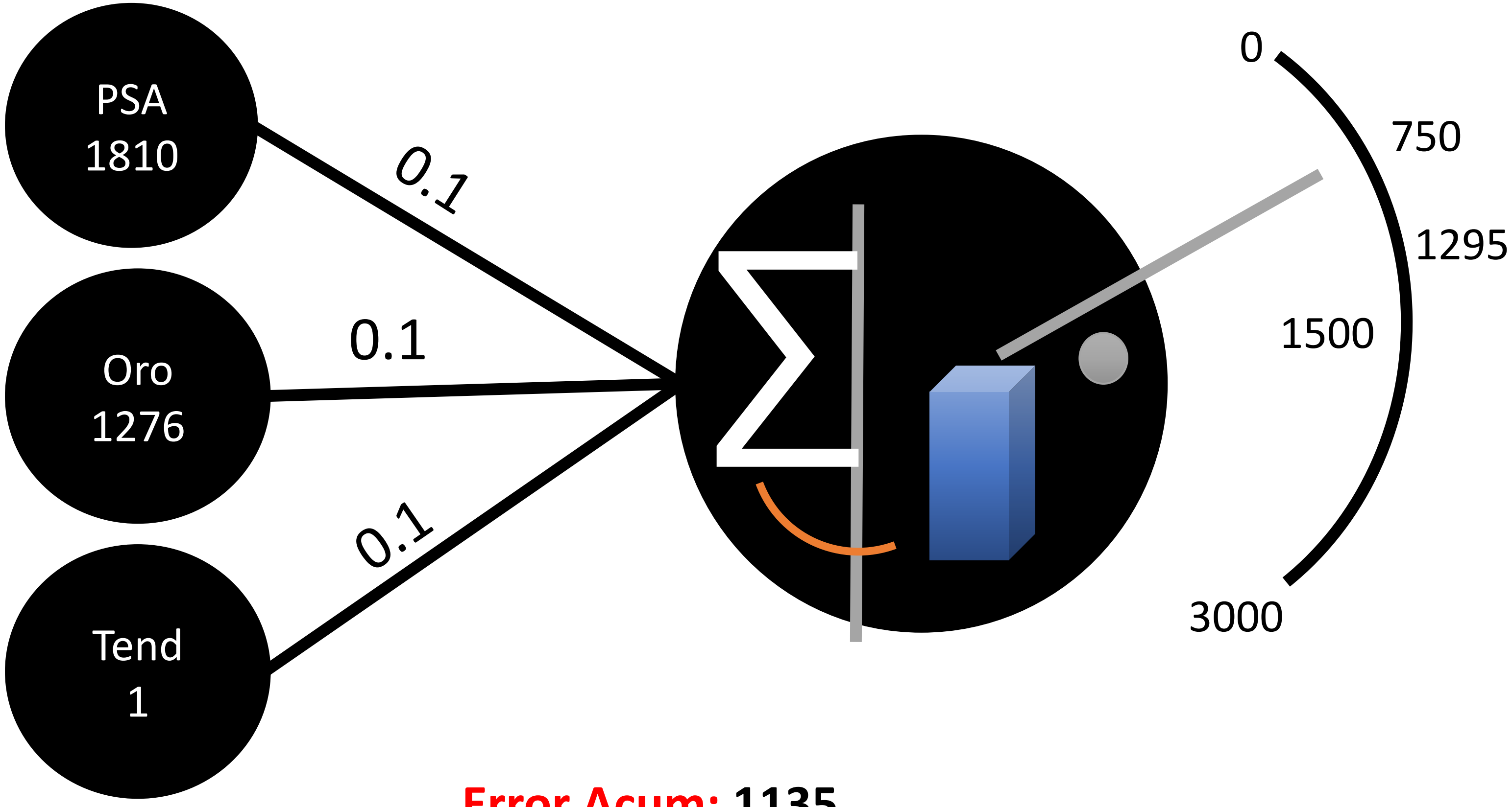


Error: 590

Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

PERCEPTRÓN

Entrenamiento (FP)

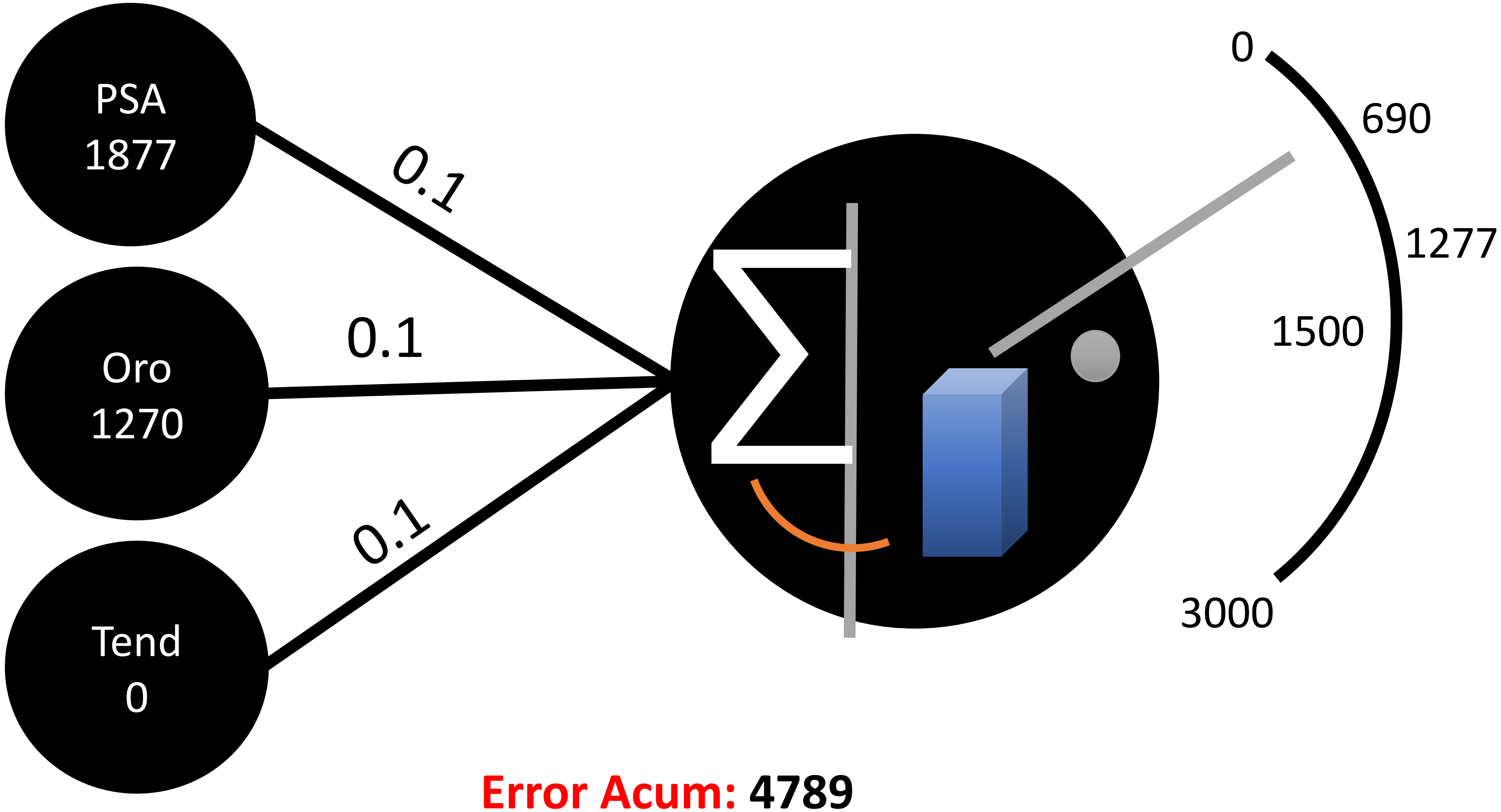


Error Acum: 1135

Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

PERCEPTRÓN

Entrenamiento (FP)

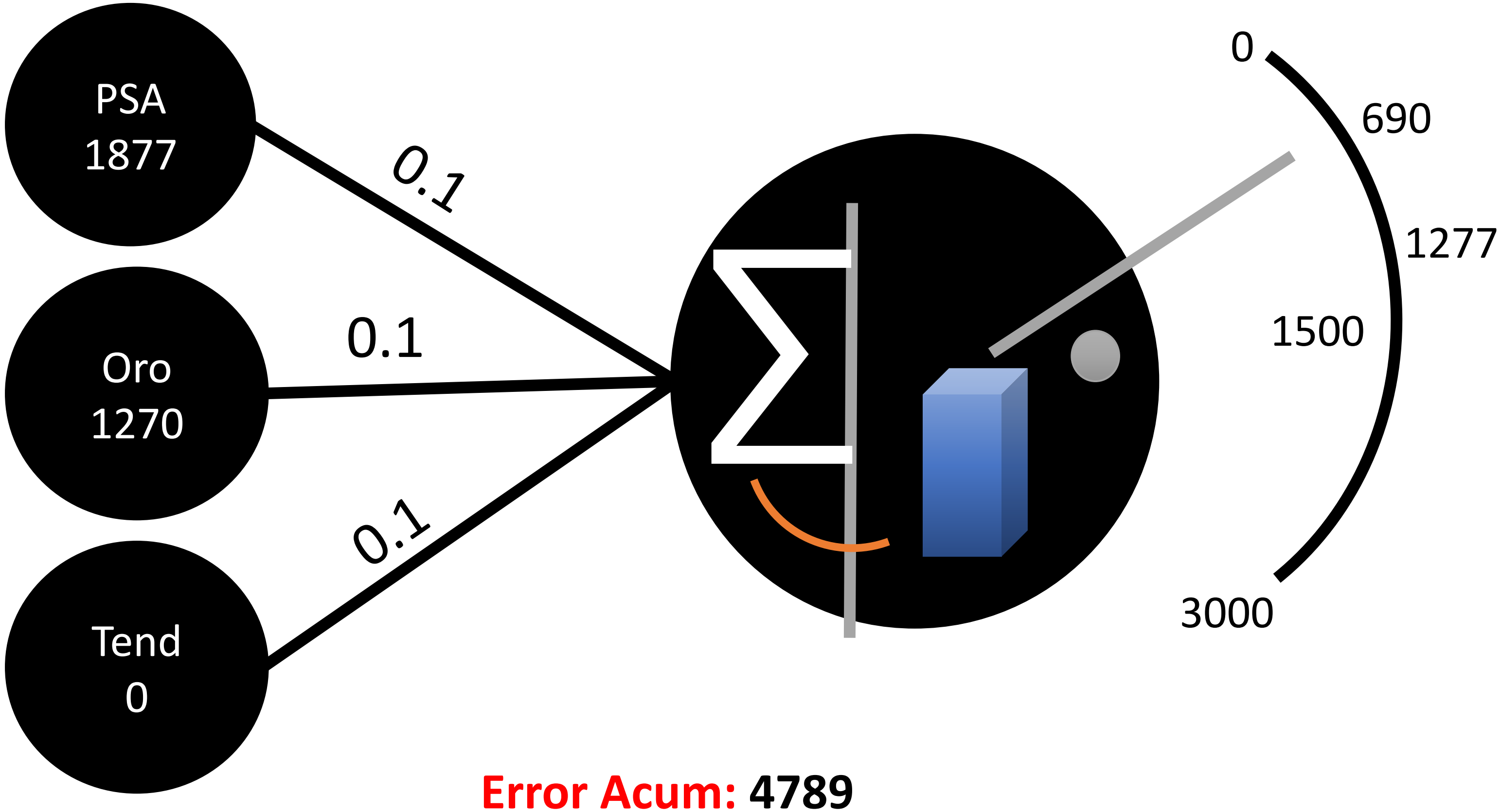


Error Acum: **4789**

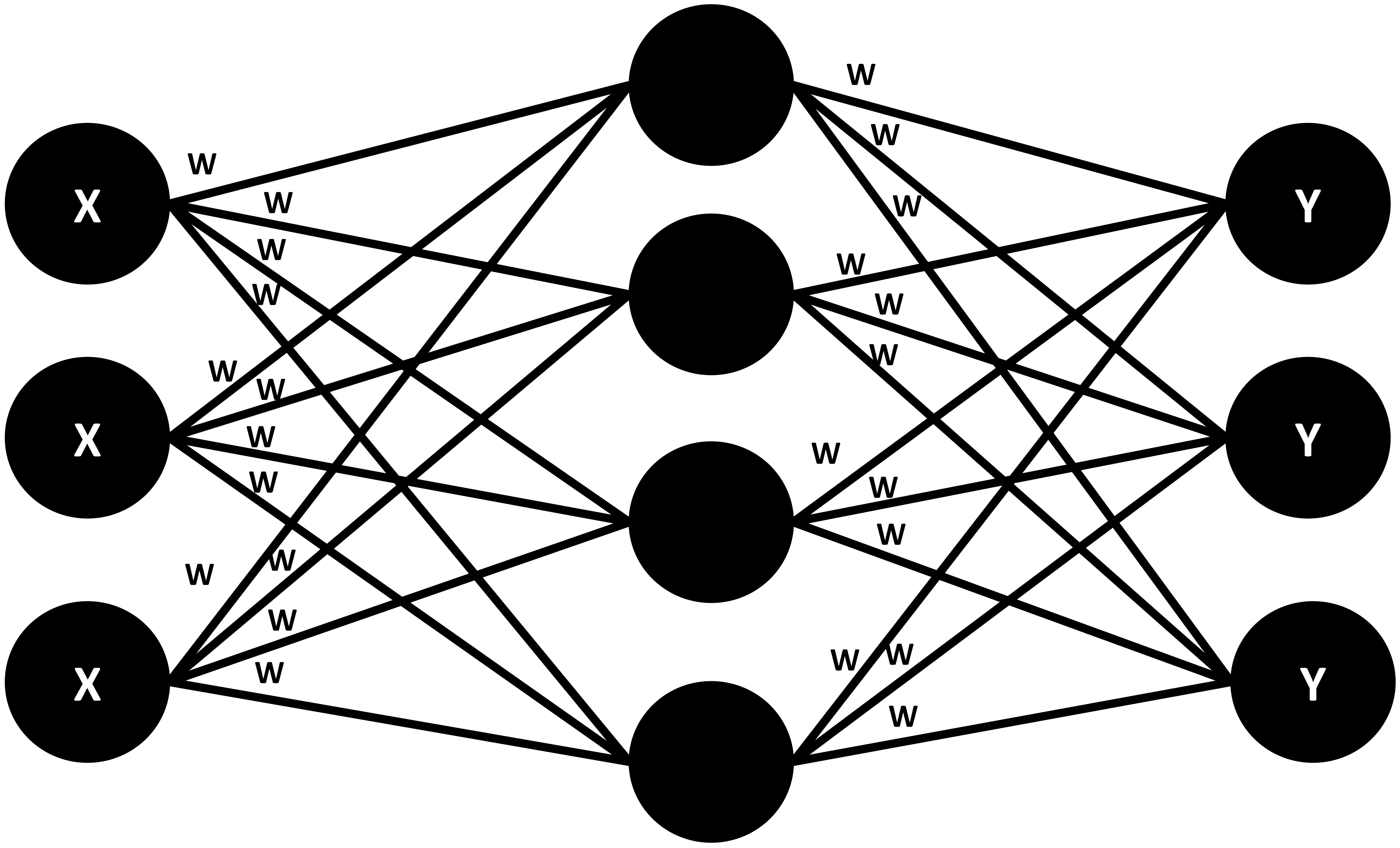
Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

PERCEPTRÓN

Ajuste de pesos
¿Fuerza bruta?



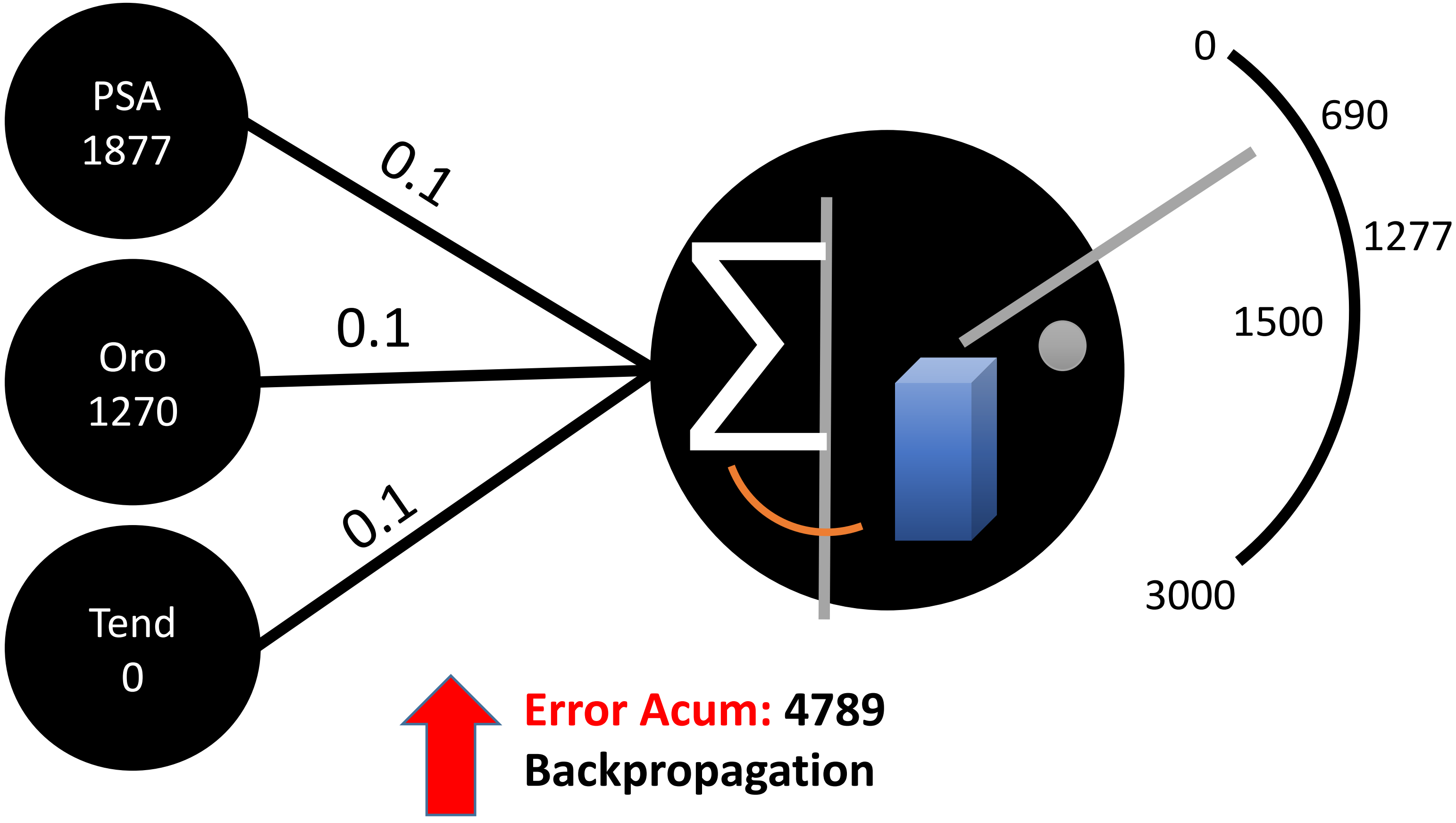
Error Acum: 4789


X
X
X
Y
Y
Y
W

Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

PERCEPTRÓN

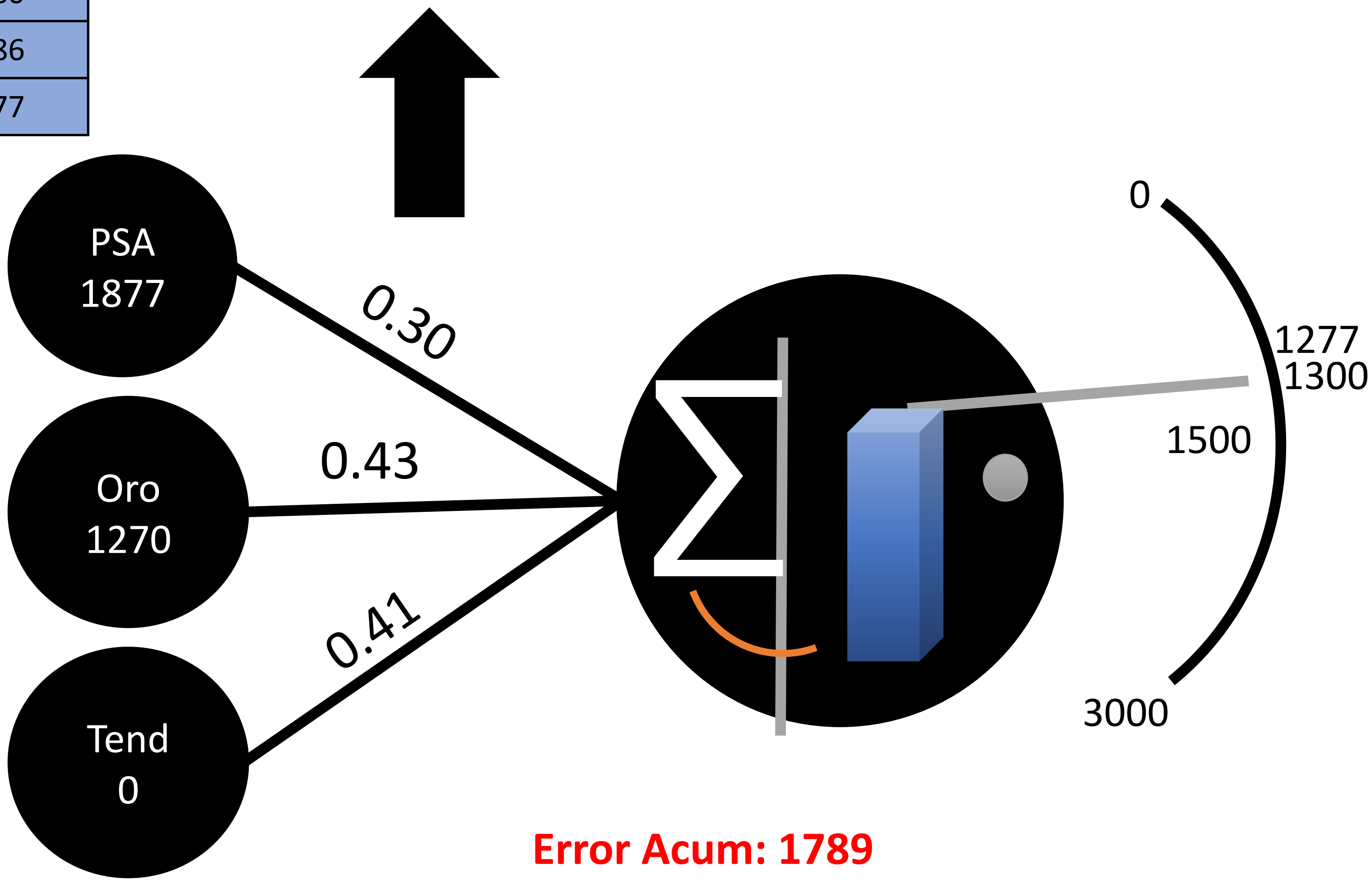
Ajuste de pesos (BP)



Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

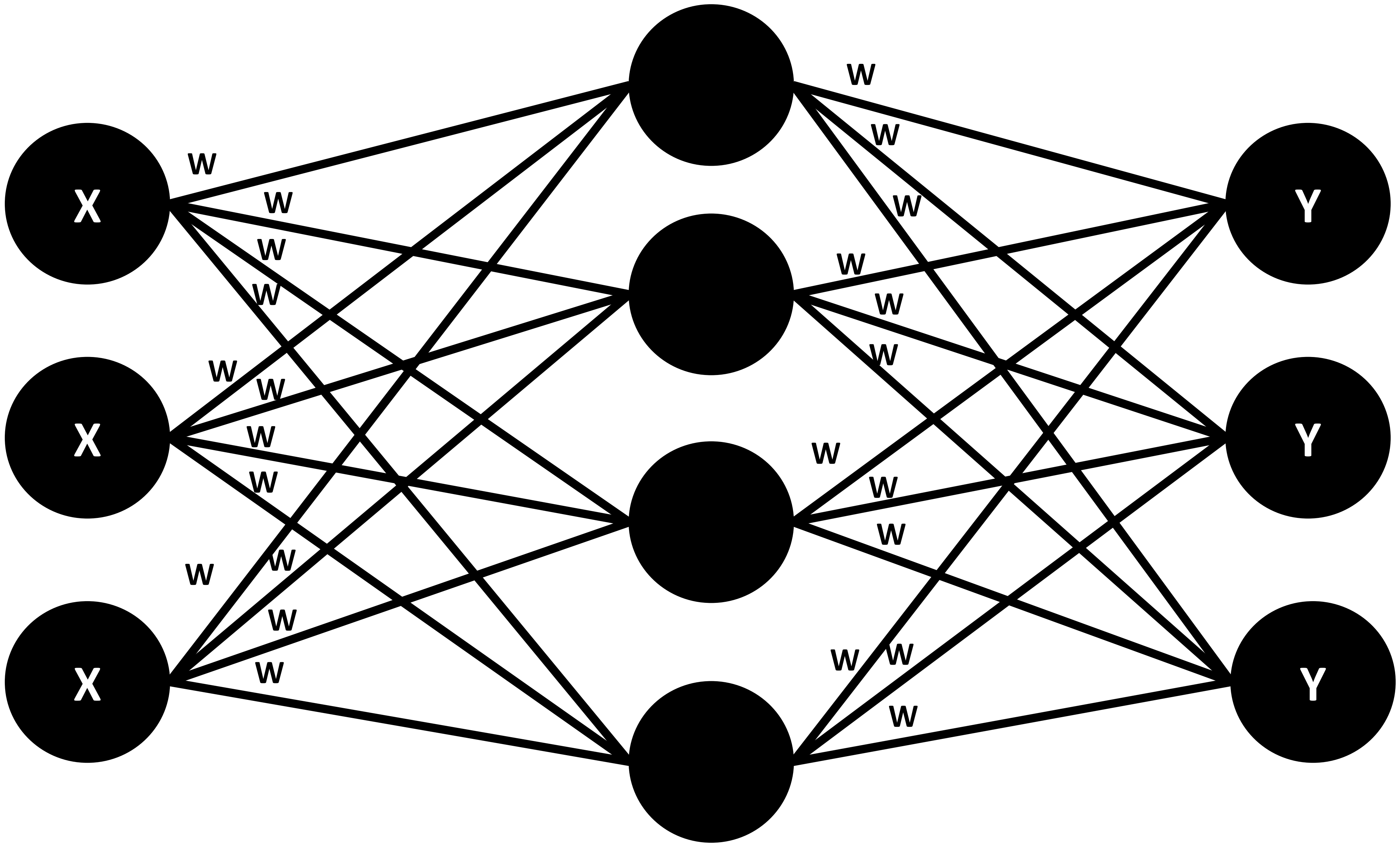
PERCEPTRÓN

Ajuste de pesos (BP)



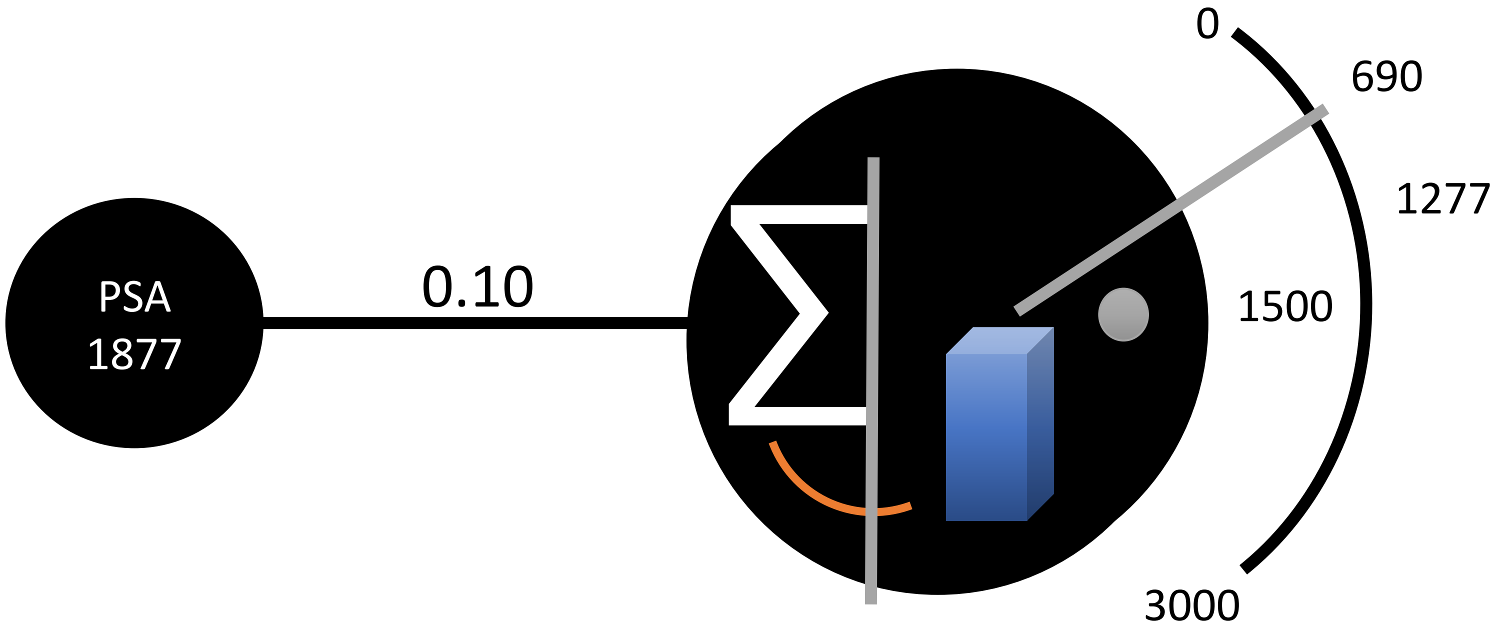
Error Acum: 1789

Backpropagation


X
X
X
Y
Y
Y
W

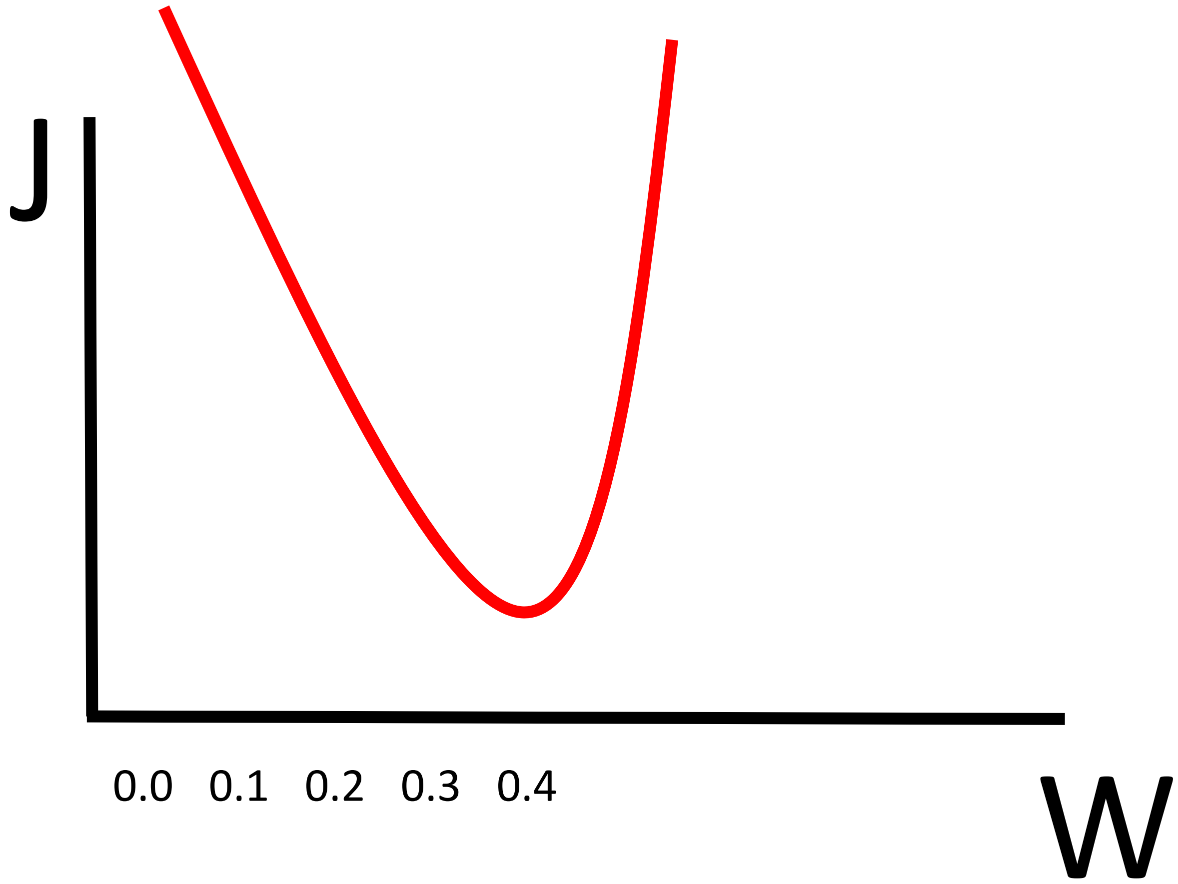
Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277

Descenso del Gradiente



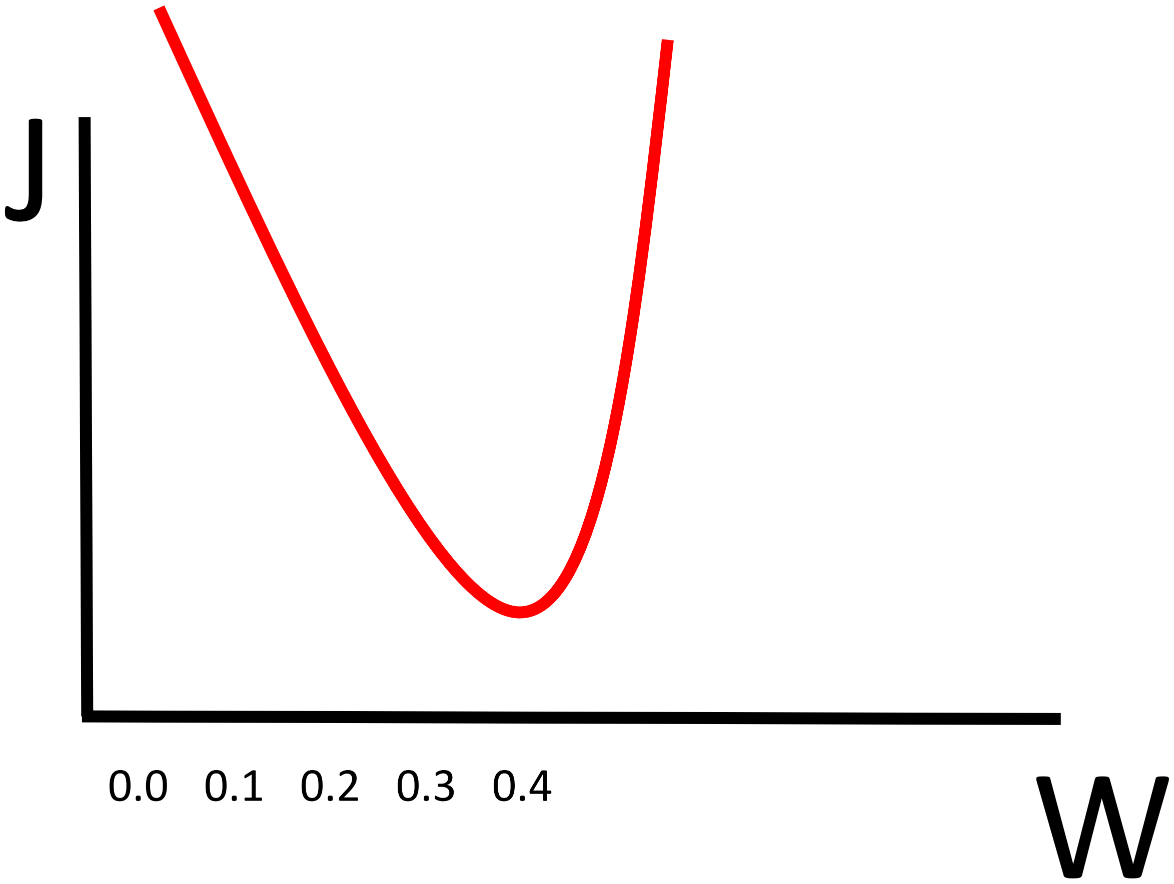
Error Acum: 4789

Descenso del Gradiente



$$J = 1/2 \,.\, \sum (y - y')^2$$

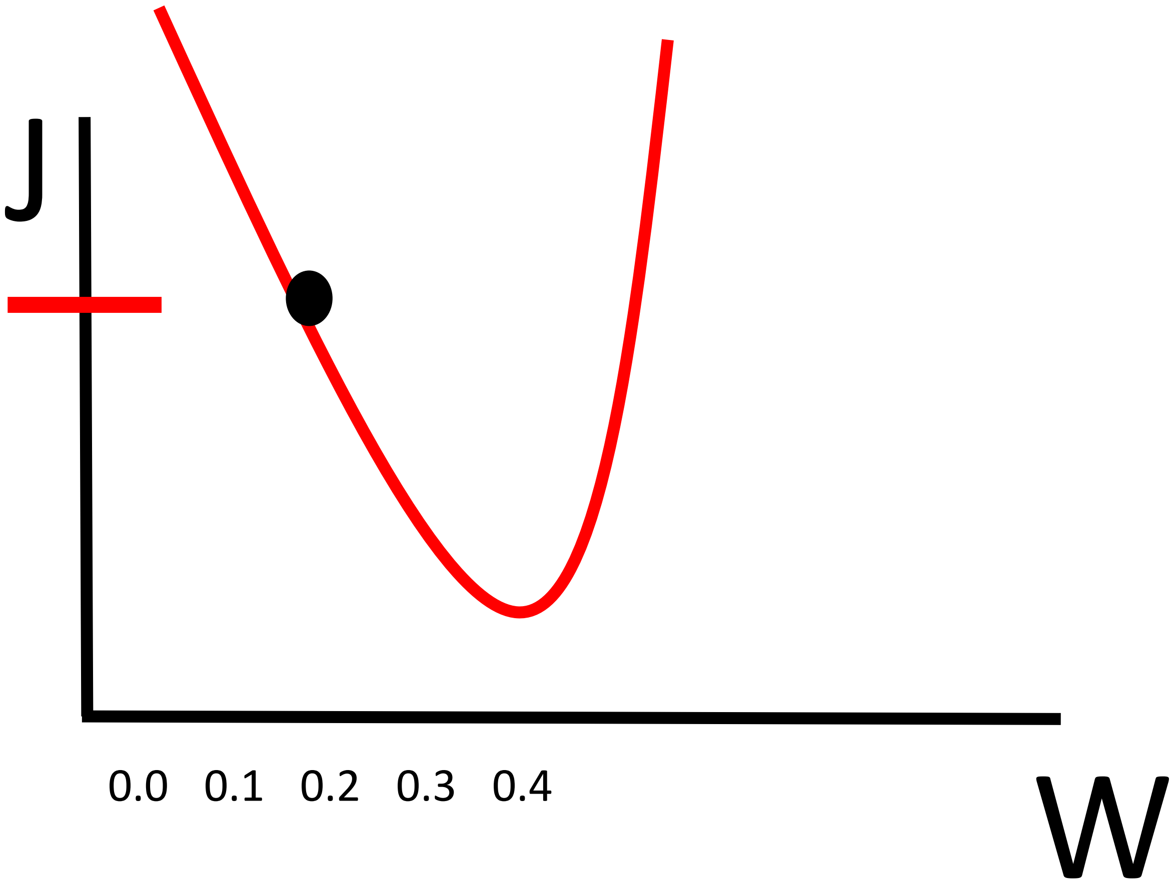
Descenso del Gradiente



Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \, . \sum (y - y')^2$$

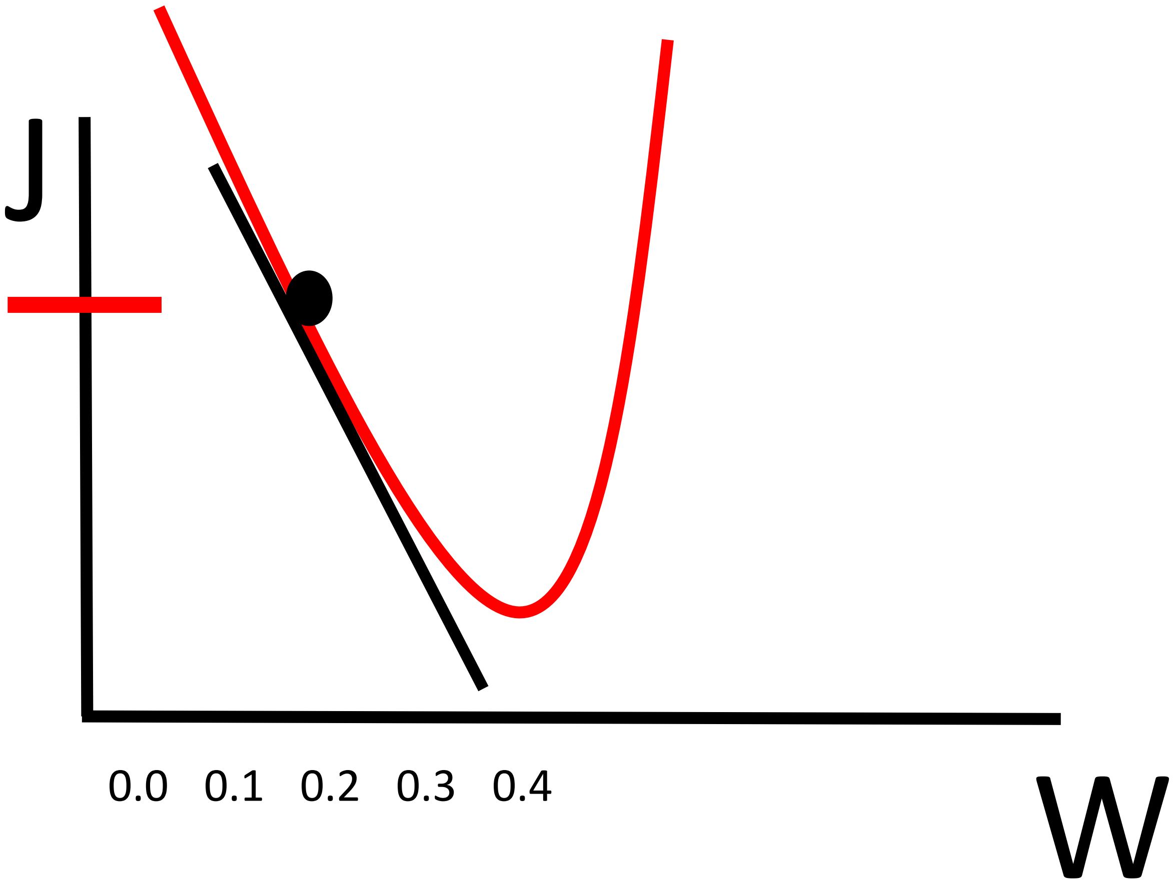
Descenso del Gradiente



Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \,.\, \sum (y - y')^2$$

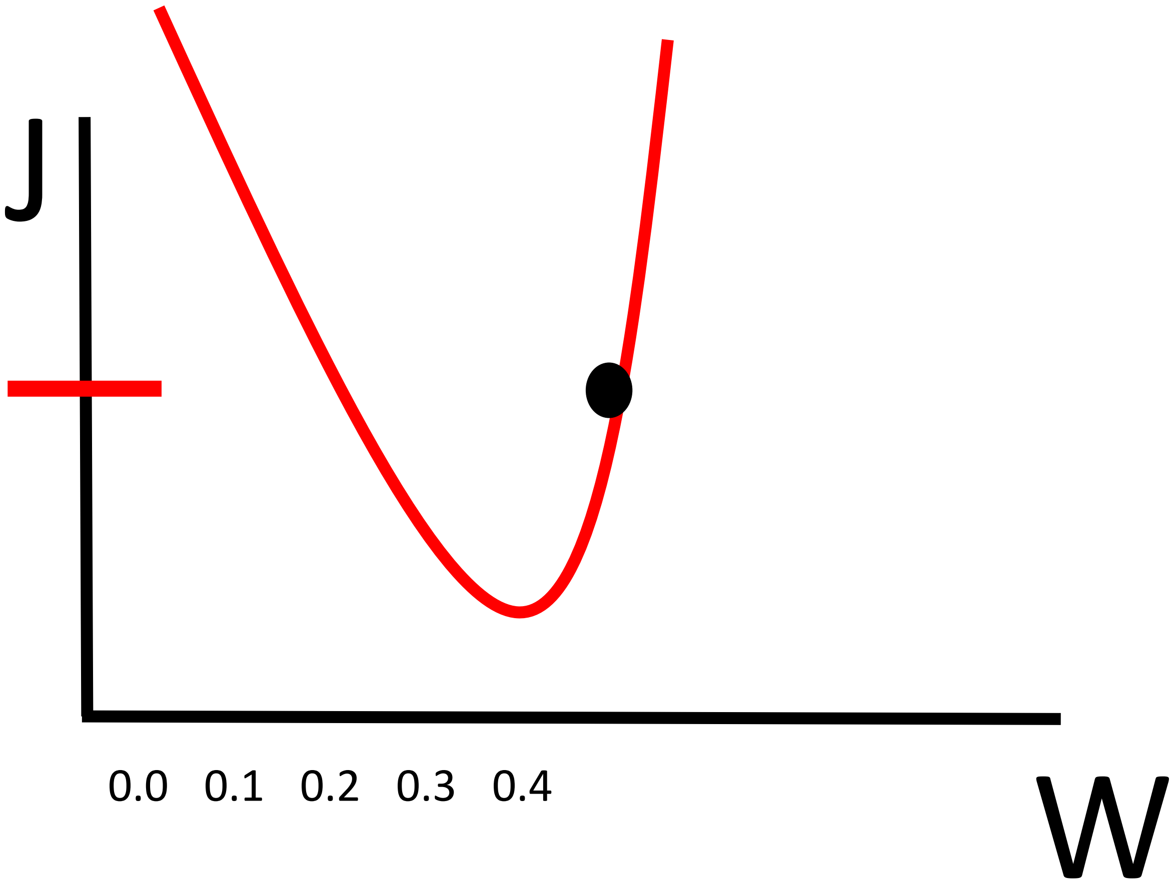
Descenso del Gradiente



Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \cdot \sum (y - y')^2$$

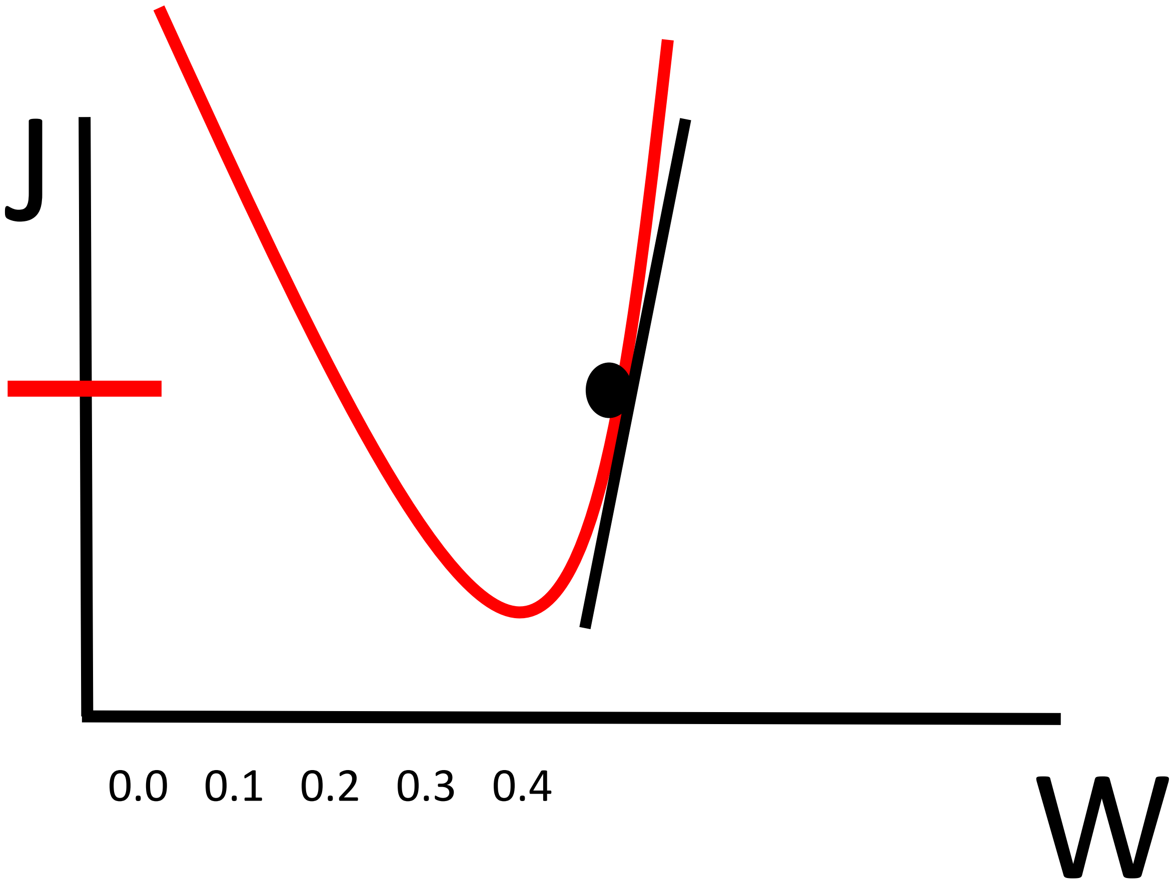
Descenso del Gradiente



Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

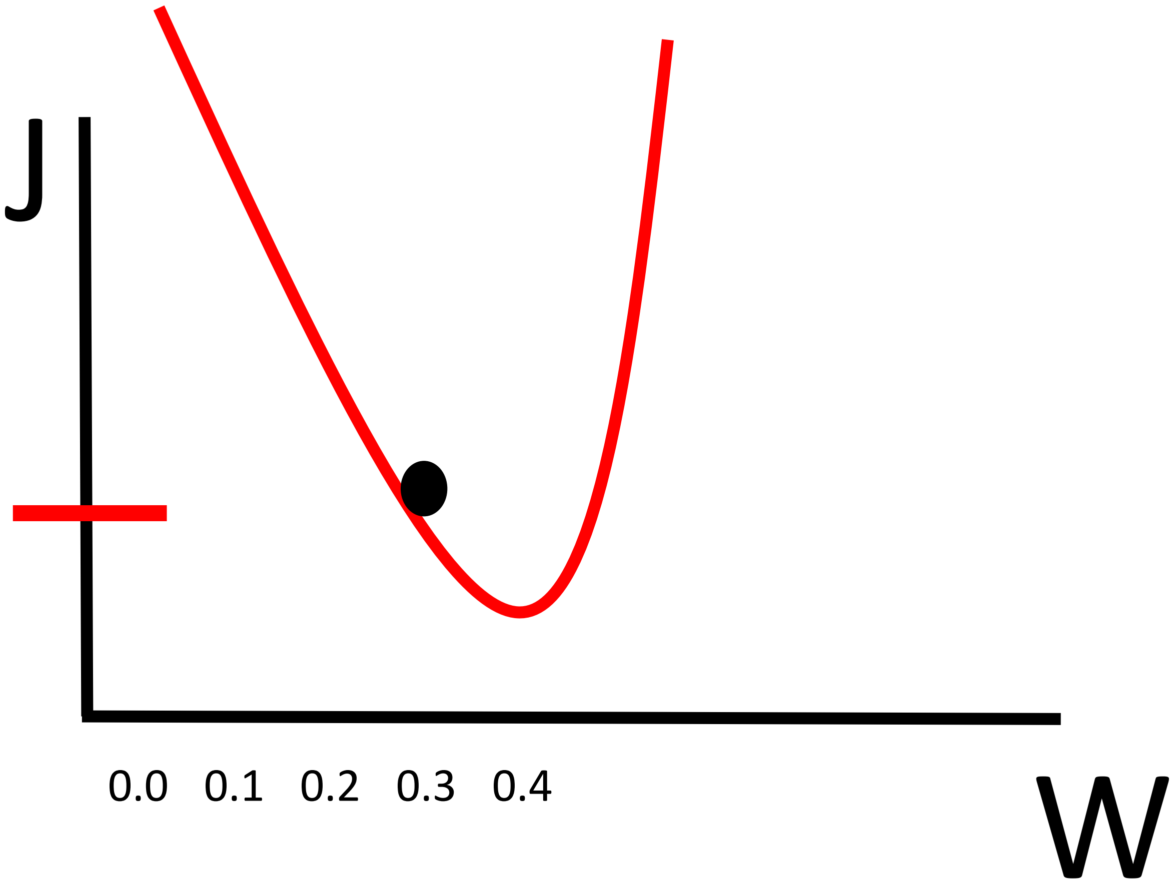
$$J = 1/2 \, . \sum (y - y')^2$$

Descenso del Gradiente



Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \cdot \sum (y - y')^2$$

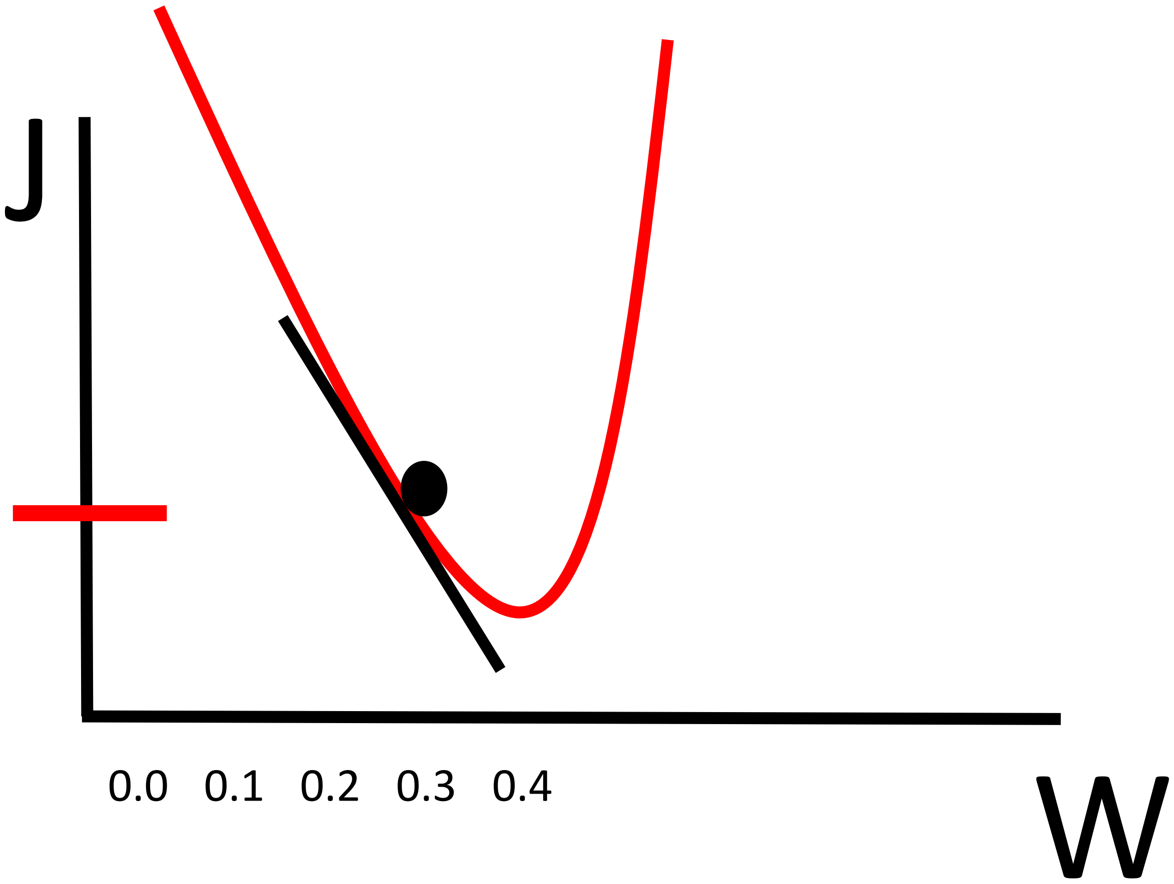


Descenso del Gradiente

Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \,.\, \sum (y - y')^2$$

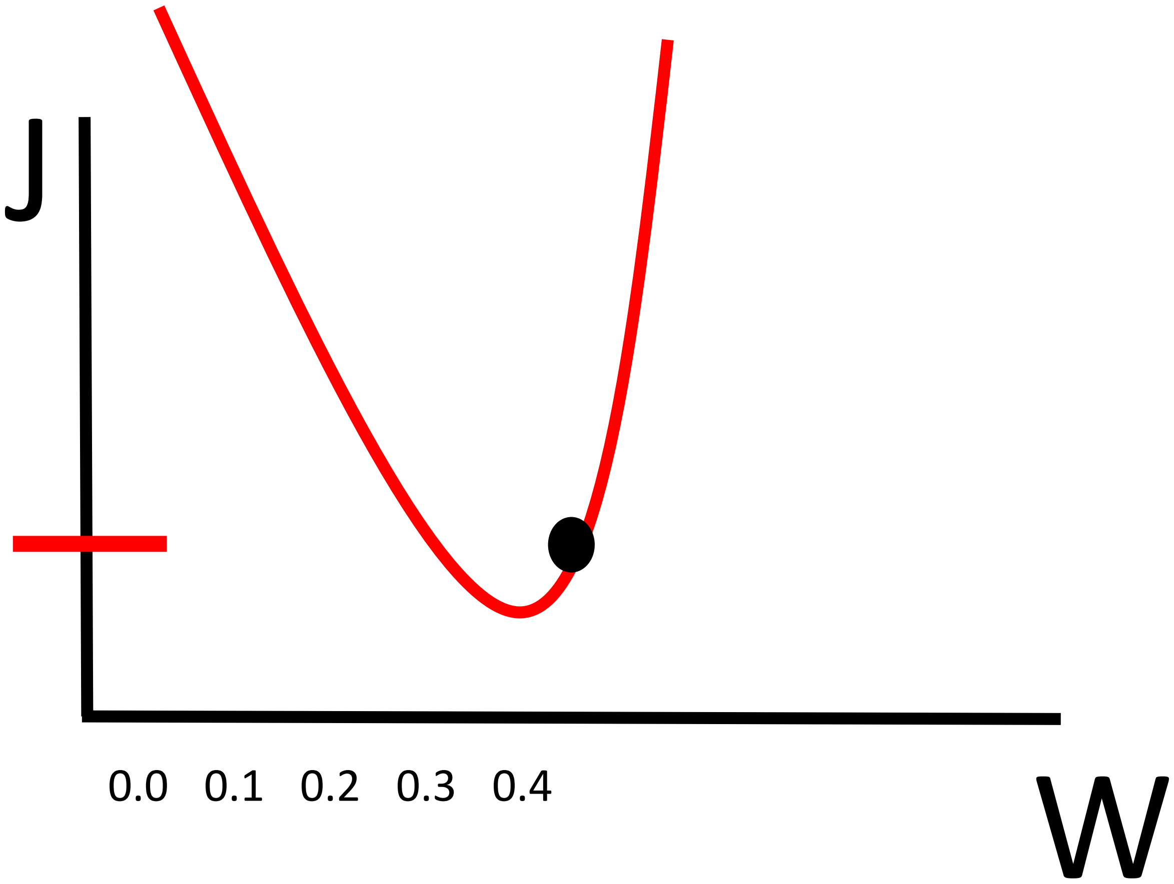
Descenso del Gradiente



Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

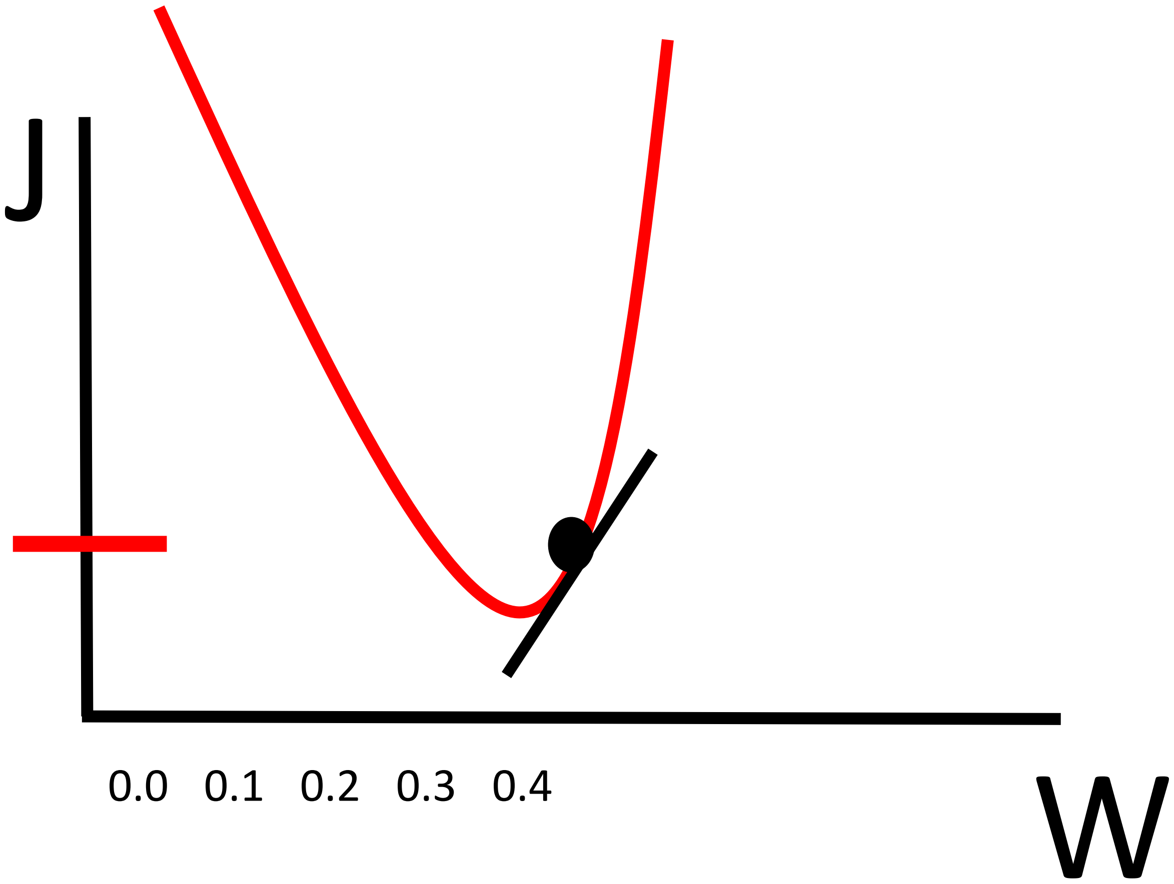
$$J = 1/2 \, . \sum (y - y')^2$$

Descenso del Gradiente



Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \cdot \sum (y - y')^2$$

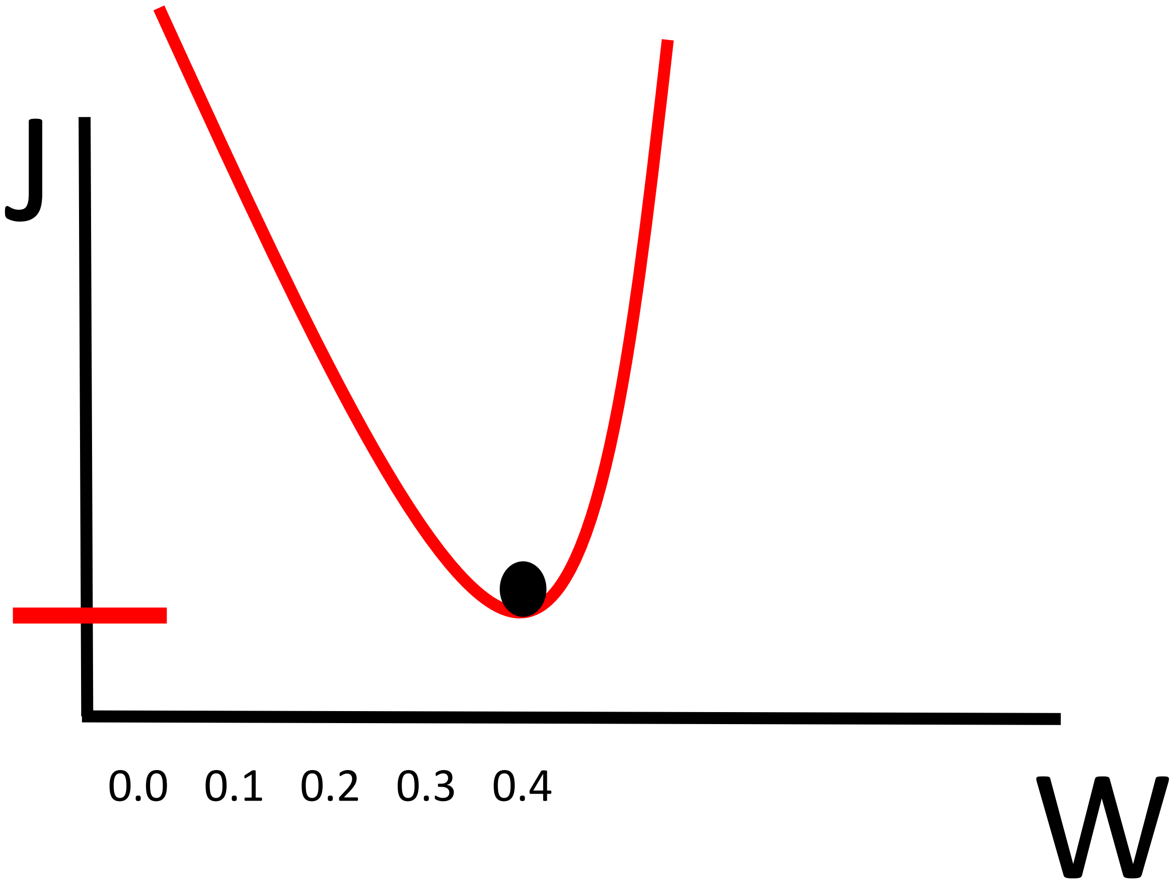


Descenso del Gradiente

Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

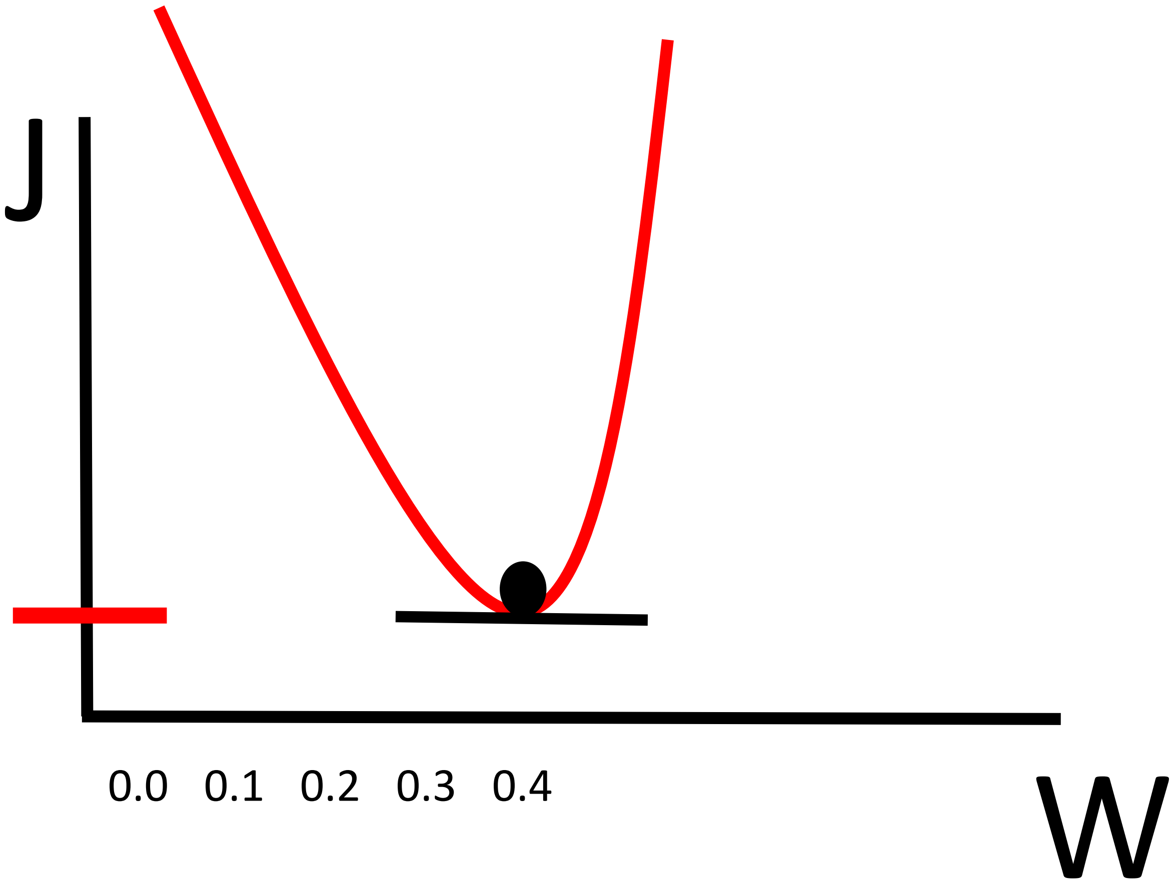
$$J = 1/2 \,.\, \sum (y - y')^2$$

Descenso del Gradiente



Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \, . \sum (y - y')^2$$



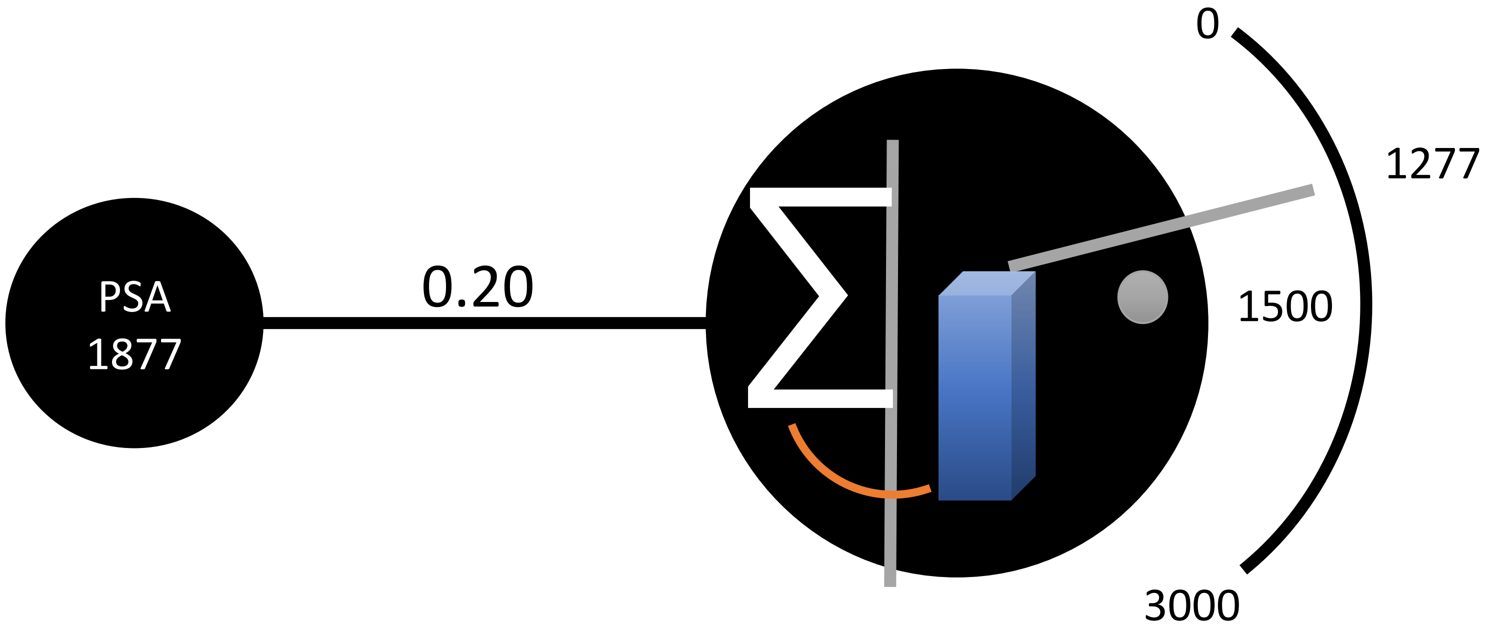
Descenso del Gradiente

Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \,.\, \sum (y - y')^2$$

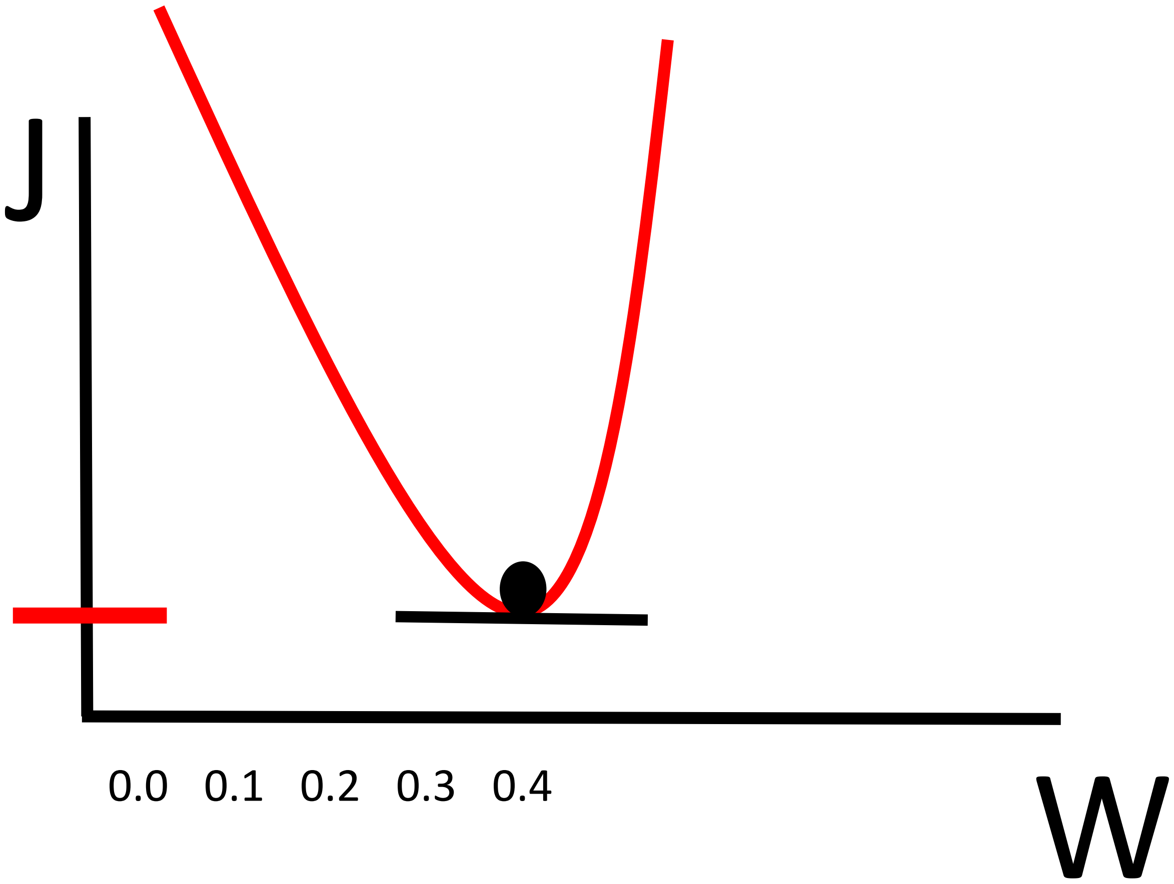
Descenso del Gradiente

Id	Sem Ant	Oro	Tend	P prox Sem
1	1816	1288	1	1290
2	1810	1276	1	1295
3	1860	1290	0	1290
4	1799	1277	0	1280
5	1790	1280	1	1286
6	1877	1270	0	1277



Error Acum: 50

Descenso del Gradiente



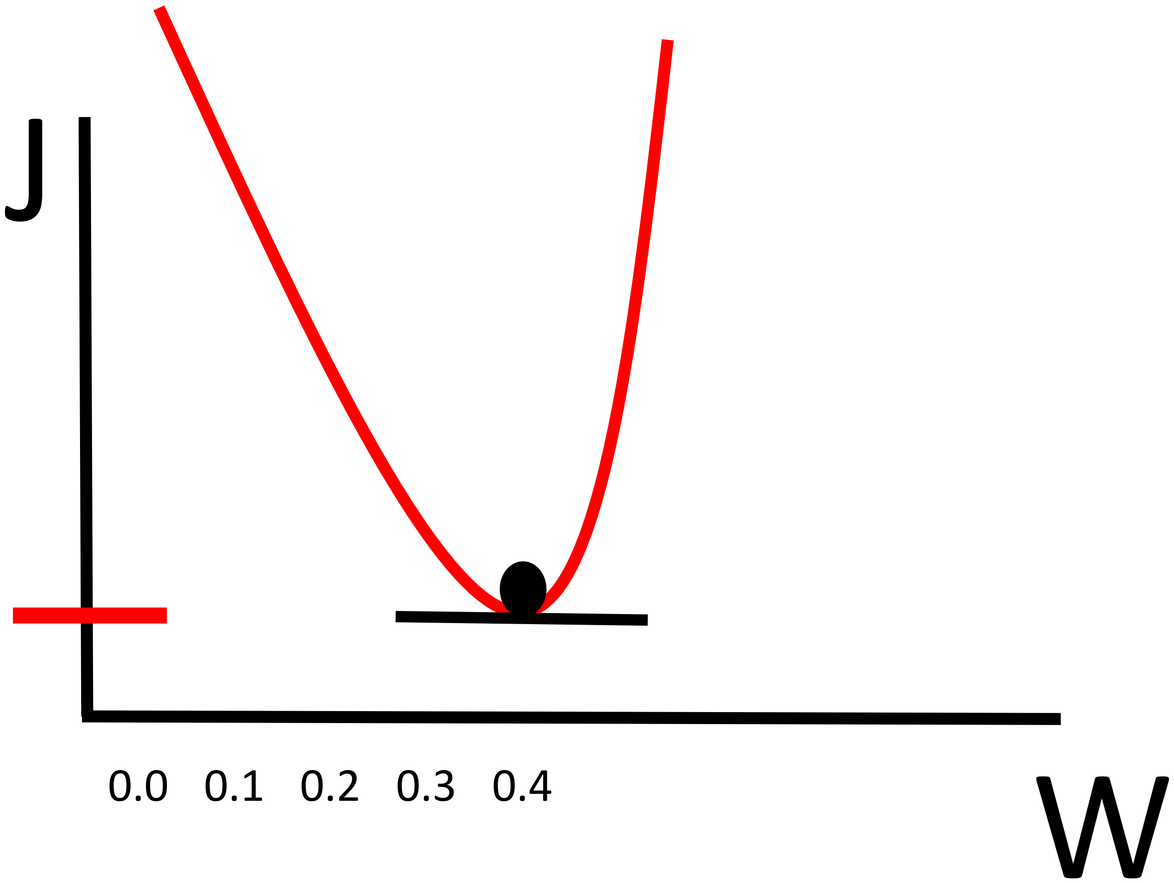
Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = 1/2 \,.\, \sum (y - y')^2$$

Para llegar a este punto, se han realizado 4 “saltos:

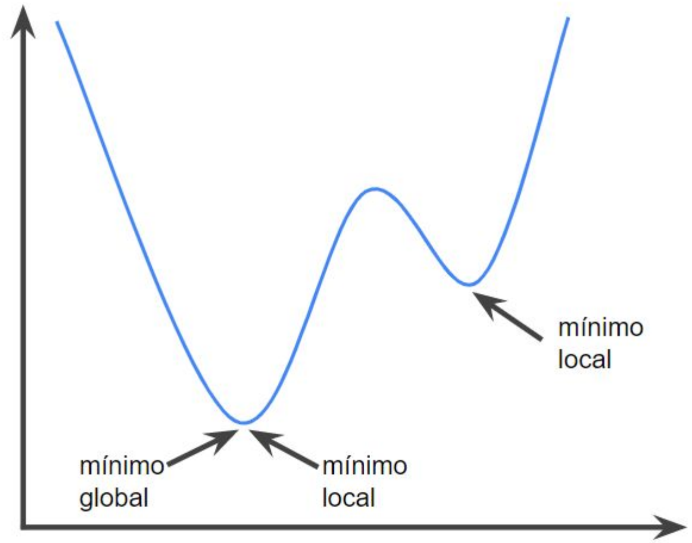
Al tamaño de estos saltos se les conoce como: **Tasa de Aprendizaje (Learn Rate - α)**

α pequeñas = fuerza bruta
α grandes = difícil de hallar el mínimo global

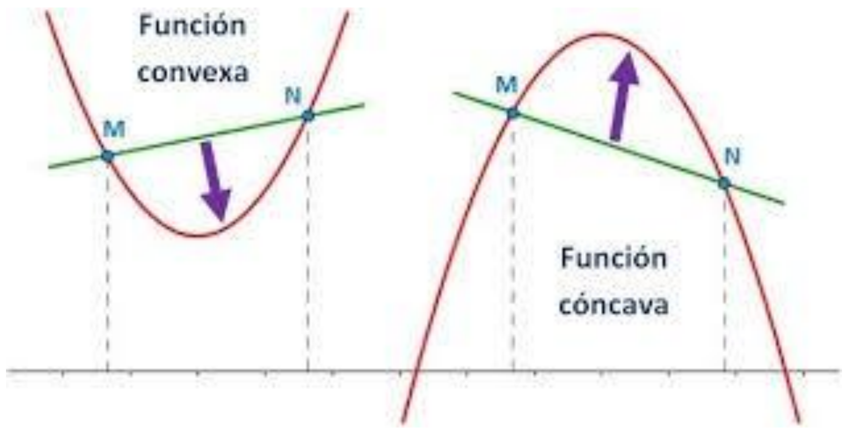


Descenso del Gradiente

- Útil cuando la función de coste es convexa.
- Puede caer en óptimos locales si la función no es convexa.
- La mayoría de las veces se requieren de modelos no lineales, que requieren a su vez funciones de pérdida que son no convexas.



Recordatorios…

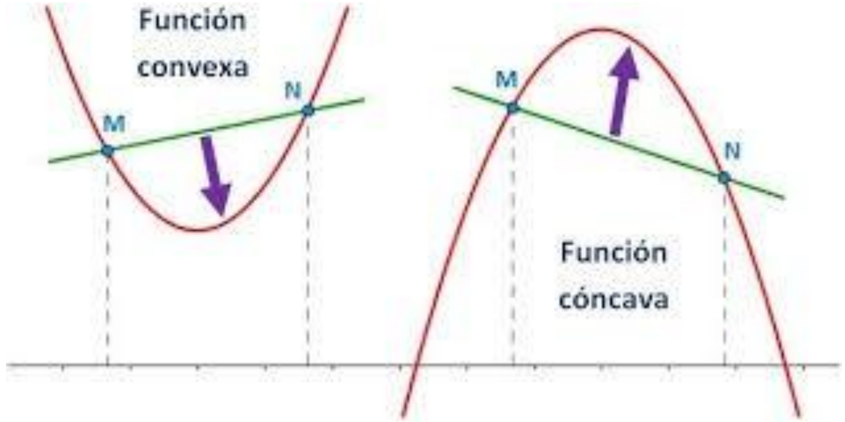


- Una función es **CÓNCAVA** o presenta su concavidad hacia abajo cuando dados dos puntos cualesquiera el segmento que los une queda por debajo de la curva.
- Una función es **CONVEXA** o presenta su concavidad hacia arriba si dados dos puntos de la curva el segmento que los une queda por encima de la curva.
- Los puntos en los que la curvatura pasa de cóncava a convexa o viceversa se llaman **PUNTOS DE INFLEXIÓN**

- **Definición #1:** Una función y=f(x) es CÓNCAVA en un intervalo cuando las tangentes a la curva en los puntos de dicho intervalo quedan por encima de la curva.

- **Definición #2:** Una función y=f(x) será CONVEXA en un intervalo cuando las tangentes a la curva en los puntos de dicho intervalo quedan por debajo de la curva.

Recordatorios…



De una forma más concreta:

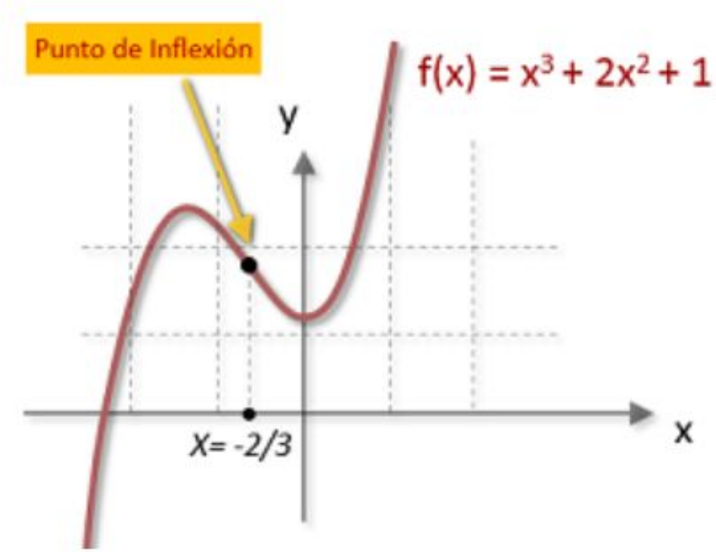
Sea una función f que posea primera y segunda derivada diferentes de cero, es decir:

$$\exists\ f'(x) \neq 0$$
$$\exists\ f''(x) \neq 0$$

Entonces:

- **Definición #1:** Una función y=f(x) es CÓNCAVA si f''(x) < 0
- **Definición #2:** Una función y=f(x) será CONVEXA si f''(x) > 0

Recordatorios…



Ejemplo:

Sea $f(x) = x^3 + 2x^2 + 1$

Tenemos que:

f es cónvaca en $(-\infty, -\frac{2}{3})$ y

f es convexa en $(-\frac{2}{3}, +\infty)$

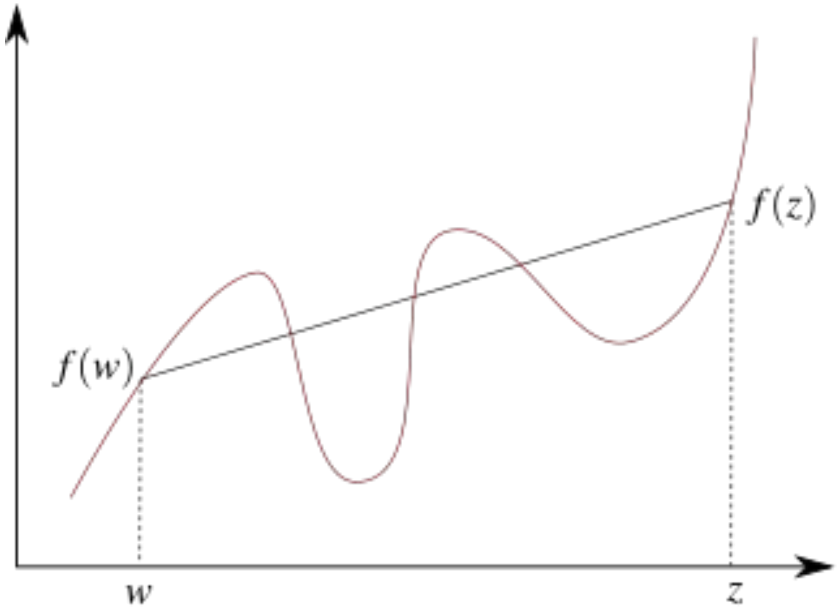
Donde: $-\frac{2}{3}$ es el punto de inflexión.

Esto es porque:

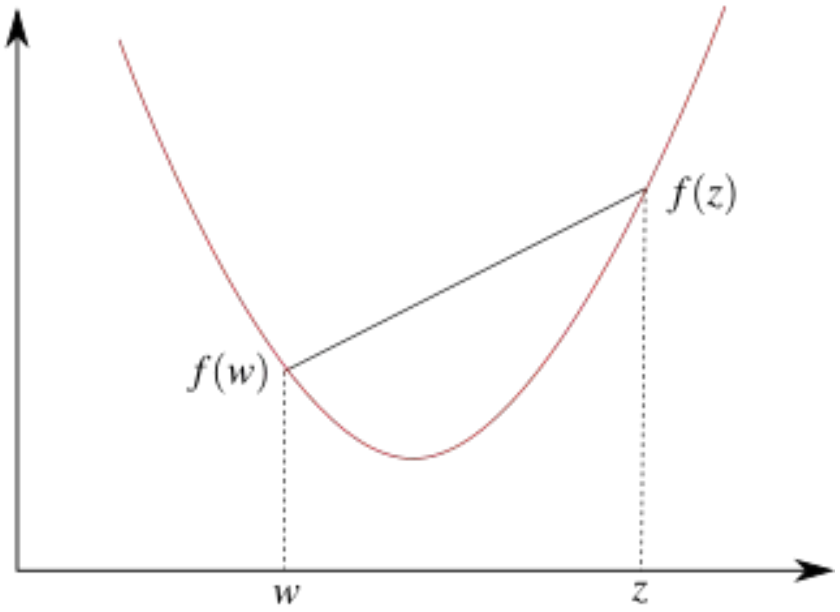
- $f'(x) = 3x^2 + 4x$ y $f''(x) = 6x + 4$ son diferentes de cero.
- $f''(x) = 0 \rightarrow 6x + 4 = 0 \rightarrow x = -\frac{2}{3}$

Recordatorios…

Función no convexa vs Función convexa

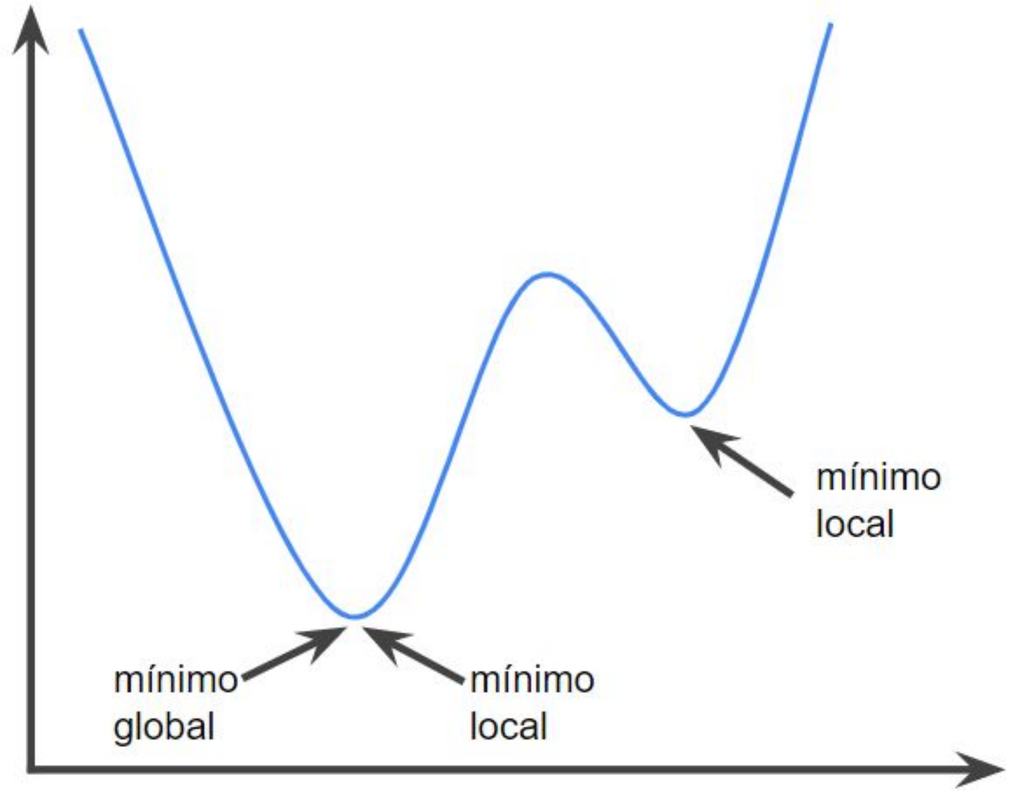


a) Función no convexa



b) Función convexa

Descenso del Gradiente estocástico

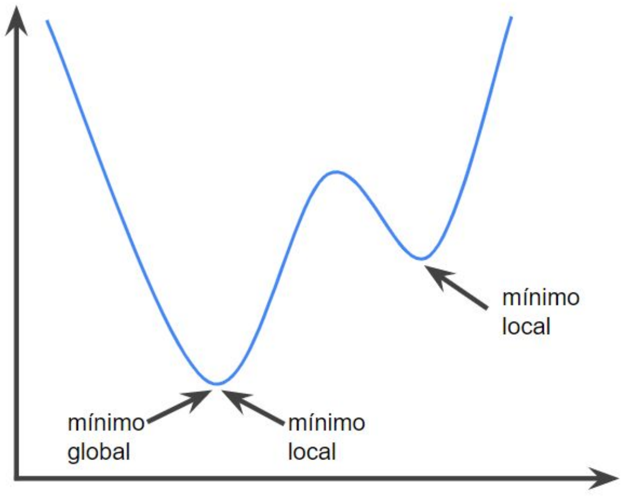


- Útil cuando la función de coste es no convexa
- La estrategia es igual que la anterior, pero ajustando los pesos por cada muestra en lugar de esperar a consumir todas las muestras.

Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

$$J = y - y'$$

Descenso del Gradiente / estocástico



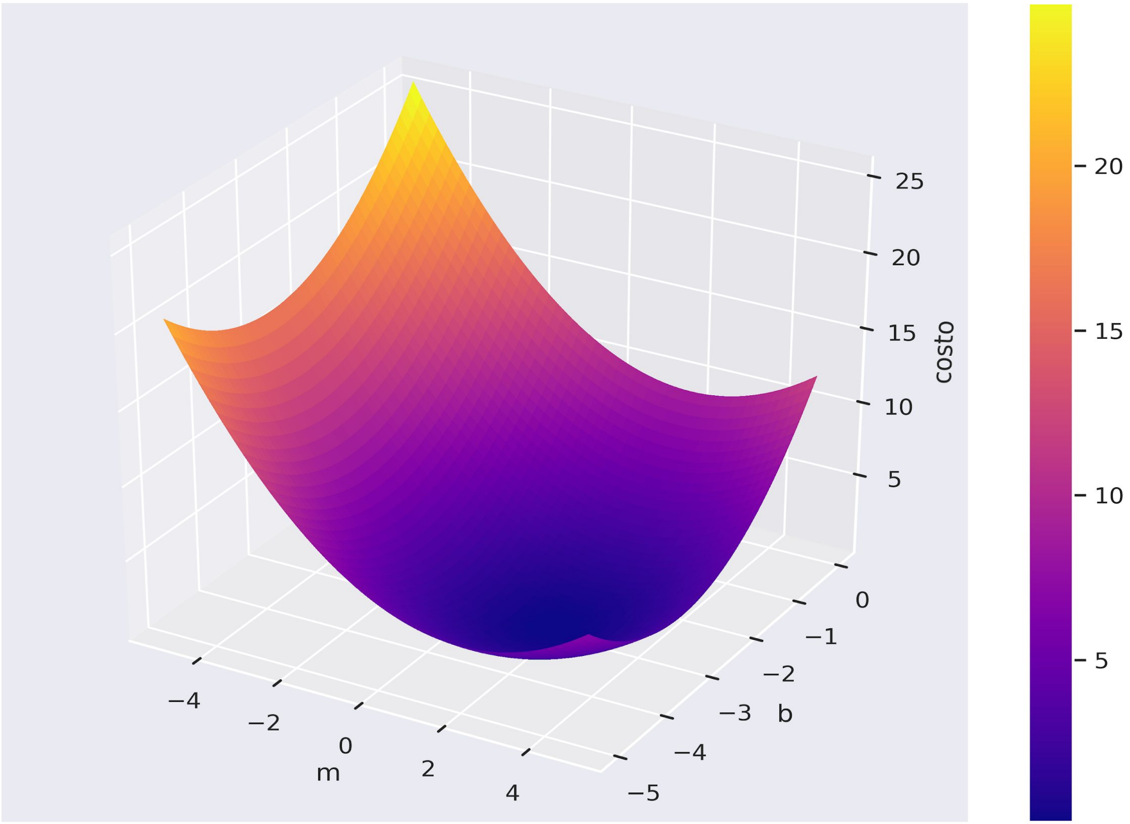
- Ambos métodos presentan ventajas y desventajas
- Para el descenso del gradiente se requiere de suficiente memoria de computador durante el proceso.
- Para el descenso del gradiente estocástico se requiere de un rápido acceso a disco.
- Para lidiar un poco con las ventajas/desventajas de ambos utilizamos una estrategia híbrida conocida como **minibatch**.

Id	Sem Ant	P prox Sem
1	1816	1290
2	1810	1295
3	1860	1290
4	1799	1280
5	1790	1286
6	1877	1277

Una función de coste no convexa

Descenso del Gradiente / estocástico

¡En 2 o más dimensiones es similar!



Funciones de coste

- La función de coste depende del tipo de problema a resolver: regresión o clasificación.
- Una función que suele funcionar muy bien en problemas de clasificación es la entropía cruzada

$$J(p,q) = \sum_{x} p(x)\log(q(x))$$

Función de entropía cruzada:

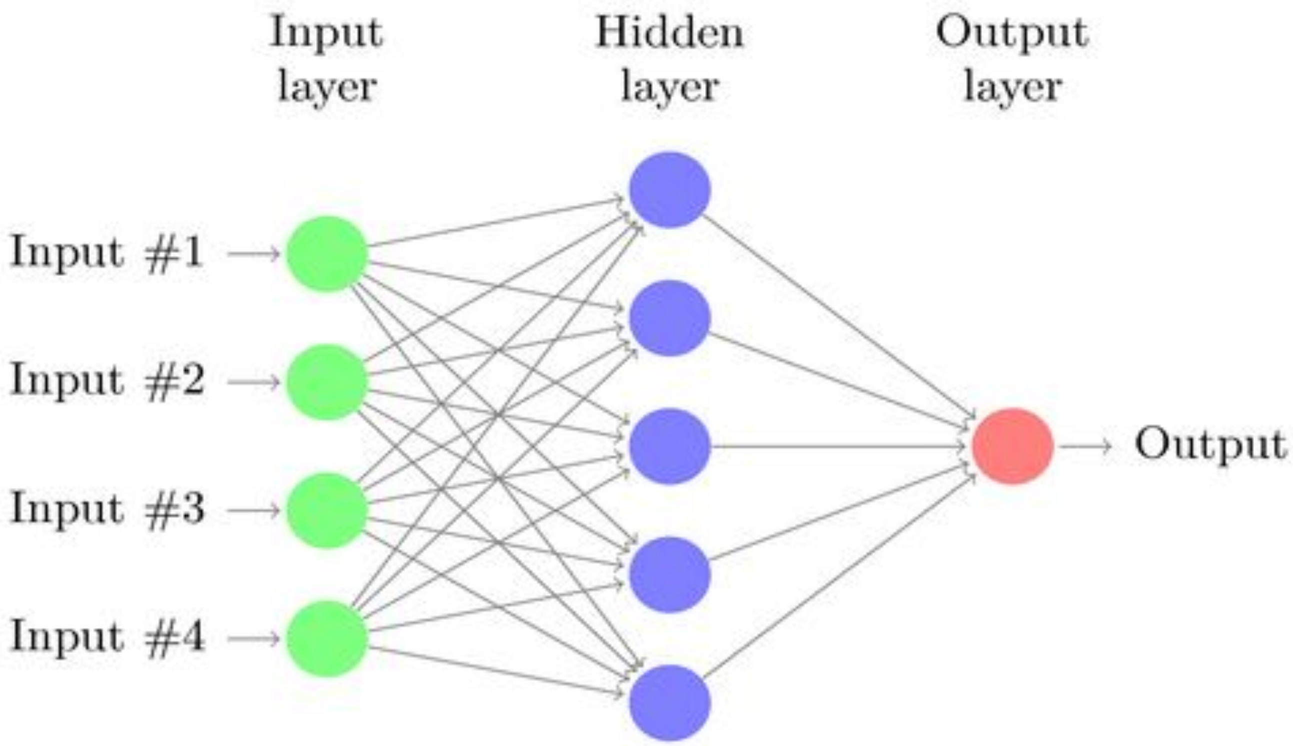
* Cuantifica la diferencia entre dos distribuciones de probabilidad.
* Describe la pérdida entre dos distribuciones de probabilidad.

Retomaremos esto más adelante…

Hasta los momentos…

- Perceptrón
- Foward Propagation
- Algunas funciones de coste
- Descenso del gradiente
- Backpropagation

Redes Neuronales



- Cada neurona tiene sus propios pesos/parámetros. En aplicaciones comunes suelen ser desde miles a millones de parámetros para toda la red.
- Deep Learning es encontrar esos pesos de manera eficiente, bajo la condición de realizar correctamente una tarea objetivo.
- Existen diversas arquitecturas de redes neuronales.
- Se pueden usar para regresión y clasificación.

A continuación…

- Redes neuronales, clasificación
- Funciones de activación
- Métricas y medidas de rendimiento
- Ejemplos avanzados
- Procesamiento de imágenes
- Introducción a las redes neuronales convolucionales