Redes neuronales

Parte 2

Redes neuronales

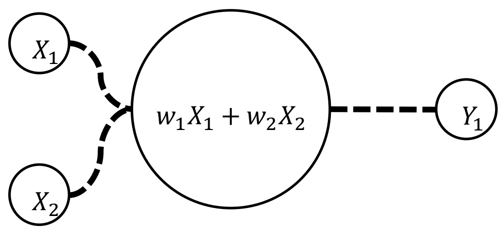
https://playground.tensorflow.org/

Redes neuronales

- Hasta ahora, hemos visto como usar redes neuronales para modelar información.
- Ya hemos visto que las neuronas pueden ser usadas fácilmente para resolver tareas de regresión o clasificación lineal.
- Incluso unimos varias neuronas para formar un red e incluimos las funciones de activación para realizar tareas más complejas.
- El detalle es nosotros debíamos modificar los parámetros de la red.

Redes neuronales

- Entre las décadas de 1950 y 1960 el científico Frank Rosenblatt creó el concepto de Perceptrón.



- Como ocurre con frecuencia, se pensó que el perceptrón ayudaría a resolver todos nuestros problemas.
- Pronto se empezaron a descubrir limitaciones (solo aplica a problemas lineales).

Redes neuronales

- En 1965 se propone el esquema de multi-capas, aparece el concepto de capas de entrada, oculta y salida.
- Lamentablemente, el valor de los pesos de cada neurona se asignaban manualmente.
- Cuantos más perceptrones en las capas, mucho más difícil conseguir los pesos para obtener salidas deseadas.

Redes neuronales

- En 1969, Marvin Minsky y Seymour Papert publican el libro “Perceptrons: an introduction to computational geometry”.
- En este libro demuestran matemáticamente las fortalezas y las limitaciones del peceptrón.
- Una de las limitaciones estaba relacionada con el cálculo de algunos predicados, como la función XOR.

Redes neuronales

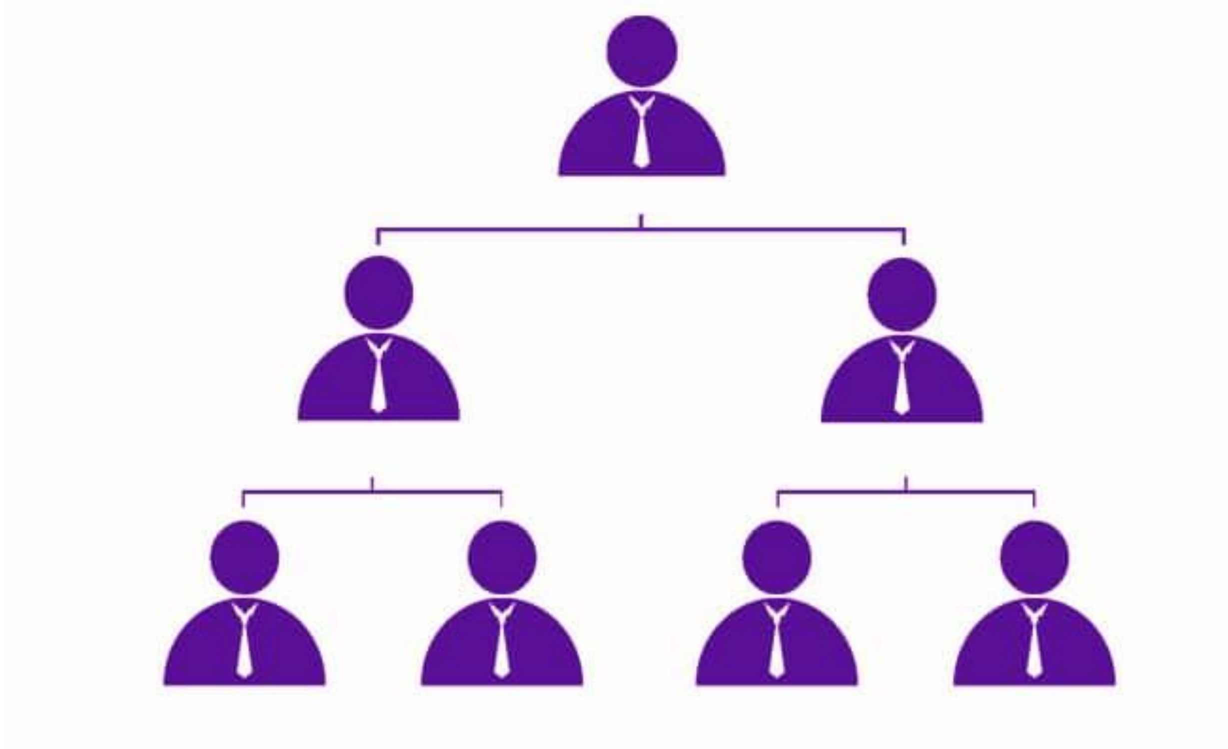
- A partir de su publicación y debate hubo un corte en el financiamiento al estudio de la inteligencia artificial, en particular a los relacionados con redes neuronales.
- Esto dio origen a un periodo de más de 15 años conocido como el invierno de la IA.

Redes neuronales

- Afortunadamente, en 1986, David E. Rumelhart, Geoffrey E. Hinton y Ronald J. Williams, publican el artículo “Learning representations by back-propagating errors”.
- Básicamente, muestran un algoritmo de aprendizaje que le permite a una red neuronal auto-ajustar sus parámetros.
- El nombre del algoritmo fue backpropagation (retro-propagación).

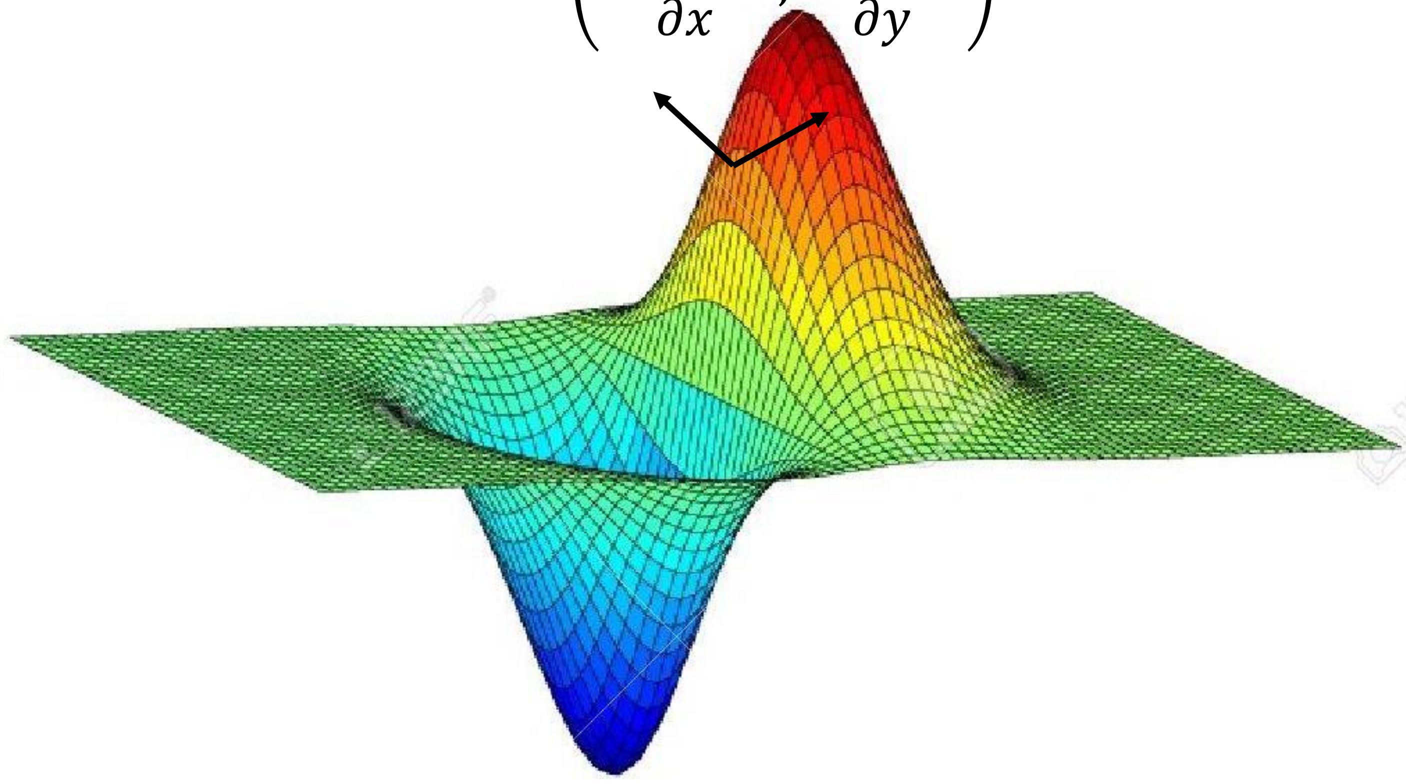
Ejemplo

- Explicar cómo mejorar el proceso productivo en una compañía ideal con una jerarquía bien detallada.



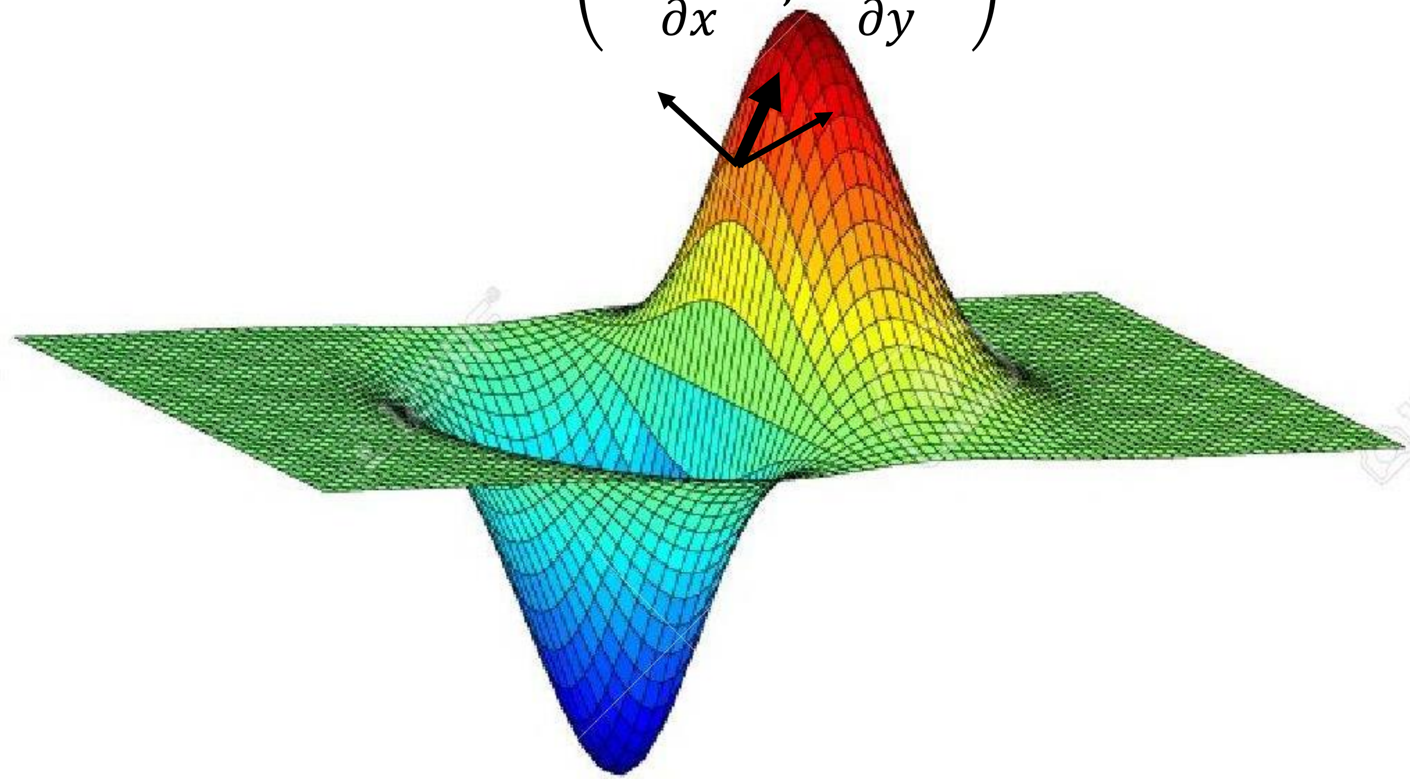
Backpropagation

$$\left(\frac{\partial Error}{\partial x}, \frac{\partial Error}{\partial y}\right)$$



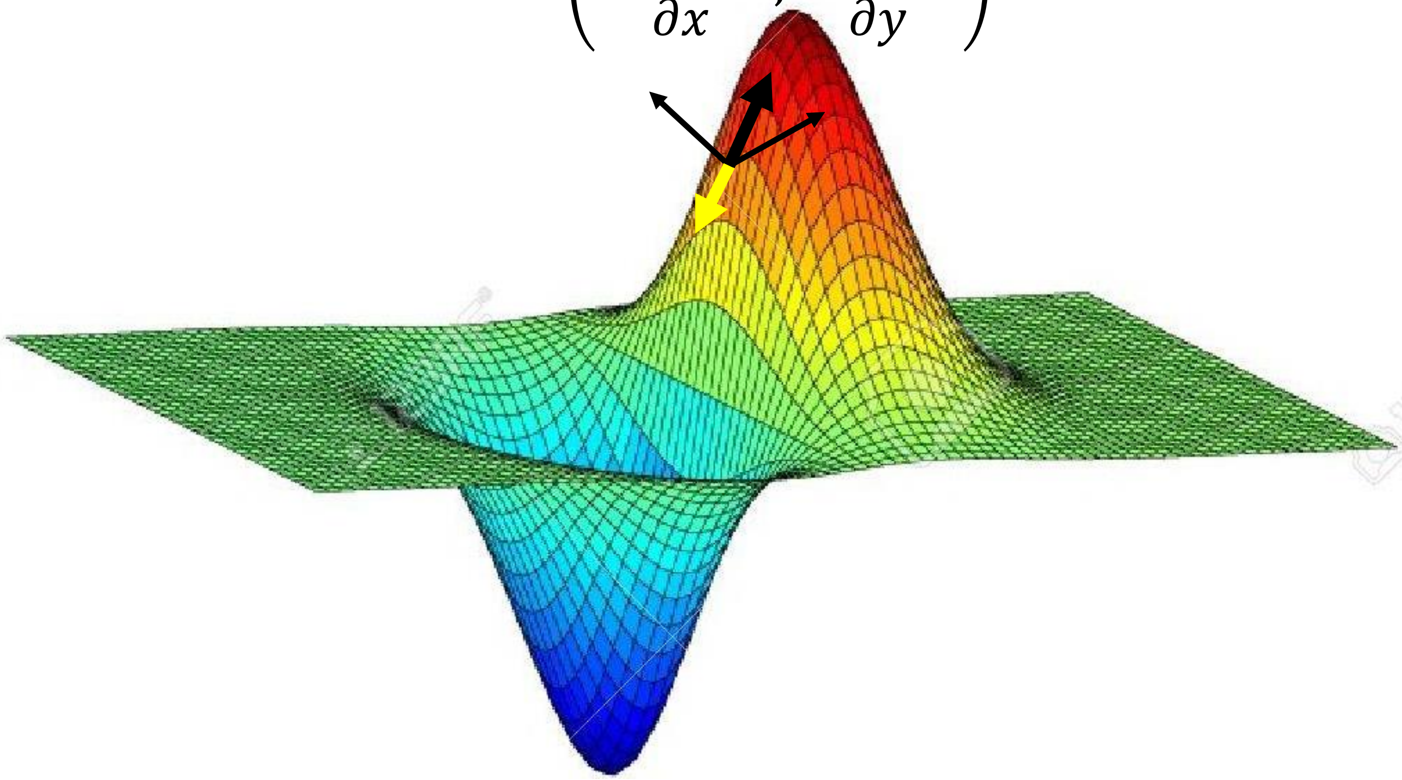
Backpropagation

$$\left(\frac{\partial Error}{\partial x}, \frac{\partial Error}{\partial y}\right)$$



Backpropagation

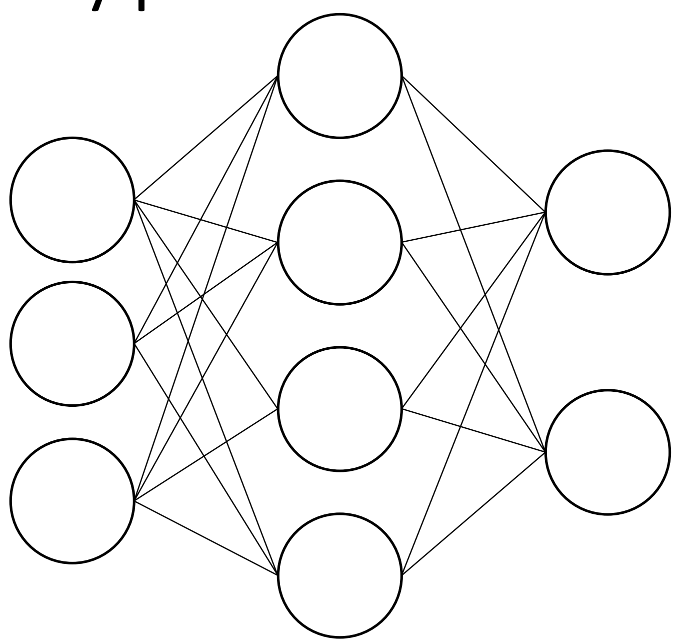
$$\left(\frac{\partial Error}{\partial x}, \frac{\partial Error}{\partial y}\right)$$



Backpropagation

- Considere una red neuronal que inicia con parámetros aleatorios.
- Por lo tanto para los datos de entrada obtendremos salidas aleatorias y posiblemente un error alto

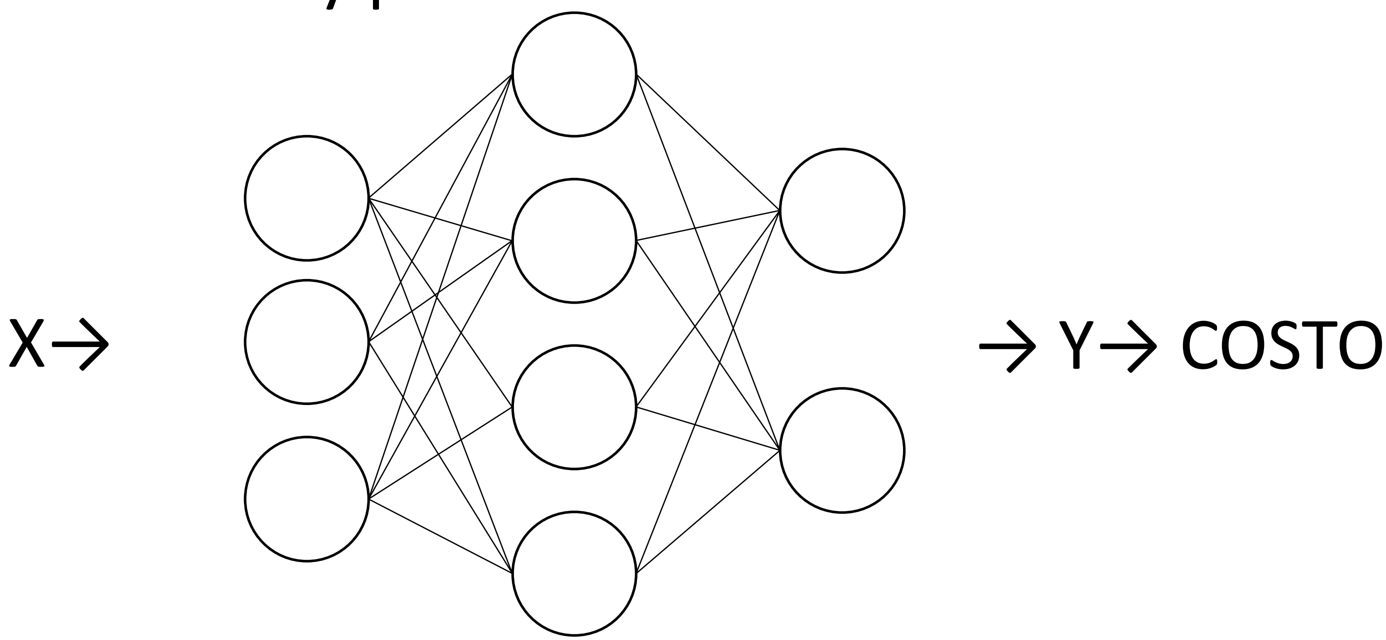
X→



→ Y→ Error

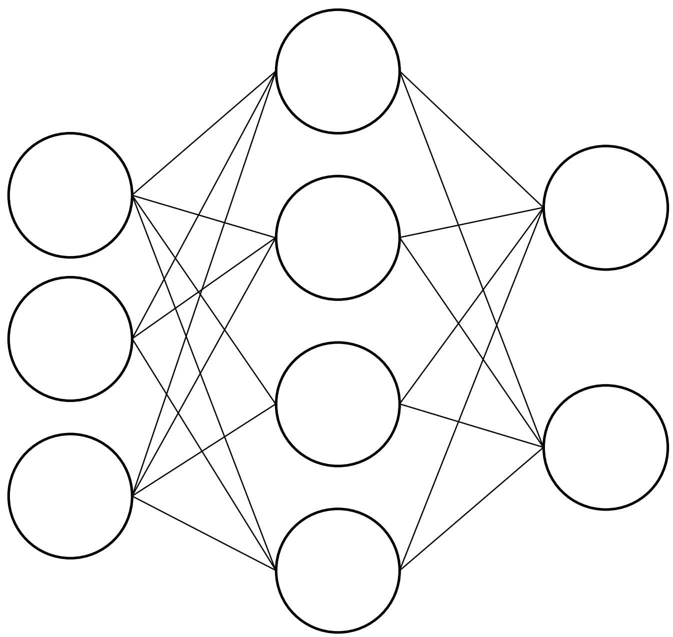
Backpropagation

- Considere una red neuronal que inicia con parámetros aleatorios.
- Por lo tanto para los datos de entrada obtendremos salidas aleatorias y posiblemente un error alto



Backpropagation

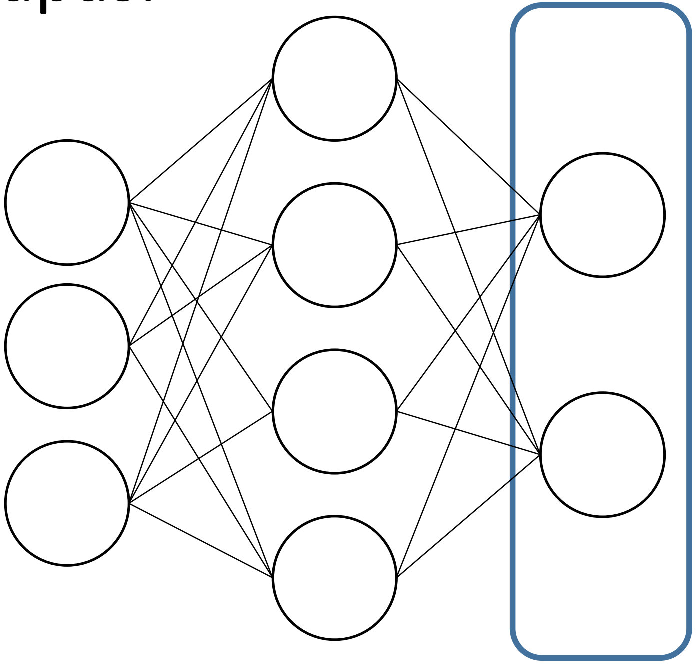
- Ahora, para tratar de mejorar el desempeño de la red, vamos a calcular las derivadas parciales de cada uno de los parámetros.



$$\frac{\partial C}{\partial w} \text{ y } \frac{\partial C}{\partial b}$$

Backpropagation

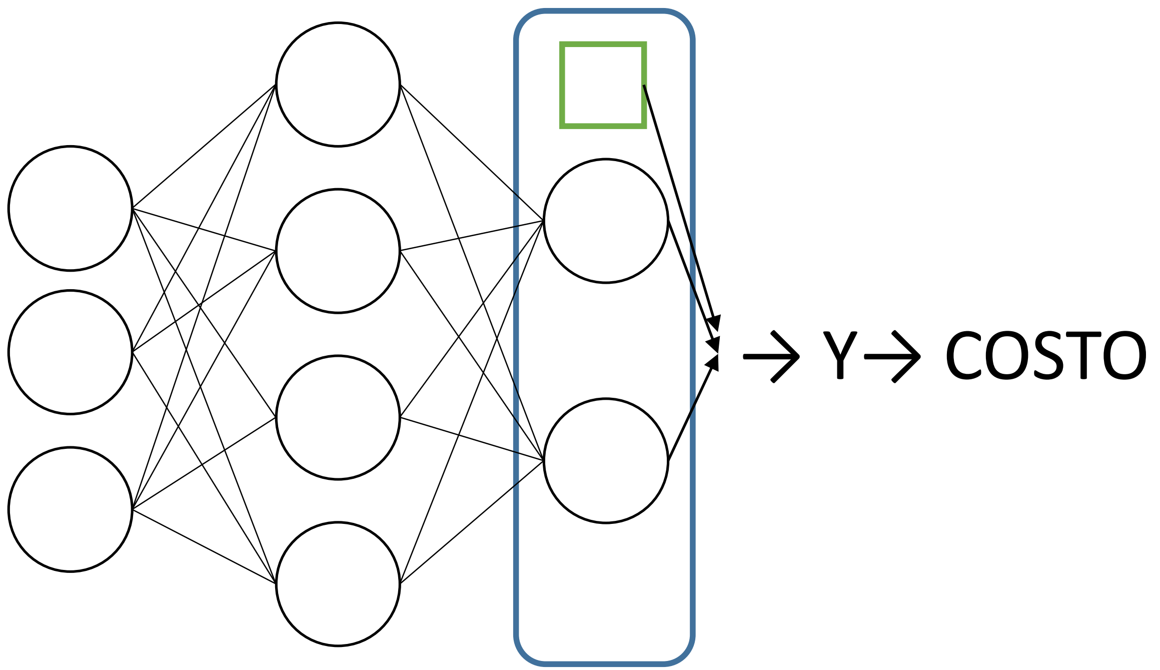
- Debemos empezar desde la última capa.
- Vamos a colocar un identificador para reconoces la capa en la cual estamos trabajando. Supongamos una red con L capas.



$$\frac{\partial C}{\partial w^L} \text{ y } \frac{\partial C}{\partial b^L}$$

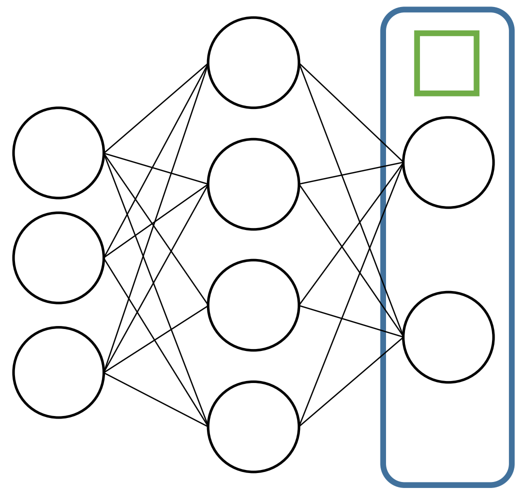
Backpropagation

- Ahora veamos cómo calculamos las derivadas parciales.
- Empecemos con la última capa.



Backpropagation

- Veamos, en la última capa se tendría la transformación lineal de la neurona.
- $Z^L = W^L X + b^L$
- Luego se aplicaría la función de activación
- $a(Z^L) = a(W^L X + b^L)$
- Por último, se calcula el costo o error
- $C\left(a(Z^L)\right) = Error$



→ Y→ COSTO

→ Z^L →$a(Z^L)$ →$C(a(Z^L))$

Backpropagation

- Recordemos que:
- $Z^L = W^L a^{L-1} + b^L; \quad C\left(a^L(Z^L)\right)$
- Ahora podemos aplicar la regla de la cadena y obtener:
- $\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L}\frac{\partial z^L}{\partial w^L}$
- $\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L}\frac{\partial z^L}{\partial b^L}$

Backpropagation

- Ahora podemos aplicar la regla de la cadena y obtener:
- $\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L}\frac{\partial z^L}{\partial w^L}$
- $\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L}\frac{\partial z^L}{\partial b^L}$
- Recordemos que:
- $Z^L = W^L a^{L-1} + b^L; \quad C\left(a^L(Z^L)\right)$
- Si la función de costo es el error cuadrático medio tenemos lo siguiente:
- $C\left(a_j^L\right) = \frac{1}{2}\sum_j\left(y_j - a_j^L\right)^2$
- Entonces,
- $\frac{\partial C}{\partial a^L} = \left(a_j^L - y_j\right)$

Backpropagation

- Ahora podemos aplicar la regla de la cadena y obtener:
- $\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z^L} \frac{\partial z^L}{\partial w^L}$
- $\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z^L} \frac{\partial z^L}{\partial b^L}$
- Recordemos que:
- $Z^L = W^L a^{L-1} + b^L; \quad C\left(a^L(Z^L)\right)$
- Si la función de activación es la sigmoide tenemos lo siguiente:
- $\frac{\partial a^L}{\partial z^L} = a^L(z^L)(1 - a^L z^L)$

Backpropagation

- Ahora podemos aplicar la regla de la cadena y obtener:
- $\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L}\frac{\partial z^L}{\partial w^L}$
- $\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L}\frac{\partial z^L}{\partial b^L}$
- Recordemos que:
- $Z^L = W^L a^{L-1} + b^L; \quad C\left(a^L(Z^L)\right)$
- Por último tenemos:
- $z^L = \sum_i a_i^{L-1} w_i^L + b^L$
- Por lo tanto,
- $\frac{\partial z^L}{\partial b^L} = 1$
- $\frac{\partial z^L}{\partial w^L} = a_i^{L-1}$

Backpropagation

- Ahora podemos aplicar la regla de la cadena y obtener:
- $\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L} a_i^{L-1}$
- $\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L} 1$
- Recordemos que:
- $Z^L = W^L a^{L-1} + b^L; \quad C\left(a^L(Z^L)\right)$

- Ahora veamos el error imputado a cada neurona:
- $\frac{\partial C}{\partial z^L} = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L} = \delta^L$
- Si este valor es grande, significa que cambios en esa neurona afectan claramente el resultado final.
- Podemos re-escribir….

Backpropagation

- Ahora podemos aplicar la regla de la cadena y obtener:
- $\frac{\partial C}{\partial w^L} = \delta^L a_i^{L-1}$
- $\frac{\partial C}{\partial b^L} = \delta^L$
- Recordemos que:
- $Z^L = W^L a^{L-1} + b^L; \quad C\left(a^L(Z^L)\right)$

Backpropagation

- Resumiendo:
- $\delta^L = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L}$
- $\frac{\partial C}{\partial w^L} = \delta^L a_i^{L-1}$
- $\frac{\partial C}{\partial b^L} = \delta^L$

- Con esto tenemos lo necesario para la última capa.
- Vamos por el resto…

Backpropagation

- Para la capa L-1 y las anteriores, debemos considerar:
- $C\left(a^L\left(W^L a^{L-1}\left(W^{L-1} a^{L-2} + b^{L-1}\right) + b^L\right)\right)$
- Aplicando la regla de la cadena
- $\dfrac{\partial C}{\partial w^L} = \boxed{\dfrac{\partial C}{\partial a^L}\dfrac{\partial a^L}{\partial z^L}}\ \boxed{\dfrac{\partial z^L}{\partial a^{L-1}}}\ \boxed{\dfrac{\partial a^{L-1}}{\partial z^{L-1}}}\ \boxed{\dfrac{\partial z^{L-1}}{\partial w^{L-1}}}$ a^{L-1}
- $\dfrac{\partial C}{\partial b^L} = \underbrace{\boxed{\dfrac{\partial C}{\partial a^L}\dfrac{\partial a^L}{\partial z^L}}}_{\delta^L}\ \boxed{\dfrac{\partial z^L}{\partial a^{L-1}}}\ \boxed{\dfrac{\partial a^{L-1}}{\partial z^{L-1}}}\ \boxed{\dfrac{\partial z^{L-1}}{\partial b^{L-1}}}$ 1

Derivada función activación

Backpropagation

- Para la capa L-1 y las anteriores, debemos considerar:
- $C\left(a^L\left(W^L a^{L-1}\left(W^{L-1} a^{L-2} + b^{L-1}\right) + b^L\right)\right)$
- Aplicando la regla de la cadena
- $\frac{\partial C}{\partial w^L} = \underbrace{\frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z^L}}_{\delta^L} \boxed{\frac{\partial z^L}{\partial a^{L-1}}} \boxed{\frac{\partial a^{L-1}}{\partial z^{L-1}}} \boxed{\frac{\partial z^{L-1}}{\partial w^{L-1}}} \leftarrow a^{L-1}$
- $\frac{\partial C}{\partial b^L} = \underbrace{\frac{\partial C}{\partial a^L} \frac{\partial a^L}{\partial z^L}}_{\delta^L} \boxed{\frac{\partial z^L}{\partial a^{L-1}}} \boxed{\frac{\partial a^{L-1}}{\partial z^{L-1}}} \boxed{\frac{\partial z^{L-1}}{\partial b^{L-1}}} \leftarrow \boxed{1}$

W^L

Derivada función activación

$\boxed{Z^L = W^L a^{L-1} + b^L}$

Backpropagation

- Para la capa L-1 y las anteriores, debemos considerar:
- $C\left(a^L\left(W^L a^{L-1}\left(W^{L-1} a^{L-2} + b^{L-1}\right) + b^L\right)\right)$
- Aplicando la regla de la cadena
- $\dfrac{\partial C}{\partial w^L} = \dfrac{\partial C}{\partial a^L}\dfrac{\partial a^L}{\partial z^L}\dfrac{\partial z^L}{\partial a^{L-1}}\dfrac{\partial a^{L-1}}{\partial z^{L-1}}\dfrac{\partial z^{L-1}}{\partial w^{L-1}}$ $\quad a^{L-1}$
- $\dfrac{\partial C}{\partial b^L} = \dfrac{\partial C}{\partial a^L}\dfrac{\partial a^L}{\partial z^L}\dfrac{\partial z^L}{\partial a^{L-1}}\dfrac{\partial a^{L-1}}{\partial z^{L-1}}\dfrac{\partial z^{L-1}}{\partial b^{L-1}}$ $\quad \boxed{1}$

$$\frac{\partial C}{\partial z^{L-1}} = \delta^{L-1}$$

Backpropagation

- Cómputo del error de la última capa

$$\delta^L = \frac{\partial C}{\partial a^L}\frac{\partial a^L}{\partial z^L}$$

- Retropropagación del error a la capa anterior

$$\delta^{l-1} = W^l \delta^l \frac{\partial a^{l-1}}{\partial z^{l-1}}$$

- Calculo de las derivadas de la capa usando en error.

$$\frac{\partial C}{\partial b^{l-1}} = \delta^{l-1} \qquad \frac{\partial C}{\partial w^{l-1}} = \delta^{l-1} a^{l-2}$$