

1 a) No escala con el tamaño de redes neuronales. de propagación y gran cantidad de datos. 10/10

2 Supervisado 10/10

Este tipo de aprendizaje tiene como característica principal el uso de datos que ~~se~~ se encuentran etiquetados para el entrenamiento

No - Supervisado

Este tipo de aprendizaje no necesita ningún tipo de etiquetado en la etapa de entrenamiento, ya que sea partir de los datos va creando sus propias resultados

3 ~~Sen(x)~~  $-x^2$  10/10

Para encontrar un punto crítico debemos derivar la función e igualarla a 0 eso significa que la pendiente de la tangente se encuentra horizontal. La derivada del  $\sin$  es  $\cos(x)$

Buscamos la  $x$  que de como resultado 0.

y la reemplazamos en la función original

~~seno~~  $\cos(90) = 0$  y el  $\sin(90) = 1$ , así

hallamos un máximo local

Para saber si es máximo o mínimo requerimos de la 2da derivada



4  $f(x)$

10/20

Sustituimos  $x$  por  $x_0 + \epsilon g$

$$F(x_0 + \epsilon g) = f(x_0) + (x_0 + \epsilon g - x_0)^T g + \frac{1}{2} ((x_0 + \epsilon g) - x_0)^T H ((x_0 + \epsilon g) - x_0)$$

se simplifica

$$F(x_0) + \epsilon g^T g + \frac{1}{2} \epsilon^2 g^T H g$$

Derivamos <sup>a partir de</sup>  $\epsilon$  y encontramos aquella que da 0

falto un signo  
e indicar que  
esto es la derivada

$$g^T g + \epsilon g^T H g = 0$$

Despejamos  $\epsilon$

$$\epsilon = -\frac{g^T g}{g^T H g}$$

sin el signo correcto  
no puedes llegar a esto

5  $ax^2 + y(ay - bx)$

Derivamos  $\frac{d}{dx} ax^2 + y(ay - bx)$

$$2ax + yb$$

$$\frac{d}{dy} ax^2 + y(ay - bx) = 2ay - bx$$

$$\Delta f = \begin{bmatrix} 2ax + yb \\ 2ay - bx \end{bmatrix} = \begin{bmatrix} .52x - .48b \\ -.48x + .52y \end{bmatrix}$$

15/15

Derivamos

$$H = \begin{bmatrix} 2a & -b \\ -b & 2a \end{bmatrix} = \begin{bmatrix} .52 & -.48 \\ -.48 & .52 \end{bmatrix}$$

20/20



5

Dada  $0.26x^2 + g(0.20 - 0.48x)$

1 Definimos en punto  $x_0$

2 Iteramos  $n$  veces lo siguiente:

15/15

3 Usamos el valor de  $x_0$  para reemplazarla en el gradiente, eso nos da un vector

4 Ese vector lo multiplicamos por una tasa ( $\epsilon$ ), que puede ser definida o una tasa cambiante definida por

$$\epsilon = \frac{g^T}{g^T H g}$$

5 ~~Definir~~ tasa

Al valor  $x_0$  le restamos el gradiente por la tasa, y ese valor es nuestro nuevo  $x_0$

6 Inicia de nuevo la iteración