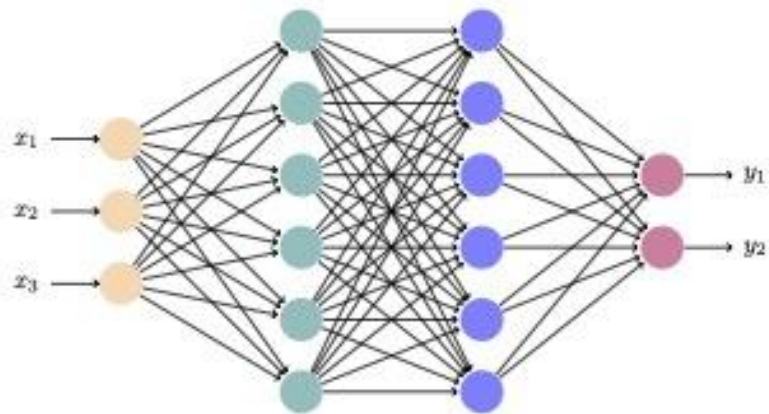


Fundamentos Matemáticos del Aprendizaje Profundo

1er cuat. 2025
Clase 3

Repaso

Una red neuronal codifica una función. Dadas las funciones de activación fijas para cada neurona, la función de salida depende de los pesos y sesgos w , b .



La salida de la red es $y = f_{w, b}(x)$.

Dada la función objetivo $z = \Phi(x)$ cómo elijo (w, b) para “aprenderla”.

Repaso (2)

Función de costo

$$C(w, b) = \text{error}(f_{w, b}(x), \Phi(x))$$

Buscamos $(w_{\text{op}}, b_{\text{op}}) = \text{argmin } C(w, b)$

En general, no conocemos $z = \Phi(x)$, sino “mediciones”, es decir

$$z_i = \Phi(x_i), i=1, \dots, K$$

Repaso (3)

Ejemplos:

- Error cuadrático medio

$$C(w, b) = \sum_{i=1}^K (f_{w,b}(x_i) - z_i)^2$$

- Entropía cruzada

$$C(w, b) = - \sum_{i=1}^K f_{w,b}(x_i) \ln z_i$$

¿Cómo hallar el mínimo de una función?

Propiedades generales de los mínimos

- Si $f(x)$ es una función de una variable real:

$f'(x_{\min}) = 0$ y si $f'(x_0)=0$ y $f''(x_0)>0$, entonces x_0 es un mínimo local de f .

- Si $f=f(x_1,\dots,x_n)$ y $x_{\min} \in K$ es mínimo, $\frac{\partial f}{\partial x_i}(x_{\min}) = 0, \quad i = 1, \dots, n$

Si ahora x_0 es tal que $\nabla f(x_0) = 0$

Usando Taylor,

$$f(x) \simeq f(x_0) + \frac{1}{2}(x - x_0)Hf(x_0)(x - x_0)^t$$

Tenemos los siguientes resultados:

PROPOSICIÓN 4.1.1. *Sea f una función de clase C^2 y $\mathbf{c} \in \mathbb{R}^n$ una solución de $\nabla f(\mathbf{c}) = 0$. Si la matriz Hessiana de f en \mathbf{c} , $Hf(\mathbf{c})$ es definida positiva, entonces \mathbf{c} es un mínimo local de f .*

Si $n=2$, recuperamos el resultado de Análisis 1

1. Si $f_{xx} > 0$ y $\det Hf > 0$, entonces Hf es definida positiva. \rightarrow mínimo
2. Si $f_{xx} < 0$ y $\det Hf > 0$, entonces Hf es definida negativa. \rightarrow máximo
3. Si $\det Hf < 0$, entonces Hf es indefinida. \rightarrow pto. silla

Algoritmos de descenso

Algoritmos de descenso

Los algoritmos de descenso se basan en el siguiente principio. Si tomamos un punto arbitrario del espacio x_1 , típicamente tendremos que $f(x_1) > \min f$.

Queremos entonces ver si podemos “mejorar” x_1 , $x_2 = x_1 + \Delta x$, de manera tal que

$$f(x_1) > f(x_2) > \min f.$$

Idealmente, encontrar Δx de manera de hacerlo de la forma más eficiente.

Descenso \rightarrow desciende el valor de f

Algoritmos de descenso (versión continua)

TEOREMA 4.2.2. *Asumamos que $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es de clase C^2 , que $\mathbf{x}^* \in D$ es un mínimo local estricto de f y que existe un entorno $\mathcal{V} \subset D$ de \mathbf{x}^* tal que $\nabla f(\mathbf{x}) \neq 0$ para $\mathbf{x} \in \mathcal{V} \setminus \{\mathbf{x}^*\}$. Entonces, dado $\mathbf{x}_0 \in \mathcal{V}$, existe una curva $\gamma: [0, 1] \rightarrow \mathbb{R}^n$ tal que*

1. $\gamma(0) = \mathbf{x}_0$;
2. $\gamma(1) = \mathbf{x}^*$;
3. $\gamma'(t)$ es perpendicular a $\mathcal{S}_{f(\gamma(t))}$ para $t \in [0, 1)$.

$$\mathcal{S}_c = f^{-1}(\{c\}) = \{\mathbf{x} \in D: f(\mathbf{x}) = c\}.$$

Idealmente, en implementaciones, necesitamos una versión discreta.