REDES NEURONALES 2024

Clase 16

Lunes 14 de octubre 2024

FAMAF, Universidad Nacional de Córdoba

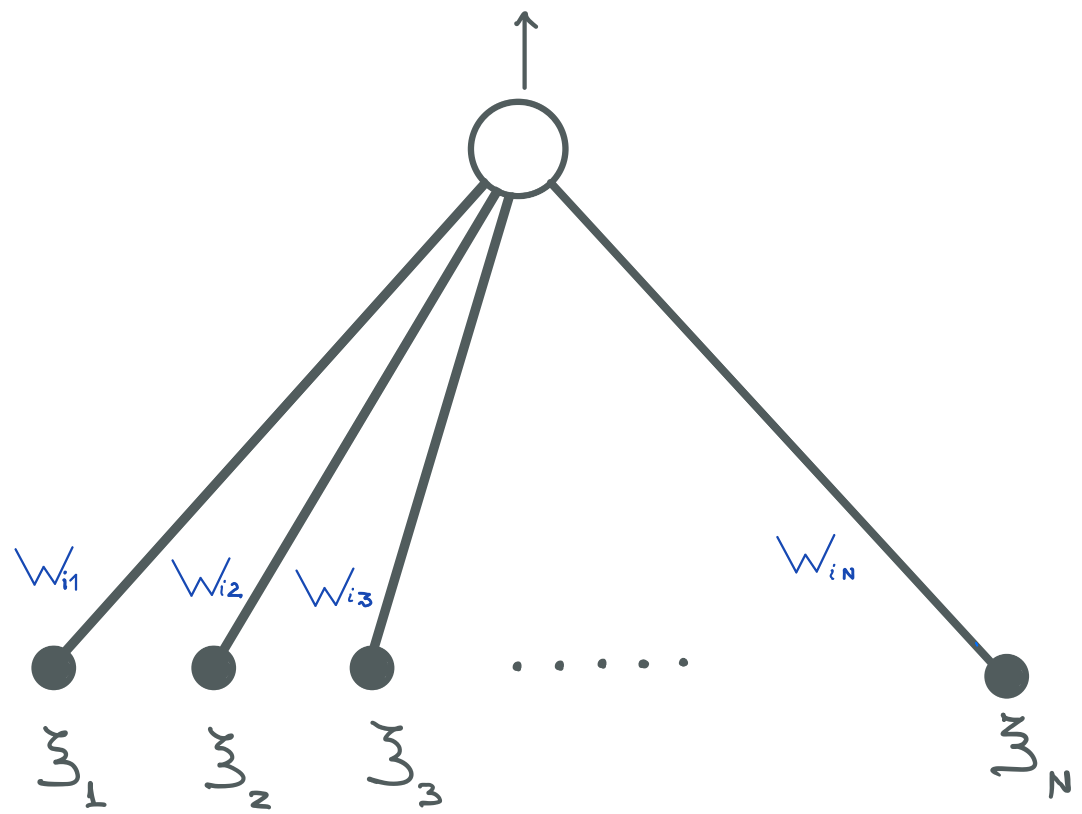
Instituto de Física Enrique Gaviola (UNC-CONICET)

M24/10/23 c18p2

PERCEPTRÓN SIMPLE CON NEURONA LINEAL

Cuando describimos las posibles funciones de activación *g(z)* que definen la salida de la neurona, no contemplamos una función lineal como posibilidad. Sin embargo ahora analizaremos, como casi didáctico, el caso del perceptrón simple con una neurona lineal de salida:

$$O_i = g(h)$$



$$g(z) = z$$

$$O_i = g(h_i) = h_i = \sum_k w_{ik} \xi_k = \overline{W}_i \cdot \overline{\xi}$$

Al igual que antes, suponemos que tenemos *p* entradas previamente etiquetadas:

$$\vec{\xi}^{\alpha} \longmapsto \zeta_i^{\alpha}$$ (valor de salida)

!

$$\zeta_i^{\alpha} = \sum_k w_{ik} \xi_k^{\alpha}$$ (valor deseado)

NOTA: **La salida puede tomar cualquier valor real ahora, por lo cual no estamos haciendo una clasificación sino una REGRESIÓN.**

Existe una solución explícita llamada PSEUDO INVERSA que nos da el valor de las componentes del vector W y tiene la forma:

$$W_{ik} = \frac{1}{N} \sum_{\mu\nu}^{p} \xi_i^{\mu} \left(Q^{-1}\right)_{\mu\nu} \xi_k^{\nu}$$

$$= \frac{1}{N} \left(\xi_i^1, \dots \xi_i^p\right) \cdot Q^{-1} \cdot \begin{pmatrix} \xi_i^1 \\ \vdots \\ \xi_i^p \end{pmatrix}$$

donde

$$Q_{\mu\nu} = \frac{1}{N} \sum_{k=1}^{N} \xi_k^{\mu} \xi_k^{\nu} = \frac{1}{N} \bar{\xi}^{\mu} \cdot \bar{\xi}^{\nu}$$

es la matriz de superposición entre los elementos del conjunto de entrenamiento.

Se puede probar que si usamos esta expresión y la reemplazamos en (1) se cumple la condición requerida.

Para que exista esta solución debe existir $\mathbf{Q}^{-1}$, y para esto los vectores del conjunto de entrenamiento deben ser linealmente independientes. O sea, si existe una relación de la forma:

$$a_1 \xi_k^1 + a_2 \xi_k^2 + \dots + a_p \xi_k^p = 0 \quad \text{para todo } k = 1, 2, \dots, N$$

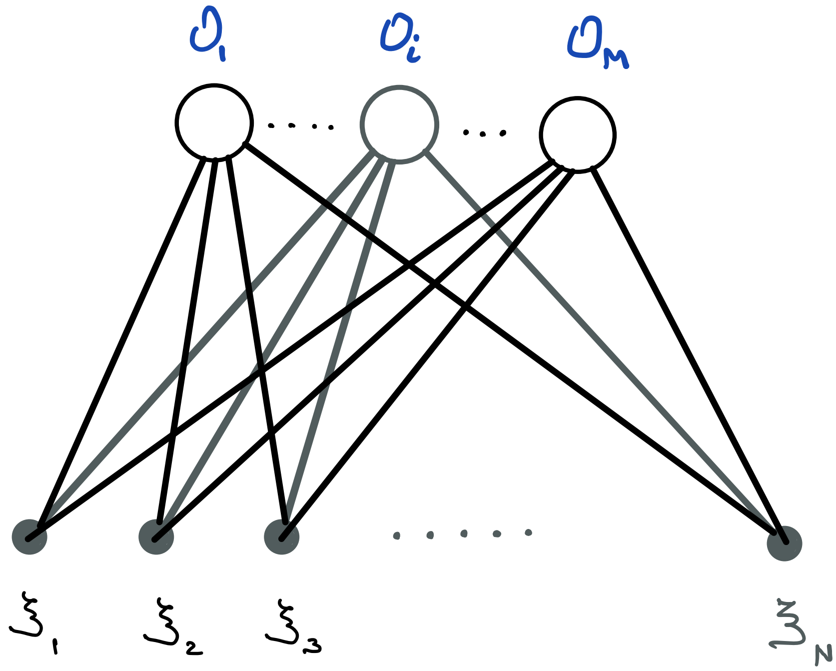
el problema no tiene solución.

El método del descenso por el gradiente

Hoy presentaremos por primera vez en este curso un elemento central en el aprendizaje supervisado moderno, el cual nos acompaña todos los días en nuestras investigaciones y nuestras aplicaciones de IA. Veremos en esta parte la estrecha relación existente entre aprender y optimizar una función, o sea, encontrar sus valores mínimos, locales y globales.

Definir como función error o función costo (loss function, en inglés).

$$E(\bar{w}) = \underbrace{\sum_{\mu=1}^{p}}_{\text{Suma sobre la p elementos del conjunto de entrenamiento}} \left[\underbrace{\sum_{i=1}^{M}}_{\text{Suma sobre la M neuronas de la capa de salida}} \frac{1}{2} \left(\zeta_i^\mu - O_i^\mu \right)^2 \right]$$



Cuando la presentamos la entrada obtenemos el resultado dado por la regla del perceptrón lineal:

$$O_i^\mu = g(h_i^\mu) = h_i^\mu = \sum_{k}^{N} w_{ik}\,\xi_k^\mu = \bar{w}_i \cdot \bar{\xi}^\mu$$

La idea es comenzar con un vector inicial arbitrario W, como hacíamos en el caso de la neurona de salida binaria y presentarle un ejemplo, el μ del conjunto de entrenamiento:

$$E : \mathbb{R}^{N\times M} \longrightarrow \mathbb{R}$$

$$\overrightarrow{\nabla E} : \mathbb{R}^{N\times M} \longrightarrow \mathbb{R}^{N\times M} \qquad \nabla E_{ik} = \frac{\partial E}{\partial w_{ik}}$$