T5 - EJERCICIOS

David Arnal García

2021 - Aprendizaje Automático

Desarrollar formalmente las ecuaciones de actualización de los pesos en el algoritmo BackProp para clasificación (transparencia 5.39).

SOLUCIÓN:

Dado un conjunto de entrenamiento S:

$$S = (x_1, t_1), \cdots, (x_N, t_N), \text{con } x_n \in \mathbb{R}^{M_0},\ t_n \in \{0,1\}^{M_2}, \text{sabiendo que } M_2 \equiv C$$

Ecuaciones para la entropía cruzada:

$$q_S(\mathbf{\Theta}) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{M_2} t_{ni}\ \log s_i^2(\boldsymbol{x}_n; \mathbf{\Theta})$$

$$q_S(\mathbf{\Theta}) = -\frac{1}{N} \sum_{n=1}^{N} q_n(\mathbf{\Theta})$$

$$q_n(\mathbf{\Theta}) = \sum_{i=1}^{M_2} t_{ni}\ \log s_i^2(\boldsymbol{x}_n; \mathbf{\Theta})$$

Aplicamos la técnica del descenso por gradiente:

$$\Delta\theta_{ij}^l = -\rho\ \frac{\partial q_S(\mathbf{\Theta})}{\partial \theta_{ij}^l}$$

Para las condiciones:

$$1 \le l \le 2,\ 1 \le i \le M_l,\ 0 \le j \le M_{l-1}$$

$$\Delta\theta_{ij}^l = -\frac{1}{N} \sum_{n=1}^{N} -\rho \frac{\partial q_n(\mathbf{\Theta})}{\partial \theta_{ij}^l} = -\frac{1}{N} \sum_{n=1}^{N} \Delta_n \theta_{ij}^l$$

Actualización de los pesos de la capa de salida θ_{ij}^2,para una muestra genérica $(\boldsymbol{x}, \boldsymbol{t}) \equiv (\boldsymbol{x}_n, \boldsymbol{t}_n)$

$$q_n(\mathbf{\Theta}) = \sum_{i=1}^{M_2} t_{ni}\ \log s_i^2(\boldsymbol{x}_n; \mathbf{\Theta})$$

$$s_i^2 = g(\phi_i^2)$$

$$\phi_i^2 = \sum_{m=0}^{M_2} \theta_{im}^2 s_m^1$$

Calculamos las derivadas parciales:

$$\frac{\partial q}{\partial \theta_{ij}^l} = \frac{\partial q}{\partial s_i^2} \frac{\partial s_i^2}{\partial \theta_{ij}^2} = \frac{\partial q}{\partial s_i^2} \frac{\partial s_i^2}{\partial \phi_i^2} \frac{\partial \phi_i^2}{\theta_{ij}^2} = t_i \frac{1}{s_i^2} g'(\phi_i^2) s_j^1 = \delta_i^2 s_j^1$$

Por último:

$$\Delta_n \theta_{ij}^l = -\rho\ \delta_i^2\ s_j^1$$

Para las condiciones $1 \le i \le M_2,\ 1 \le i \le M_2,\ 0 \le j \le M_1$

Actualización de los pesos de la capa oculta θ^1_{ij},para una muestra genérica $(\boldsymbol{x}, \boldsymbol{t}) \equiv (\boldsymbol{x}_n, \boldsymbol{t}_n)$

$$q_n(\boldsymbol{\Theta}) = \sum_{i=1}^{M_2} t_{ni} \ \log s^2_i(\boldsymbol{x}_n; \boldsymbol{\Theta})$$

$$s^2_i = g(\phi^2_i)$$

$$\phi^2_i = \sum_{m=0}^{M_1} \theta^2_{im} s^1_m$$

$$s^1_m = g(\phi^1_i)$$

$$\phi^1_i = \sum_{k=0}^{M_0} \theta^1_{mk} x_k$$

Calculamos las derivadas parciales:

$$\frac{\partial q}{\partial \theta^1_{ij}} = \sum_{r=1}^{M_2} \frac{\partial q}{\partial s^2_r} \frac{\partial s^2_r}{\partial \theta^1_{ij}} = \sum_{r=1}^{M_2} \frac{\partial q}{\partial s^2_r} \frac{\partial s^2_r}{\partial \phi^2_r} \frac{\partial \phi^2_r}{\partial s^1_i} \frac{\partial s^1_i}{\partial \phi^1_i} \frac{\partial \phi^1_i}{\theta^1_{ij}} = \sum_{r=1}^{M_2} \delta^2_r \ \theta^2_{ri} \ g'(\phi^1_i) \ x_j =$$

$$= \left(g'(\phi^1_i) \sum_{r=1}^{M_2} \delta^2_r \theta^2_{ri} \right) x_j = \delta^1_i x_j$$

Por último:

$$\Delta_n \theta^l_{ij} = -\rho \ \delta^1_i \ x_j$$

Para las condiciones $1 \leq i \leq M_1, \ 0 \leq j \leq M_0$