T4 EJERCICIOS

Neus Molero

27 de diciembre de 2020

Ejercicio 1

En el formato de las transparencias 5.27 y 5.28 (el número de transparencia puede cambiar (su cabecera es Algoritmo BackProp),

1) Escribir el algoritmo BackProp (batch e incremental) con momentum;

SOLUCIÓN BATCH:

Entrada: Topología, pesos iniciales θ_{ij}^l, $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_l - 1$, momentum ν, $0 \leq \nu < 1$, factor de aprendizaje ρ, condiciones de convergencia, N datos de entrenamiento S
Salidas: Pesos de las conexiones que minimizan el error cuadrático medio de S

Mientras no se cumplan las condiciones de convergencia:

Para $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$, inicializar $\Delta\theta_{ij}^l = 0$

Para cada muestra de entrenamiento $(x, t) \in S$

Desde la capa de entrada a la de salida $(l = 0, \ldots, L)$:
Para $1 \leq i \leq M_l$ si $l == 0$ entonces $s_i^0 = x_i$ sino calcular ϕ_i^l y $s_i^l = g(\phi_i^l)$

Desde la capa de salida a la de entrada $(l = L, \ldots, 1)$,
Para cada nodo $1 \leq i \leq M_l$:

Calcular $\delta_i^l = \begin{cases} g\prime(\phi_i^l)(t_{ni} - s_i^L), & \text{si } l == L \\ g\prime(\phi_i^l)\left(\sum_r \delta_i^{l+1}\delta_{ri}^{l+1}\right), & \text{en otro caso} \end{cases}$

Para cada peso θ_{ij}^l $(0 \leq j \leq M_{l-1})$ calcular $\Delta\theta_{ij}^l = \nu\Delta\theta_{ij}^l + \rho\delta_i^l s_j^{l-1}$

Para $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$ actualizar pesos: $\theta_{ij}^l = \theta_{ij}^l + \frac{1}{N}\Delta\theta_{ij}^l$

SOLUCIÓN INCREMENTAL:

Entrada: Topología, pesos iniciales θ_{ij}^l, $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$, momentum ν, $0 \leq \nu < 1$, factor de aprendizaje ρ, condiciones de convergencia, N datos de entrenamiento S
Salidas: Pesos de las conexiones que minimizan el error cuadrático medio de S

Mientras no se cumplan las condiciones de convergencia:

Para cada muestra de entrenamiento $(x, t) \in S$ (en orden aleatorio)

Desde la capa de entrada a la de salida $(l = 0, \ldots, L)$:
Para $1 \leq i \leq M_l$ si $l == 0$ entonces $s_i^0 = x_i$ sino calcular ϕ_i^l y $s_i^l = g(\phi_i^l)$

Desde la capa de salida a la de entrada $(l = L, \ldots, 1)$,
Para cada nodo $1 \leq i \leq M_l$:

Calcular $\delta_i^l = \begin{cases} g\prime(\phi_i^l)(t_{ni} - s_i^L), & \text{si } l == L \\ g\prime(\phi_i^l)\left(\sum_r \delta_i^{l+1}\delta_{ri}^{l+1}\right), & \text{en otro caso} \end{cases}$

Para cada peso θ^l_{ij} $(0 \leq j \leq M_{l-1})$ calcular $\Delta\theta^l_{ij} = \rho\delta^l_i s^{l-1}_j$

Para $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$ actualizar pesos: $\theta^l_{ij} = \nu\theta^l_{ij} + \frac{1}{N}\Delta\theta^l_{ij}$

2) Escribir el algoritmo BackProp (batch e incremental) con amortiguamiento;

Solución BATCH:

Entrada: Topología, pesos iniciales θ^l_{ij}, $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$, factor de aprendizaje ρ, factor de regularización λ, condiciones de convergencia, N datos de entrenamiento S

Salidas: Pesos de las conexiones que minimizan el error cuadrático medio de S

Mientras no se cumplan las condiciones de convergencia:

Para $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$, inicializar $\Delta\theta^l_{ij} = 0$

Para cada muestra de entrenamiento $(x,t) \in S$

Desde la capa de entrada a la de salida $(l = 0, ..., L)$:
Para $1 \leq i \leq M_l$ si $l == 0$ entonces $s^0_i = x_i$ sino calcular ϕ^l_i y $s^l_i = g(\phi^l_i)$

Desde la capa de salida a la de entrada $(l = L, ..., 1)$,
Para cada nodo $1 \leq i \leq M_l$:

Calcular $\delta^l_i = \begin{cases} \text{g}\prime(\phi^l_i)(t_{ni} - s^L_i), & \text{si } l == L \\ \text{g}\prime(\phi^l_i)\left(\sum_r \delta^{l+1}_i \theta^{l+1}_{ri}\right), & \text{en otro caso} \end{cases}$

Para cada peso θ^l_{ij} $(0 \leq j \leq M_{l-1})$ calcular $\Delta\theta^l_{ij} = \rho\delta^l_i s^{l-1}_j - \rho\lambda\theta^l_{ij}$

Para $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$ actualizar pesos: $\theta^l_{ij} = \theta^l_{ij} + \frac{1}{N}\Delta\theta^l_{ij}$

Solución INCREMENTAL:

Entrada: Topología, pesos iniciales θ^l_{ij}, $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$, factor de aprendizaje ρ, factor de regularización λ, condiciones de convergencia, N datos de entrenamiento S

Salidas: Pesos de las conexiones que minimizan el error cuadrático medio de S

Mientras no se cumplan las condiciones de convergencia:

Para cada muestra de entrenamiento $(x,t) \in S$ (en orden aleatorio)

Desde la capa de entrada a la de salida $(l = 0, ..., L)$:
Para $1 \leq i \leq M_l$ si $l == 0$ entonces $s^0_i = x_i$ sino calcular ϕ^l_i y $s^l_i = g(\phi^l_i)$

Desde la capa de salida a la de entrada $(l = L, ..., 1)$,
Para cada nodo $1 \leq i \leq M_l$:

Calcular $\delta^l_i = \begin{cases} \text{g}\prime(\phi^l_i)(t_{ni} - s^L_i), & \text{si } l == L \\ \text{g}\prime(\phi^l_i)\left(\sum_r \delta^{l+1}_i \theta^{l+1}_{ri}\right), & \text{en otro caso} \end{cases}$

Para cada peso θ^l_{ij} $(0 \leq j \leq M_{l-1})$ calcular $\Delta\theta^l_{ij} = \rho\delta^l_i s^{l-1}_i$

Para $1 \leq l \leq L$, $1 \leq i \leq M_l$, $0 \leq j \leq M_{l-1}$ actualizar pesos: $\theta^l_{ij} = -\rho\lambda\theta^l_{ij} + \frac{1}{N}\Delta\theta^l_{ij}$

Ejercicio 2

Desarrollar formalmente las ecuaciones de actualización de los pesos en el algoritmo BackProp para clasificación (transparencia 6.39).

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:

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

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

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

Aplicamos la técnica del descenso por gradiente:

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

Para las condiciones:

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

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

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

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

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

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

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 $1 \leq i \leq M_2,\ 1 \leq i \leq M_2,\ 0 \leq j \leq M_1$

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

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

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

$$\phi_i^2 = \sum_{m=0}^{M_1} \theta_{im}^2 s_m^1$$

$$s_m^1 = g(\phi_i^1)$$

$$\phi_i^1 = \sum_{k=0}^{M_0} \theta_{mk}^1 x_k$$

Calculamos las derivadas parciales:

$$\frac{\partial q}{\partial \theta_{ij}^1} = \sum_{r=1}^{M_2} \frac{\partial q}{\partial s_r^2} \frac{\partial s_r^2}{\partial \theta_{ij}^1} = \sum_{r=1}^{M_2} \frac{\partial q}{\partial s_r^2} \frac{\partial s_r^2}{\partial \phi_r^2} \frac{\partial \phi_r^2}{\partial s_i^1} \frac{\partial s_i^1}{\partial \phi_i^1} \frac{\partial \phi_i^1}{\theta_{ij}^1} = \sum_{r=1}^{M_2} \delta_r^2 \ \theta_{ri}^2 \ g'(\phi_i^1) \ x_j =$$

$$= \left(g'(\phi_i^1) \sum_{r=1}^{M_2} \delta_r^2 \theta_{ri}^2 \right) x_j = \delta_i^1 x_j$$

Por último:

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

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