TRABAJO TEMA 5 APR

1.- En el formato de las transparencias 5.27 y 5.28,

1) escribir el algoritmo BackProp (batch e incremental) con momentum.

VERSIÓN BATCH

Entrada: topología, pesos iniciales $\theta_{ij}^{l}, 1 \leq l \leq L, 1 \leq i \leq M_l, 0 \leq j \leq M_{l-1}, 0 \leq v < 1$ (v siendo el peso del **momentum**), 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 ($l = 0, \dots, L$):

Para $1 \leq i \leq M_l \begin{cases} s_i^0 = x_i, si\ l == 0 \\ \phi_i^l \text{ y } s_i^L = g(\phi_i^l), en\ otro\ caso \end{cases}$

Desde la capa de salida a la de entrada ($l = L, \dots, 1$):

Para cada nodo ($1 \leq i \leq M_l$)

Calcular $\delta_i^l = \begin{cases} g'(\phi_i^l)(t_{ni} - s_i^L), si\ l == L \\ g'(\phi_i^l)(\sum_r \delta_r^{l+1}\theta_{ri}^{l+1}), en\ otro\ caso \end{cases}$

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

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

VERSIÓN INCREMENTAL

Entrada: topología, pesos iniciales $\theta_{ij}^{l}, 1 \leq l \leq L, 1 \leq i \leq M_l, 0 \leq j \leq M_{l-1}, 0 \leq v < 1$ (v siendo el peso del **momentum**), 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 ($l = 0, \dots, L$):

Para $1 \leq i \leq M_l \begin{cases} s_i^0 = x_i, si\ l == 0 \\ \phi_i^l \text{ y } s_i^L = g(\phi_i^l), en\ otro\ caso \end{cases}$

Desde la capa de salida a la de entrada ($l = L, \dots, 1$):

Para cada nodo $(1 \leq i \leq M_l)$

Calcular $\delta_i^l = \begin{cases} g'(\phi_i^l)(t_{ni} - s_i^L), si\ l == L \\ g'(\phi_i^l)(\sum_r \delta_r^{l+1}\theta_{ri}^{l+1}), en\ otro\ caso \end{cases}$

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

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

2) escribir el algoritmo BackProp (batch e incremental) con amortiguamiento.

VERSIÓN BATCH

Entrada: topología, pesos iniciales $\theta_{ij}^l, 1 \leq l \leq L, 1 \leq i \leq M_l, 0 \leq j \leq M_{l-1}$, factor de **amortiguamiento** λ, 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 $1 \leq i \leq M_l \begin{cases} s_i^0 = x_i, si\ l == 0 \\ \phi_i^l \text{ y } s_i^L = g(\phi_i^l), en\ otro\ caso \end{cases}$

Desde la capa de salida a la de entrada $(l = L, \ldots, 1)$:

Para cada nodo $(1 \leq i \leq M_l)$

Calcular $\delta_i^l = \begin{cases} g'(\phi_i^l)(t_{ni} - s_i^L), si\ l == L \\ g'(\phi_i^l)(\sum_r \delta_r^{l+1}\theta_{ri}^{l+1}), en\ otro\ caso \end{cases}$

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

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

VERSIÓN INCREMENTAL

Entrada: topología, pesos iniciales $\theta_{ij}^l, 1 \leq l \leq L, 1 \leq i \leq M_l, 0 \leq j \leq M_{l-1}$, factor de **amortiguamiento** λ, 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 $(l = 0, \dots, L)$:

Para $1 \leq i \leq M_l \begin{cases} s_i^0 = x_i, si\ l == 0 \\ \phi_i^l \text{ y } s_i^L = g(\phi_i^l), en\ otro\ caso \end{cases}$

Desde la capa de salida a la de entrada $(l = L, \dots, 1)$:

Para cada nodo $(1 \leq i \leq M_l)$

Calcular $\delta_i^l = \begin{cases} g'(\phi_i^l)(t_{ni} - s_i^L), si\ l == L \\ g'(\phi_i^l)(\sum_r \delta_r^{l+1}\theta_{ri}^{l+1}), en\ otro\ caso \end{cases}$

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

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

2.- Desarrollar formalmente las ecuaciones de actualización de los pesos en el algoritmo BackProp para clasificación.

Dado un conjunto de entrenamiento:

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

Entropía cruzada:

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

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

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

Formulación:

$$\Delta\theta_{ij}^l = -\rho\frac{\partial q_S(\Theta)}{\partial\theta_{ij}^l} \quad 1 \leq l \leq 2, 1 \leq i \leq M_l, 0 \leq j \leq M_{l-1}$$

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

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

$$q_n(\Theta) = \sum_{l=1}^{M_2} t_l \log s_l^2\ ; s_l^2 = g(\phi_l^2); \phi_l^2 = \sum_{m=0}^{M_1}\theta_{lm}^2 s_m^1$$

$$\frac{\partial q}{\partial \theta_{ij}^2} = \frac{\partial q}{\partial s_i^2}\frac{\partial s_i^2}{\partial \theta_{ij}^2} = \frac{\partial q}{\partial s_i^2}\frac{\partial s_i^2}{\partial \phi_i^2}\frac{\partial \phi_i^2}{\partial \theta_{ij}^2}$$

$$= t_i \frac{1}{s_i^2} g'(\phi_i^2) s_j^1 = \delta_i^2 s_j^1 \; si \; definimos \; que \; \delta_i^2 = \; t_i \frac{1}{s_i^2} g'(\phi_i^2)$$

$$\Delta_n \theta_{ij}^2 = -\rho \frac{\partial q_n}{\partial \theta_{ij}^2} = -\rho \delta_i^2 s_j^1 \quad 1 \le i \le M_2, 0 \le j \le M_1$$

Finalmente queda:

$$\Delta \theta_{ij}^2 = -\rho \frac{\partial q_S(\Theta)}{\partial \theta_{ij}^2} = \frac{\rho}{N}\sum_{n=1}^{N} \delta_i^2(x_n) s_j^1(x_n) \quad 1 \le i \le M_2, 0 \le j \le M_1$$

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

$$q_n(\Theta) = \sum_{l=1}^{M_2} t_l \log s_l^2 \, ; s_l^2 = g(\phi_l^2); \phi_l^2 = \sum_{m=0}^{M_1} \theta_{lm}^2 \, s_m^1; s_m^1 = g(\phi_m^1); \phi_m^1 = \sum_{k=0}^{M_0} \theta_{mk}^1 \, x_k$$

$$\frac{\partial q}{\partial \theta_{ij}^1} = \sum_{r=1}^{M_2} \frac{\partial q}{\partial s_r^2}\frac{\partial s_r^2}{\partial \theta_{ij}^1} = \sum_{r=1}^{M_2} \frac{\partial q}{\partial s_r^2}\frac{\partial s_r^2}{\partial \phi_r^2}\frac{\partial \phi_r^2}{\partial s_i^1}\frac{\partial s_i^1}{\partial \phi_i^1}\frac{\partial \phi_i^1}{\partial \theta_{ij}^1}$$

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

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

$$\Delta_n \theta_{ij}^1 = -\rho \frac{\partial q_n}{\partial \theta_{ij}^1} = -\rho \delta_i^1 x_j \quad 1 \le i \le M_1, 0 \le j \le M_0$$

Finalmente queda:

$$\Delta \theta_{ij}^1 = -\rho \frac{\partial q_S(\Theta)}{\partial \theta_{ij}^1} = \frac{\rho}{N}\sum_{n=1}^{N} \delta_i^1(x_n) x_{nj} \quad 1 \le i \le M_1, 0 \le j \le M_0$$