Estimación del Gradiente en Perceptrones Multicapa

Juan David Velásquez Henao
jdvelasq@unal.edu.co
Universidad Nacional de Colombia, Sede Medellín
Facultad de Minas
Medellín, Colombia

Haga clic aquí para acceder al repositorio online.

A continuación, se realiza la derivación del gradiente para un perceptrón multicapa.

SUPUESTOS USADOS EN LA DERIVACIÓN DEL GRADIENTE DE LA FUNCIÓN DE ERROR:

1. Se limita un máximo de 3 capas de procesamiento (2 capas ocultas)
2. Ecuaciones para perceptrones más simples (menos capas) se obtienen directamente de la formulación.
3. Los contadores para las capas de procesamiento 0, 1, 2 y 3 son i, j, k y m respectivamente.
4. $z_L[\cdot]$ es la salida de una neurona de la capa L. Si L es la última capa, entonces equivale a la salida del modelo. $z_0[\cdot]$ es la entrada al perceptrón ($x[\cdot]$).
5. $a_L[\cdot]$ es la entrada a una neurona de la capa L.
6. $w_L[\cdot,\cdot]$ son los pesos de la capa $L-1$ hacia la capa L.
7. $b_L[\cdot]$ son los pesos de la neurona bias hacia cada una de las neuronas de procesamiento de la capa L
8. $z_L[\cdot] = \sigma_L(\cdot)$ es la activación de la neurona y $\sigma_L'(\cdot)$ es su derivada para una neurona de la capa L.

ECUACIONES BÁSICAS:

$a_L[\tau] = b_L[\tau] + \sum_\kappa w_L[\kappa,\tau]\, z_{L-1}[\kappa]$; $z_L[\tau] = \sigma(a_L[\tau])$; $\frac{\partial}{\partial a_L[\tau]} z_L[\tau] = \sigma_L'[\tau]$; $\frac{\partial}{\partial w_L[\kappa,\tau]} a_L[\tau] = z_{L-1}[\kappa]$

$\frac{\partial}{\partial b_L[\kappa,\tau]} a_L[\tau] = 1$; $\delta_L[\tau] = e_L[\tau]\, \sigma_L'[\tau]$; $e_{L-1}[\tau] = \sum_\kappa \delta_L[\kappa]\ w_L[\tau,\kappa]$

DERIVACIÓN ANALÍTICA DEL GRADIENTE:

$E()$ es el criterio de error utilizado. La derivación es independiente de la función específica utilizada. Por simplicidad se usará la sumatoria del error cuadrático para el patrón actual:

$$E_p = \frac{1}{2}\sum_{m=1}^{M} e_3[m]^2 = \sum_{m=1}^{M} (d[m] - z_3[m])^2$$

Entonces:

$$\frac{\partial}{\partial z_3[q]} E_p = \sum_{m=1}^{M} \frac{1}{2} \frac{\partial}{\partial z_3[q]} (d[m] - z_3[m])^2 = \sum_{m=1}^{M} e_3[m] \frac{\partial}{\partial z_3[q]} (d[m] - z_3[m])$$

La sumatoria anterior sólo es diferente de cero para $q = m$.

$$\frac{\partial E_p}{\partial z_3[m]} = -e_3[m]$$

Gradiente para la 3ª capa de procesamiento.

$$\frac{\partial}{\partial w_3[k,m]}E_p = \left(\frac{\partial z_3[m]}{\partial z_3[m]}\frac{\partial a_3[m]}{\partial a_3[m]}\right)\frac{\partial E_p}{\partial w_3[k,m]} = \left(\frac{\partial a_3[m]}{\partial w_3[k,m]}\right)\left(\frac{\partial z_3[m]}{\partial a_3[m]}\right)\left(\frac{\partial E_p}{\partial z_3[m]}\right) = (z_2[k])\ (\sigma_3'[m])\ (-\,e_3[m])$$

$$= -\,(e_3[m]\,\sigma_3'[m])\,(z_2[k]) = -\,\delta_3[m]\,z_2[k]$$

$$\frac{\partial}{\partial b_3[m]}E_p = \left(\frac{\partial z_3[m]}{\partial z_3[m]}\frac{\partial a_3[m]}{\partial a_3[m]}\right)\frac{\partial E_p}{\partial b_3[m]} = \left(\frac{\partial a_3[m]}{\partial b_3[m]}\right)\left(\frac{\partial z_3[m]}{\partial a_3[m]}\right)\left(\frac{\partial E_p}{\partial z_3[m]}\right) = (1)\ (\sigma_3'[m])\ (-\,e_3[m]) = -\ e_3[m]\,\sigma_3'[m] = -\delta_3[m]$$

Gradiente para la 2ª capa de procesamiento (note la estructura recursiva de las ecuaciones).

$$\frac{\partial}{\partial w_2[j,k]}E_p = \sum_{m=1}^{M}\left(\frac{\partial z_3[m]}{\partial z_3[m]}\frac{\partial a_3[m]}{\partial a_3[m]}\right)\left(\frac{\partial z_2[k]}{\partial z_2[k]}\frac{\partial a_2[k]}{\partial a_2[k]}\right)\frac{\partial E_p}{\partial w_2[j,k]} = \sum_{m=1}^{M}\left(\frac{\partial z_3[m]}{\partial a_3[m]}\right)\left(\frac{\partial a_3[m]}{\partial z_2[k]}\right)\left(\frac{\partial z_2[k]}{\partial a_2[k]}\right)\left(\frac{\partial a_2[k]}{\partial w_2[j,k]}\right)\left(\frac{\partial E_p}{\partial z_3[m]}\right)$$

$$= \sum_{m=1}^{M}(\sigma_3'[m])\ (w_3[k,m])\ (\sigma_2'[k])\ (z_1[j])\ (-\,e_3[m]) = -\sigma_2'[k]\ z_1[j]\ \sum_{m=1}^{M}(e_3[m]\,\sigma_3'[m])\ w_3[k,m]$$

$$= -\,\sigma_2'[k]\ z_1[j]\ \sum_{m=1}^{M}\delta_3[m]w_3[k,m] = -\,\sigma_2'[k]\ z_1[j]\ e_2[k] = -\,\delta_2[k]\ z_1[j]$$

$$\frac{\partial}{\partial b_2[k]}E_p = -\,\delta_2[k]$$

Gradiente para la 1ª capa de procesamiento

$$\frac{\partial}{\partial w_1[i,j]}E_p = \sum_k\sum_m\left(\frac{\partial z_3[m]}{\partial z_3[m]}\frac{\partial a_3[m]}{\partial a_3[m]}\right)\left(\frac{\partial z_2[k]}{\partial z_2[k]}\frac{\partial a_2[k]}{\partial a_2[k]}\right)\left(\frac{\partial z_1[j]}{\partial z_1[j]}\frac{\partial a_1[j]}{\partial a_1[j]}\right)\frac{\partial E_p}{\partial w_1[i,j]}$$

$$= \sum_k\sum_m\left(\frac{\partial z_3[m]}{\partial a_3[m]}\right)\left(\frac{\partial a_3[m]}{\partial z_2[k]}\right)\left(\frac{\partial z_2[k]}{\partial a_2[k]}\right)\left(\frac{\partial a_2[k]}{\partial z_1[j]}\right)\left(\frac{\partial z_1[j]}{\partial a_1[j]}\right)\left(\frac{\partial a_1[j]}{\partial w_1[i,j]}\right)\left(\frac{\partial E_p}{\partial z_3[m]}\right)$$

$$= \sum_k\sum_m(\sigma_3'[m])\,(w_3[k,m])\,(\sigma_2'[k])\,(w_2[j,k])\,(\sigma_1'[j])\,(z_0[i])\,(-\,e_3[m])$$

$$= -\sum_k\sum_m(e_3[m]\sigma_3'[m])\,(w_3[k,m])\,(\sigma_2'[k])\,(w_2[j,k])\,(\sigma_1'[j])\,(z_0[i])$$

$$= -\sum_k(\sigma_2'[k])\,(w_2[j,k])\,(\sigma_1'[j])\,(z_0[i])\sum_m(\delta_3[m])\,(w_3[k,m]) = -\sum_k(\sigma_2'[k])\,(w_2[j,k])\,(\sigma_1'[j])\,(z_0[i])\,e_2[k]$$

$$= -\sum_k(e_2[k]\,\sigma_2'[k])\,(w_2[j,k])\,(\sigma_1'[j])\,(z_0[i]) = -\sigma_1'[j]\,z_0[i]\sum_k\ \delta_2[k]\,w_2[j,k] = -\sigma_1'[j]\,z_0[i]\,e_1[j] = -\,\delta_1[j]\,z_0[i]$$

$$\frac{\partial}{\partial b_1[j]}E_p = -\,\delta_1[j]$$

Para un perceptrón de una sola capa de procesamiento (monocapa) con k entradas y m neuronas de salida:

$$\frac{\partial}{\partial w_3[k,m]}E(\cdot) = (\sigma_3'[m]\,z_2[k])\ \frac{\partial E(\cdot)}{\partial z_3[m]} = -2\,e[m]\,(\sigma_3'[m]\,x[k])$$

Regla de aprendizaje:

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \mu\frac{\partial}{\partial \mathbf{w}(t)}E(\cdot)$$