Para un modelo cuadrático y n puntos tenemos una función de costo

$$\chi^2(a_0, a_1, a_2) = \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\right)^2$$

De la que podemos obtener un sistema de ecuaciones al minimizar $\chi^2(a_0, a_1, a_2)$ de la siguiente manera

- Para a_0

$$\frac{\partial \chi^2}{\partial a_0} = -2 \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\right) = 0$$

podemos despejar y_i y cancelar el 2 en ambos lados de la igualdad

$$\sum_{i=1}^{n} a_0 + a_1 x_i + a_2 x_i^2 = \sum_{i=1}^{n} y_i \qquad (1)$$

- Para a_1

$$\frac{\partial \chi^2}{\partial a_1} = -2 \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\right) x_i = 0$$

$$-2 \sum_{i=1}^{n} \left(y_i x_i - (a_0 x_i + a_1 x_i^2 + a_2 x_i^3)\right) = 0$$

podemos despejar la expresión con y_i y cancelar el 2 en ambos lados de la igualdad

$$\sum_{i=1}^{n} a_0 x_i + a_1 x_i^2 + a_2 x_i^3 = \sum_{i=1}^{n} y_i x_i \qquad (2)$$

• Para a_2

$$\frac{\partial X^2}{\partial a_2} = -2\sum_{i=1}^{n}\left(y_i - (a_0 + a_1 x_i + a_2 x_i^2)\right) x_i^2 = 0$$

$$-2\sum_{i=1}^{n}\left(y_i x_i^2 - (a_0 x_i^2 + a_1 x_i^3 + a_2 x_i^4)\right) = 0$$

podemos despejar la expresión con y_i y cancelar el 2 en ambos lados de la igualdad

$$\sum_{i=1}^{n} a_0 x_i^2 + a_1 x_i^3 + a_2 x_i^4 = \sum_{i=1}^{n} y_i x_i^2 \qquad (3)$$

Las ecuaciones (1), (2) y (3) conforman nuestro sistema de ecuaciones

Scanned by TapScanner

Vectorialmente el decenso del gradiente está definido por

$$\vec{\theta}_{j+1} = \vec{\theta}_j - \gamma \left(-2 \sum_{i=1}^{N} \left(y_i - M(x_i, \vec{\theta}_j) \right) \nabla_\theta M(x_i, \theta_j) \right)$$

Puesto que el parámetro $\vec{\theta}$ es un vector con componentes $\theta_0, \theta_1, \theta_2$ implementadas en el modelo de ajuste

$$M(x, \vec{\theta}) = \frac{\theta_0}{\theta_1 + e^{-\theta_2 x}}$$

al cual le corresponde una función de costo

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{N} \left(\frac{y_i - M(x_i, \vec{\theta})}{\sigma_i} \right)^2$$

que se deriva para obtener el decenso del gradiente.

Haciendo el error $\sigma_i = 1$ se puede llegar a la siguiente expresión para el método mencionado

$$\vec{\theta}_{j+1} = \vec{\theta}_j - \gamma \left(-2 \sum_{i=1}^{N} \left(y_i - M(x_i, \vec{\theta}_j) \right) \frac{\partial M(x_i, \vec{\theta}_j)}{\partial \theta_k} \right)$$

donde $\vec{\theta}_{j+1}$ es el vector $\vec{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix}$ actualizado,

$\vec{\theta}_j$ es el vector precedente, γ es la taza de aprendizaje,

$M(x_i, \vec{\theta}_j)$ es el modelo de ajuste evaluado en x_i y $\vec{\theta}_j$

y $\dfrac{\partial M(x_i, \vec{\theta}_j)}{\partial \theta_k}$ es la derivada parcial del modelo evaluado

con respecto a cada componente de $\vec{\theta}_j$ $(k = 0, 1, 2)$.

Esta última expresión equivale al gradiente

$$\nabla_{\theta_i} M(x_i, \vec{\theta}_j) = \left[\frac{\partial M(x_i, \vec{\theta}_j)}{\partial \theta_0}, \frac{\partial M(x_i, \vec{\theta}_j)}{\partial \theta_1}, \frac{\partial M(x_i, \vec{\theta}_j)}{\partial \theta_2} \right]$$

Que también se expresa de forma compacta en la métrica derivada parcialmente:

$$\frac{\partial \chi^2 \vec{\theta}}{\partial \theta_i} = -2 \sum_{i=1}^{N} \left(y_i - M(x_i, \vec{\theta}) \right) \frac{\partial M(x_i, \vec{\theta})}{\partial \theta_i}$$

De esta forma se puede expresar el decenso del gradiente como

$$\begin{bmatrix} \theta^0_{j+1} \\ \theta^1_{j+1} \\ \theta^2_{j+1} \end{bmatrix} = \begin{bmatrix} \theta^0_j \\ \theta^1_j \\ \theta^2_j \end{bmatrix} - \gamma \left(-2 \sum_{i=1}^{N} (y_i - M(x_i, \vec{\theta}_j)) \begin{bmatrix} \frac{\partial M(x_i, \vec{\theta}_j)}{\theta^0} \\ \frac{\partial M(x_i, \vec{\theta}_j)}{\theta^1} \\ \frac{\partial M(x_i, \vec{\theta}_j)}{\theta_2} \end{bmatrix} \right)$$

lo que equivale a la primera expresión.