https://github.com/cristobalstrange/Tarea-IMT2220-Grupo-5

Integrantes: Cristobal Strange, Sergio Ramirez, Bastian Baez Oses.

Aproximación lineal

Para este ejercicio se leen los M datos $(x_i, y_i) \in \mathbb{R}^2$ con $i = 1, 2, \dots, M$.
Luego se crea la siguiente matriz $A \in \mathcal{M}^{M\times 3}$ y el vector $\vec{b} \in \mathbb{R}^M$

$$A = \begin{pmatrix} 1 & x_1 & (x_1)^2 \\ 1 & x_2 & (x_2)^2 \\ 1 & x_3 & (x_3)^3 \\ \vdots & \vdots & \vdots \\ 1 & x_M & (x_M)^3 \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_M \end{pmatrix}$$

y se busca minimizar el siguiente problema:

$$\min_{\vec{x} \in \mathbb{R}^3} \|A\vec{x} - \vec{b}\|_2^2$$

Donde $\vec{x}$ es el vector de los coeficientes del polinomio que estamos buscando $(y = \beta_0 + \beta_1 x + \beta_2 x^2)$.

En resumen $\vec{x} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$.

Si

$$J(\vec{x}) = \|A\vec{x} - \vec{b}\|_2^2$$

Entonces.

$$\nabla J(\vec{x}) = 2(A\vec{x} - \vec{b})^T A$$

Donde obtendremos un punto crítico si $\nabla J(\vec{x}) = 0$, es decir:

Obs: Este punto crítico es mínimo dado que $\|\cdot\|_2^2$ siempre mayor o igual que 0.

$$2(A\vec{x} - \vec{b})^T A = 0$$

$$((A\vec{x})^T - \vec{b}^T) A = 0$$

$$(\vec{x}^T A^T - \vec{b}^T) A = 0$$

$$\Rightarrow \vec{x}^T A^T A = \vec{b}^T A$$

Si se asume que $A^T A$ es invertible.

$$\Rightarrow \vec{x}^T A^T A = \vec{b}^T A \quad / \cdot (A^T A)^{-1}$$

$$\vec{x}^T = \vec{b}^T A (A^T A)^{-1} \quad / (\ )^T$$

$$\Rightarrow \vec{x} = (A^T A)^{-1} A^T \vec{b}.$$

$A^T A$ es invertible $\Leftrightarrow \forall i = 1, 2, \dots, M$, $\lambda_i \neq 0$

Es decir, todos sus valores propios son distintos de 0.

Entonces si $A^T A$ no es invertible, podemos plantear el siguiente problema.

$$J(\vec{x}) = \|A\vec{x} - \vec{b}\|_2^2 + \hat{\lambda} \|\vec{x}\|_2^2$$

$$\nabla J(\vec{x}) = 2(A\vec{x} - \vec{b})^T A + 2\lambda \vec{x}^T$$

y el punto crítico (que a la vez sera mínimo) seria:

$$2\,(A\vec{x}-\vec{b})\,A + 2\hat{\lambda}\,\vec{x}^T = 0$$

$$\vec{x}^T A^T A - b^T A + \hat{\lambda}\,\vec{x}^T = 0$$

$$\vec{x}^T (A^T A + \hat{\lambda} I) = b^T A$$

Si $(A^T A + \hat{\lambda} I)$ es invertible, el punto crítico seria

$$\vec{x} = (A^T A + \hat{\lambda} I)\, A^T \vec{b}$$

Obs: $A^T A + \hat{\lambda} I = P D P^{-1} + \lambda\, P I P^{-1}$

$$= P\,\underbrace{(D + \hat{\lambda} I)}_{\begin{pmatrix} \lambda_1+\hat{\lambda} & 0 & \cdots & 0 \\ 0 & \lambda_2+\hat{\lambda} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \lambda_M+\hat{\lambda} \end{pmatrix}}\,P^{-1}$$

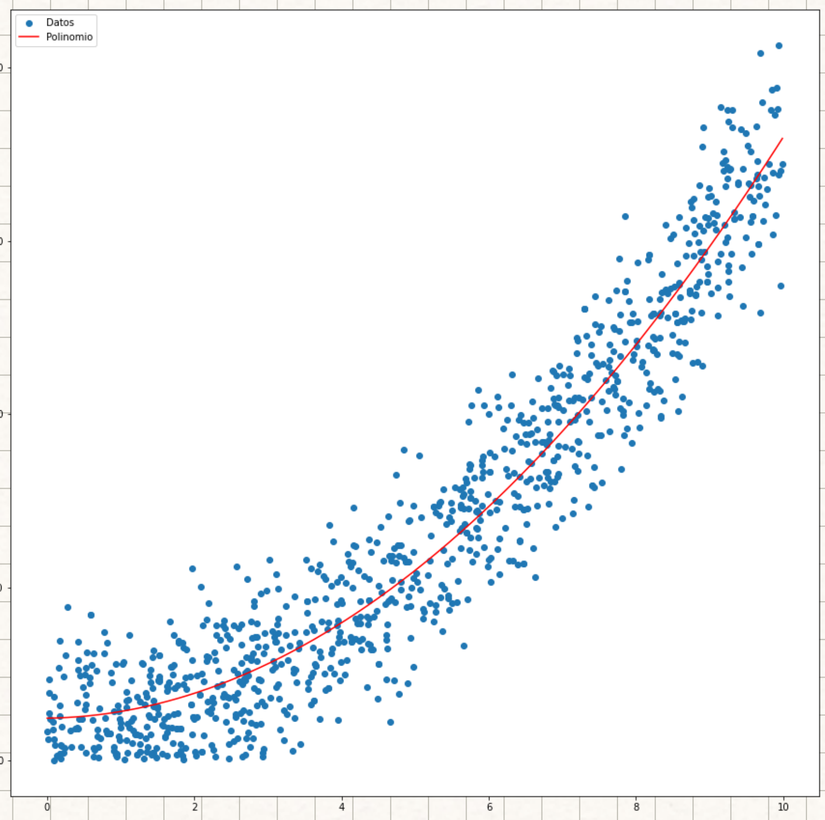
Entonces para que este metodo funcione se tiene que escoger un $\hat{\lambda}$ tal que

$$\lambda_i + \hat{\lambda} \neq 0 \quad \forall\, i = 1, 2, \ldots, M.$$

Aplicando este metodo en los dato obtenemos que

$\beta_0 \approx 2{,}4 \qquad \beta_1 \approx 0{,}1 \qquad \beta_2 \approx 0{,}3$

y al graficarlo se veria de la siguiente forma:



Red Neuronal

a) $\sigma(t) = \dfrac{1}{1+e^{-t}}$

Entonces:

$$\sigma'(t) = \left(\frac{1}{1+e^{-t}}\right)'$$

$$= \frac{-(1+e^{-t})'}{(1+e^{-t})^2}$$

$$= \frac{-(e^{-t})'}{(1+e^{-t})^2}$$

$$= \frac{e^{-t}}{(1+e^{-t})^2} \quad //$$

b)

$$J(a,b,c,d,e,f) = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \sigma(d\cdot\sigma(a u_i + b\cdot v_i + c w_i + e) + f)\right)^2$$

Sea. $\lambda = d\cdot\sigma(a u_i + b\cdot v_i + c w_i + e) + f$

$p = a u_i + b\cdot v_i + c w_i + e$

$$\Rightarrow J(a,b,c,d,e,f) = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \sigma(\lambda)\right)^2$$

$$\frac{\partial J}{\partial a} = \frac{1}{n}\sum_{i=1}^{n} 2\left(z_i - \sigma(\lambda)\right)\cdot\left(-\sigma'(\lambda)\cdot d\,\sigma'(p)\cdot u_i\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n}\left(\sigma(\lambda) - z_i\right)\cdot\left(\sigma'(\lambda)\cdot d\,\sigma'(p)\cdot u_i\right)$$

$$\frac{\partial J}{\partial b} = \frac{1}{n}\sum_{i=1}^{n} 2\left(z_i - \sigma(\lambda)\right)\cdot\left(-\sigma'(\lambda)\cdot d\,\sigma'(p)\cdot w_i\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n}\left(\sigma(\lambda) - z_i\right)\cdot\left(\sigma'(\lambda)\cdot d\,\sigma'(p)\cdot v_i\right)$$

$$\frac{\partial J}{\partial c} = \frac{1}{n}\sum_{i=1}^{n} 2\left(z_i - \sigma(\lambda)\right)\cdot\left(-\sigma'(\lambda)\cdot d\,\sigma'(p)\cdot w_i\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n}\left(\sigma(\lambda) - z_i\right)\cdot\left(\sigma'(\lambda)\cdot d\,\sigma'(p)\cdot w_i\right)$$

$$\frac{\partial J}{\partial d} = \frac{1}{n}\sum_{i=1}^{n} 2\left(z_i - \sigma(\lambda)\right)\cdot\left(-\sigma'(\lambda)\cdot\sigma(p)\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n} \left(\sigma(\lambda) - z_i\right)\cdot\left(\sigma'(\lambda)\cdot\sigma(p)\right)$$

$$\frac{\partial J}{\partial e} = \frac{1}{n}\sum_{i=1}^{n} 2\left(z_i - \sigma(\lambda)\right)\cdot\left(-\sigma'(\lambda)\cdot d\,\sigma'(p)\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n} \left(\sigma(\lambda) - z_i\right)\cdot\left(\sigma'(\lambda)\cdot d\,\sigma'(p)\right)$$

$$\frac{\partial J}{\partial f} = \frac{1}{n}\sum_{i=1}^{n} 2\left(z_i - \sigma(\lambda)\right)\cdot\left(-\sigma'(\lambda)\right)$$

$$= \frac{2}{n}\sum_{i=1}^{n} \left(\sigma(\lambda) - z_i\right)\cdot\left(\sigma'(\lambda)\right)$$

Grafico error por iteraciones: