Aprendizagem

Instituto Superior Técnico

outubro de 2023

Homework 3 - Report

Joana Pimenta (103730), Rodrigo Laia (102674)

Pen and Paper

1. a) Uma função radial basis permite mapear observações para um novo espaço baseando-se na distância entre as observações e os centróides.

$$\phi_j(x) = \exp\left(-\frac{\|\vec{x} - c_j\|^2}{2}\right) \tag{1}$$

Cálculo dos vetores transformados:

$$\vec{\phi_i} = \left(\exp\left(-\frac{\|\vec{x}_i - c_1\|^2}{2}\right), \exp\left(-\frac{\|\vec{x}_i - c_2\|^2}{2}\right), \exp\left(-\frac{\|\vec{x}_i - c_3\|^2}{2}\right)\right) \tag{2}$$

Assim os vetores transformados obtidos foram:

$$\phi_1 = (0.74826, 0.74826, 0.10127)$$

$$\phi_2 = (0.81465, 0.27117, 0.33121)$$

$$\phi_3 = (0.71177, 0.09633, 0.71177)$$

Para fazer regressão de Ridge é necessário minimizar a função de erro:

$$E(\vec{w}) = \frac{1}{2}\sum_{i=1}^{n}(z_i - \vec{w}^T \cdot x_i)^2 + \frac{\lambda}{2}||\vec{w}||^2 \tag{3}$$

Sendo que isso é equivalente a calcular $\vec{w}$ através da seguinte fórmula:

$$\vec{w} = (X^T \cdot X + \lambda \cdot I)^{-1} \cdot X^T \cdot \vec{z} \tag{4}$$

Uma vez que estamos a trabalhar com uma transformação de espaços, é necessário calcular a matriz transformada Φ colocando para cada linha um 1 na primeira coluna e depois o vetor transformado de cada observação. Utilizamos então as fórmulas acima com Φ no lugar de X, assumindo que após a transformação a relação entre as variáveis e o target é linear.

Cálculos intermédios:

$$\Phi = \begin{bmatrix} 1 & 0.74826 & 0.74826 & 0.10127 \\ 1 & 0.81465 & 0.27117 & 0.33121 \\ 1 & 0.71177 & 0.09633 & 0.71177 \\ 1 & 0.88250 & 0.16122 & 0.65377 \end{bmatrix}$$

$$\Phi^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0.74826 & 0.81465 & 0.71177 & 0.88250 \\ 0.74826 & 0.27117 & 0.09633 & 0.16122 \\ 0.10127 & 0.33121 & 0.71177 & 0.65377 \end{bmatrix}$$

$$(\Phi^T \cdot \Phi + \lambda \cdot I)^{-1} \cdot \Phi^T = \begin{bmatrix} 0.14105 & 0.35022 & 0.35575 & -0.30185 \\ -0.09064 & 0.43823 & -0.50361 & 0.53370 \\ 0.99394 & -0.50615 & -0.13690 & -0.16477 \\ -0.31222 & -0.65246 & 0.72647 & 0.42436 \end{bmatrix}$$

$$\vec{w} = \begin{bmatrix} 0.33914 \\ 0.19945 \\ 0.40096 \\ -0.29600 \end{bmatrix}$$

Assim, a regressão de Ridge obtida foi:

$$\hat{z} = 0.33914 + 0.19945 \cdot \phi_1 + 0.40096 \cdot \phi_2 - 0.29600 \cdot \phi_3$$

b) Para calcular o RMSE (root mean square error) foi utilizada a seguinte fórmula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (z_i - \hat{z}_i)^2} \tag{5}$$

$$\hat{z}_i = \vec{w}^T \cdot \vec{\phi_i} \tag{6}$$

Targets estimados:

$$\hat{z}_1 = 0.75843$$

$$\hat{z}_2 = 0.51231$$

$$\hat{z}_3 = 0.30905$$

$$\hat{z}_4 = 0.38629$$

Assim, o RMSE obtido foi:

$$RMSE = 0.06508$$

2. É importante referir que para este exercício se utilizou a seguinte notação: $L^{\text{label}}_{observation}$

   As fórmulas utilizadas foram:

$$\vec{x}^{[p]} = \phi(\vec{W}^{[p]} \cdot \vec{x}^{[p-1]} + \vec{b}^{[p]})$$

$$\vec{\delta}^{[p]} = \frac{\partial E}{\partial \vec{x}^{[p]}} \circ \frac{\partial \vec{x}^{[p]}}{\partial \vec{z}^{[p]}} = (\vec{x}^{[p]} - \vec{t}) \circ \phi'(\vec{z}^{[p]}), \text{ para a última layer}$$

$$\vec{\delta}^{[p]} = (\frac{\partial \vec{z}^{[p+1]}}{\partial \vec{x}^{[p]}})^T \cdot \vec{\delta}^{[p+1]} \circ \frac{\partial \vec{x}^{[p]}}{\partial \vec{z}^{[p]}} = (\vec{W}^{[p+1]})^T \cdot \vec{\delta}^{[p+1]} \circ \phi'(\vec{z}^{[p]}), \text{ para as outras layers}$$

$$\vec{W}^{[p]} = \vec{W}^{[p]} - \eta \cdot \frac{\partial E}{\partial \vec{W}^{[p]}} = \vec{W}^{[p]} - \eta \cdot \vec{\delta}^{[p]} \cdot (\vec{x}^{[p-1]})^T$$

$$\vec{b}^{[p]} = \vec{b}^{[p]} - \eta \cdot \frac{\partial E}{\partial \vec{b}^{[p]}} = \vec{b}^{[p]} - \eta \cdot \vec{\delta}^{[p]}$$

   Estas expressões são válidas para squared error loss function.

   Dados necessários para começar o algoritmo:

$$\vec{x}_1^{[0]} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, t_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \vec{x}_2^{[0]} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}, t_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$W^{[1]} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, b^{[1]} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, W^{[2]} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 4 & 1 \end{bmatrix}, b^{[2]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, W^{[3]} = \begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 1 & 1 \end{bmatrix}, b^{[3]} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\phi = tanh(0.5x - 2),\ \phi' = 0.5 * (1 - tanh^2(0.5x - 2))$$

POR GUIA DE REDE!!!!!!!!

Primeiro é necessário realizar (forward) propagation para obter os valores das observações:

- Para a primeira observação:

$$z_1^{[1]} = W^{[1]} \cdot x_1^{[0]} + b^{[1]} = \begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix} \implies x_1^{[1]} = \phi(z_1^{[1]}) = \begin{bmatrix} 0.46212 \\ 0.76159 \\ 0.46212 \end{bmatrix}$$

$$z_1^{[2]} = W^{[2]} \cdot x_1^{[1]} + b^{[2]} = \begin{bmatrix} 2.68583 \\ 4.97061 \end{bmatrix} \implies x_1^{[2]} = \phi(z_1^{[2]}) = \begin{bmatrix} -0.57642 \\ 0.45048 \end{bmatrix}$$

$$z_1^{[3]} = W^{[3]} \cdot x_1^{[2]} + b^{[3]} = \begin{bmatrix} 0.87406 \\ -0.27878 \\ 0.87406 \end{bmatrix} \implies x_1^{[3]} = \phi(z_1^{[3]}) = \begin{bmatrix} -0.9159 \\ -0.97266 \\ -0.9159 \end{bmatrix}$$

- Para a segunda observação:

Depois é necessário realizar (backward) propagation dos erros da última camada para a primeira:

- Para a primeira observação:

$$\delta_1^{[3]} = (x_1^{[3]} - t_1) \circ \phi'(z_1^{[3]}) = \begin{bmatrix} -0.07379 \\ -0.05320 \\ -0.07379 \end{bmatrix}$$

$$\delta_1^{[2]} = (W^{[3]})^T \cdot \delta_1^{[3]} \circ \phi'(z_1^{[2]}) = \begin{bmatrix} -0.10255 \\ -0.08001 \end{bmatrix}$$

$$\delta_1^{[1]} = (W^{[2]})^T \cdot \delta_1^{[2]} \circ \phi'(z_1^{[1]}) = \begin{bmatrix} -0.07179 \\ -0.08874 \\ -0.07179 \end{bmatrix}$$

- Para a segunda observação:

Programming - Código Python e Resultados Obtidos

1.