EFC3 - MLP e SVM

Rafael Gonçalves - RA: 186062

12 de Junho de 2019

Parte I - Retropropagação de erro

A rede apresentada pode ser descrita como:

$$\mathbf{y} = \sum \mathbf{W}\mathbf{F}\left(\sum \mathbf{V}\mathbf{x}\right) \tag{1}$$

em que $\mathbf{y}$ é o vetor de saídas, $\mathbf{x}$ é o vetor de entradas incluindo um valor unitário para facilitar o cálculo do bias, V e W são as matrizes de peso e $\mathbf{F}(\mathbf{x})$ uma matriz diagonal com os únicos elementos não nulos sendo $\mathbf{F_{i,i}}(\mathbf{x}) = f(x_i)$
Podemos definir $\mathbf{u} = \sum \mathbf{V}\mathbf{x}$ e $\mathbf{a} = \mathbf{F}(\mathbf{u})$ para facilitar os cálculos do gradiente.
A função de custo é:

$$J = \mathbf{e}^{\mathbf{T}}\mathbf{e} \tag{2}$$

Com $\mathbf{e} = \mathbf{y_{target}} - \mathbf{y}$, $\mathbf{y}$ sendo o vetor coluna que representa a saída da rede e $\mathbf{y_{target}}$ o vetor coluna com os rótulos.
Então o cálculo do gradiente se dá da seguinte forma:

$$\frac{\partial J}{\partial \mathbf{W}} = \frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{W}} = 2\mathbf{e}(-1)\mathbf{a} = \boldsymbol{\delta_W}\mathbf{a}$$

$$\frac{\partial J}{\partial \mathbf{V}} = \frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{a}}\frac{\partial \mathbf{a}}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{V}} = 2\mathbf{e}(-1)\mathbf{W}\dot{\mathbf{F}}(\mathbf{u})\mathbf{x} = \boldsymbol{\delta_V}\mathbf{x}$$

em que $\dot{\mathbf{F}}_{i,j}(\mathbf{u}) = \dot{f}(u_i)$ se $i = j$ e vale 0 caso o contrário.

Desta maneira podemos calcular:

$$\frac{\partial J}{\partial v_{21}} = -2e_1 w_{20} \dot{f}(u_1) x_2$$

em que $u_1 = v_{01} + v_{11}x_1 + v_{12}x_2$ e $e_1 = y_{target_1} - y_1$

Parte II - Classificação binária com MLPs e SVMs

Multilayer Perceptrons

Support-vector Machines