Examen de Teoría de Redes Neuronales

Máster MIARFID, Universitat Politècnica de València, Diciembre de 2015

Apellidos: | | **Nombre:** | |

Cuestiones (2 puntos, 30 minutos, sin apuntes)

1 $\boxed{\text{C}}$ La frontera de decisión entres dos clases i, j es el conjunto de puntos $\boldsymbol{x} \in E$ para los que $g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$, donde $g_i : \mathcal{R}^d \to \mathcal{R}$ y $g_i : \mathcal{R}^d \to \mathcal{R}$ son las funciones discriminantes asociadas a las clases i y j. Indicar qué afirmación es cierta:

A) Si $d = 2$ y g_i y g_j son funciones discriminantes cualesquiera, la frontera de decisión es una línea recta.

B) Si $d = 1$ y g_i y g_j son funciones discriminantes lineales, la frontera de decisión es una línea recta.

C) Si $d = 2$ y g_i y g_j son funciones discriminantes lineal, la frontera de decisión es una línea recta.

D) Si $d = 3$ y g_i y g_j son funciones discriminantes cualesquiera, la frontera de decisión es un punto.

2 $\boxed{\text{B}}$ Sea un problema de predicción de la temperatura en 10 ciudades. Se propone resolverlo con una red neuronal a partir de diferentes variables de entrada y 10 neuronas en la capa de salida. En este sentido, cuál de las siguientes funciones de activación para dicha capa de salida es la más adecuada

A) Sigmoide

B) Lineal

C) ReLu

D) Softmax

3 $\boxed{\text{C}}$ Dada una neurona definimos su valor pre-activación como x y su valor post-activación como y, por lo tanto $y = f(x)$ siendo $f()$ la función de activación. Si la derivada parcial del error con respecto a y (delta en y) es 0.5 cuál es la derivada con respecto a x (delta en x) siendo $f()$ una ReLU:

A) 1

B) 0

C) 0.5

D) -1

4 $\boxed{\text{A}}$ Las funciones radiales son un caso particular de las funciones discriminantes generalizadas dónde dado un $\boldsymbol{x}$ se obtiene una representación alternativa empleando funciones de cambio de espacio Φ_k, donde en general:

A) $\Phi_k = \Phi(r)$ siendo $r =|| \boldsymbol{x} - \boldsymbol{\mu}_k ||$

B) $\Phi_k =|| \boldsymbol{x} - \boldsymbol{\mu}_k ||$

C) $\Phi_k = K(x, x)$ siendo $K(\cdot, \cdot)$ una función kernel

D) Ninguna de las anteriores

5 $\boxed{\text{C}}$ Dado un conjunto de muestras de aprendizaje $X = \{(\boldsymbol{x}_1, c_1), (\boldsymbol{x}_2, c_2), \cdots, (\boldsymbol{x}_n, c_n)\}$, donde $\boldsymbol{x} \in \mathcal{R}^d$ y $c_i \in \{-1, +1\}$ los SVM plantean la siguiente función objetivo:

A) $\arg\min_{(\mathbf{w}, w_0)} \frac{1}{2}\|\mathbf{w}\|^2$ s.t. $c_i(\mathbf{w} \cdot \mathbf{x_i} + w_0) \leq 1, \quad 1 \leq i \leq n$

B) $\arg\max_{(\mathbf{w}, w_0)} \frac{1}{2}\|\mathbf{w}\|^2$ s.t. $c_i(\mathbf{w} \cdot \mathbf{x_i} + w_0) \leq 1, \quad 1 \leq i \leq n$

C) $\arg\min_{(\mathbf{w}, w_0)} \frac{1}{2}\|\mathbf{w}\|^2$ s.t. $c_i(\mathbf{w} \cdot \mathbf{x_i} + w_0) \geq 1, \quad 1 \leq i \leq n$

D) $\arg\max_{(\mathbf{w}, w_0)} \frac{1}{2}\|\mathbf{w}\|^2$ s.t. $c_i(\mathbf{w} \cdot \mathbf{x_i} + w_0) \geq 1, \quad 1 \leq i \leq n$

6 $\boxed{D}$ En el algoritmo passive aggressive lineal con vector de pesos $\boldsymbol{w}_t$ en el instante t, dado un nuevo par $(\boldsymbol{x}_t, y_t)$ se emplea la función de pérdida Hinge-loss definida como:

A) $max(0, \boldsymbol{w}_t\boldsymbol{x}_t)$
B) $max(0, y_t\boldsymbol{w}_t\boldsymbol{x}_t)$
C) $max(0, \boldsymbol{w}_t\boldsymbol{x}_t)$
D) $max(0, 1 - y_t\boldsymbol{w}_t\boldsymbol{x}_t)$

7 $\boxed{A}$ Dado un vector de pesos, $\boldsymbol{w} = [1,\ 0,\ 1,\ 1,\ 0,\ 1]$ se desea aplicar una regularización maxnorm con parámetro $c = 3$. El resultado sería:

A) $\boldsymbol{w} = [1,\ 0,\ 1,\ 1,\ 0,\ 1]$
B) $\boldsymbol{w} = [\frac{3}{2},\ 0,\ \frac{3}{2},\ \frac{3}{2},\ 0,\ \frac{3}{2}]$
C) $\boldsymbol{w} = [3,\ 0,\ 3,\ 3,\ 0,\ 3]$
D) $\boldsymbol{w} = [2,\ 0,\ 2,\ 2,\ 0,\ 2]$

8 $\boxed{C}$ En esencia la regularización l_2 lo que hace es:

A) Restar una constante λ a los pesos
B) Restar o sumar, según el signo, una constante λ a los pesos
C) Restar un factor proporcional al valor de los pesos
D) Sumar un factor proporcional al valor de los pesos

9 $\boxed{C}$ Dada una capa convolucional 64@3x3 que se aplica a una entrada 32@32x32, asumiendo stride=1 y no padding, el resultados sería

A) 32@32x32
B) 64@32x32
C) 64@30x30
D) 32@30x30

10 $\boxed{A}$ En el entrenamiento de redes recurrentes simple con el 'Exact Gradient Algorithm" (EGA) o con el "Back-propagation through time algorithm" (BPTT), indicar qué afirmación es correcta

A) El EGA presenta un coste computacional temporal superior al BPTT, pero un coste computacional espacial inferior.
B) El EGA presenta un coste computacional temporal y espacial superior al BPTT. inferior.
C) El EGA presenta un coste computacional temporal inferior al BPTT, pero un coste computacional espacial superior.
D) El EGA presenta un coste computacional temporal y espacial inferior al BPTT.

Examen de Teoría de Redes Neuronales

Máster MIARFID, Universitat Politècnica de València, Diciembre de 2015

Apellidos: | | **Nombre:** | |

Problemas (4 puntos, 90 minutos, con apuntes)

Problema 1 (2 puntos)

Dada una red neuronal con una capa oculta de 2 neuronas. La red tiene 3 neuronas de entrada y una de salida. Implementa una ReLu en la capa oculta y dado que tenemos un problema de regresión no limitado hemos seleccionado una función lineal en la capa de salida. Los pesos y bias que conectan la entrada con la capa oculta son los siguientes:

$$W_1 = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & -1 \end{pmatrix}$$

$$b_1 = [-1\ 1]$$

Mientras que los pesos y bias que conectan la capa oculta con la capa de salida son los siguientes:

$$W_2 = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

$$b_2 = [0]$$

Dada la entrada $\boldsymbol{x} = [1\ 0\ 0]$ cuyo valor objetivo (target) es $t = 1$ y asumiendo un criterio de aprendizaje que minimize el error cuadrático medio ($error = \frac{1}{2}(t-y)^2)$) siendo y la salida de la red. **Se pide**:

a) (0.5 puntos) Calcula el error que proporciona la red para esta muestra en concreto

b) (1 punto) Dada la constante de aprendizaje $\mu = 0.1$ actualiza los parámetros W_2 y b_2 para dicha muestra

c) (0.5 puntos) Repite el paso anterior suponiendo una regularización l_2 con parámetro $\lambda = 0.01$

a) Calculamos la activación de la capa oculta que llamaremos Z

$$Z = (1\ 0\ 0)\begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & -1 \end{pmatrix}^t + (-1\ 1) = (0\ 3)$$

Caculamos la salida

$(0\ 3)(1\ 0)^t + (0) = 0$

El error sería$= \frac{1}{2}(1-0)^2) = 0.5$

b) Calculamos la derivada del error con respecto a la salida

$\frac{\partial error}{\partial y} = -(t-y) = -(1-0) = -1$

Con la regla de la cadena obtenemos la derivada con respecto a los parámetros a actualizar. Como la función de activación de la salida es lineal se simplifica bastante dado que activación y pre-activación son los mismo y:

$\frac{\partial error}{\partial W_2} = \frac{\partial error}{\partial y}\frac{\partial y}{\partial W_2} = -1 * Z = -(0\ 3)$

$\frac{\partial error}{\partial b_2} = \frac{\partial error}{\partial y}\frac{\partial y}{\partial b_2} = -1 * 1 = -1$

Con los gradientes calculados aplicamos las actualizaciones:

$W_2 = W_2 - \mu\frac{\partial error}{\partial W_2} = (1\ 0) - 0.1(-(0\ 3)) = (1\ 0.3)$

$b_2 = b_2 - \mu\frac{\partial error}{\partial b_2} = (0) - 0.1(-1) = (0.1)$

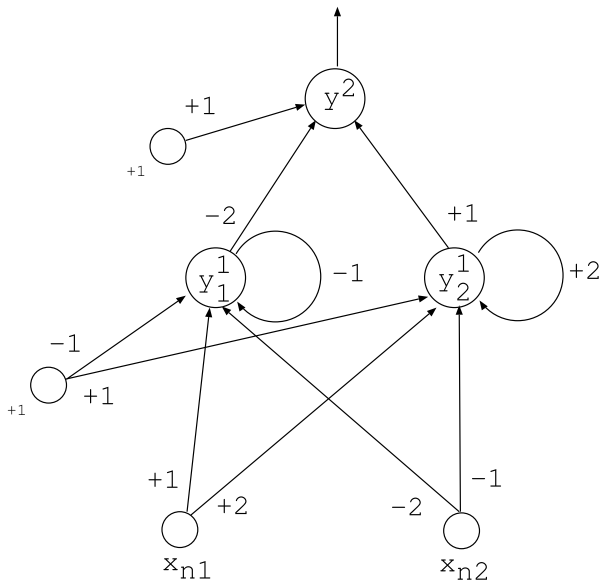
c) Con los mismos gradientes que en el caso anterior pero la ecuación de actualización tiene en cuenta la regularización:

$W_2 = W_2 - \mu\frac{\partial error}{\partial W_2} - \lambda\mathbf{W_2} = (1\ 0) - 0.1(-(0\ 3)) - 0.01(1\ 0) = (0.99\ 0.3)$

$b_2 = b_2 - \mu\frac{\partial error}{\partial b_2} - \lambda\mathbf{b_2} = (0) - 0.1(-1) - 0.01(0) = (0.1)$

Problema 2 (2 puntos)

Las funciones de activación en los nodos de la figura son del tipo ReLU: $f(z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases}$.



Dada una secuencia de entrada $(+1,+1), (0,+1), (+1,-1)$

a) (1 punto) Calcular la secuencia de los valores de los nodos ocultos Y_1^1 y Y_2^1

b) (1 punto) Calcular la secuencia de los valores del nodo de salida Y^2

a)

$$y_{n,1}^1 = f((-1) + (+1) \cdot x_{n,1} + (-2) \cdot x_{n,2} + (-1) \cdot y_{n-1,1})$$
$$y_{n,2}^1 = f((+1) + (+2) \cdot x_{n,1} + (-1) \cdot x_{n,2} + (+2) \cdot y_{n-1,2})$$
$$y_{1,1}^1 = f((-1) + (+1) \cdot (+1) + (-2) \cdot (+1) + (-1) \cdot (0)) = f(-2) = 0$$
$$y_{1,2}^1 = f((+1) + (+2) \cdot (+1) + (-1) \cdot (+1) + (+2) \cdot (0)) = f(+2) = 2$$
$$y_{2,1}^1 = f((-1) + (+1) \cdot (0) + (-2) \cdot (+1) + (-1) \cdot (0)) = f(-3) = 0$$
$$y_{2,2}^1 = f((+1) + (+2) \cdot (0) + (-1) \cdot (+1) + (+2) \cdot (+1)) = f(4) = 4$$
$$y_{3,1}^1 = f((-1) + (+1) \cdot (+1) + (-2) \cdot (-1) + (-1) \cdot (0)) = f(2) = 2$$
$$y_{3,2}^1 = f((+1) + (+2) \cdot (+1) + (-1) \cdot (-1) + (+2) \cdot (+2)) = f(12) = 12$$

b)

$$y_n^2 = f((+1) + (-2) \cdot y_{n,1}^1 + (+1) \cdot y_{n,2}^1)$$
$$y_1^2 = f((+1) + (-2) \cdot (0) + (+1) \cdot (+2)) = f(3) = 3$$
$$y_2^2 = f((+1) + (-2) \cdot (0) + (+1) \cdot (+4)) = f(5) = 5$$
$$y_3^2 = f((+1) + (-2) \cdot (2) + (+1) \cdot (+12)) = f(9) = 9$$