INTRODUCCION A LOS MODELOS COMPUTACIONALES 13 de enero 2020

Alumno/a D.....

Cuestiones.- A) (5 puntos)

i) (**1.5 puntos**) Explique brevemente el sentido que tiene el tamaño del bloque en redes neuronales convolucionales. ¿Qué efecto produce?

Solución.- El tamaño de bloque está asociado al número de patrones del conjunto de entrenamiento que se evalúan antes de lanzar el algoritmo de retropropagación del error. Es un parámetro del algoritmo que hay que entrenar por crosvalidación sobre el conjunto de entrenamiento. SI el tamaño del conjunto de entrenamiento es pequeño, el tamaño de bloque puede ser más pequeño. Cuanto más pequeño es el tamaño de bloque el coste computacional es mayor.

(ii) (**1 punto**) Dado el algoritmo de retropropagación del error en las redes neuronales de unidades sigmoides. Calcule sus derivadas y explique brevemente cómo se implementa.

Solución.-

Algorithm Backpropagation;

Start with randomly chosen weights;

while MSE is unsatisfactory

and computational bounds are not exceeded, do

for each input pattern x_p, $1 \leq p \leq P$,

Compute hidden node inputs ($net_{p,j}^{(1)}$);

Compute hidden node outputs ($x_{p,j}^{(1)}$);

Compute inputs to the output nodes ($net_{p,k}^{(2)}$);

Compute the network outputs ($o_{p,k}$);

Modify outer layer weights:

$$\Delta w_{k,j}^{(2,1)} = \eta(d_{p,k} - o_{p,k})\mathcal{S}'(net_{p,k}^{(2)})x_{p,j}^{(1)}$$

Modify weights between input & hidden nodes:

$$\Delta w_{j,i}^{(1,0)} = \eta \sum_k \left((d_{p,k} - o_{p,k})\mathcal{S}'(net_{p,k}^{(2)})w_{k,j}^{(2,1)}\right) \mathcal{S}'(net_{p,j}^{(1)})x_{p,i}$$

end-for

end-while.

S es una función sigmoide, entonces S'(x) = S(x)(1 – S(x)); mientras que $net^{(1)}_{p.j}$ es la salida del valor del patrón p por el nodo j de la capa oculta.

(iii) (**1 punto**) Ponga un ejemplo de red de unidades de base radial con tres variables de entrada, un nodo en capa oculta y uno de salida. Implemente la fórmula de un clasificador binario con una red de unidades RBF como la anterior.

Solución.- Definimos la función de clasificación en la forma

$f(\mathbf{x},\boldsymbol{\theta}) = \beta_0 + \beta_1 B(\mathbf{x};(\mathbf{c}\,|\,r))$ para $\mathbf{x}=(x_1,x_2,x_3)^T$ y $\mathbf{c}=(c_1,c_2,c_3)^T$, siendo

$B(\mathbf{x};(\mathbf{c}\,|\,r)) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|^2}{2r^2}\right)$, donde **c** es el centroide y r es el radio del núcleo

La función de decisión es de la forma

$$p(\mathbf{x},\boldsymbol{\theta}) = \frac{\exp(f(\mathbf{x},\boldsymbol{\theta}))}{1+\exp(f(\mathbf{x},\boldsymbol{\theta}))} \qquad Si\ p(\mathbf{x},\boldsymbol{\theta}) > 0,5 \text{ entonces } \mathbf{x} \in C_1 \text{ (clase positiva)}$$

(iv) **(1.5 puntos)** La función de activación ReLU ¿Que ventaja tiene sobre la función sigmoide en redes neuronales profundas?.

Solución.-

La función de rectificación lineal, **Rectified Linear Units, ReLU**, es de la forma

y = f (x)= max(0,x).

La derivada de esta función coincide con la función escalón utilizada por las neuronas de Mc Culloch y Pitts. Es una función asimétrica puesto que la respuesta a un patrón de entrada inhibidor es 0, esto es, no hay respuesta.

Como ventaja esta función permite que una red obtenga fácilmente representaciones dispersas. Por ejemplo, si hacemos una inicialización uniforme de los pesos, por ejemplo en el intervalo [-1, 1] o una Normal (0,1), alrededor del 50% de los valores de salida continuos de las unidades ocultas son ceros y este porcentaje puede aumentar fácilmente con la regularización inducida por la dispersión
La función de activación tradicional en redes neuronales superficiales tiene la forma de la función sigmoide fs, definida como

$$f(x) = \frac{1}{1+\exp(-x)},\ de\ donde\ f'(x) = f(x)(1-f(x))$$

Sin embargo, la saturación generalizada de la derivada de la función sigmoide hace que el aprendizaje basado en gradiente y sus variantes tengan un bajo rendimiento en la red neuronal de entrenamiento. Esto hace que las neuronas en las primeras capas de la red aprendan mucho más lentamente que las neuronas de las últimas capas. Es el problema de la desaparición del gradiente, también conocido como el desvanecimiento del gradiente.

Ejercicio 1.- (1,5 puntos) En un modelo de análisis discriminante lineal tenemos la siguiente regla de decisión.

$$Si\ (\mathbf{m}_2 - \mathbf{m}_1)^T \boldsymbol{\Sigma}^{-1}\mathbf{x} + 1/2(\mathbf{m}_1^T\boldsymbol{\Sigma}^{-1}\mathbf{m}_1 - \mathbf{m}_2^T\boldsymbol{\Sigma}^{-1}\mathbf{m}_2) - \ln\frac{P(C_1)}{P(C_2)} > 0$$

$$entonces\ \mathbf{x} \in C_1$$

¿Qué hipótesis asumimos para construir el modelo? ¿Qué significado tienen $P(C_1)$ y $P(C_2)$? Cuál sería la decisión en el caso de que el vector asociado a un patrón sea $\mathbf{x}^T$= (-0.5, 1), $\mathbf{m}_1^T$= (-1,1) $\mathbf{m}_2^T$=(1,-1)

$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}$, sabiendo además que tenemos 30 patrones de la clase 1 y 70 de la clase 2 de la muestra de entrenamiento.

Solución.-

$$Si\ (\mathbf{m}_2 - \mathbf{m}_1)^T \boldsymbol{\Sigma}^{-1}\mathbf{x} + 1/2(\mathbf{m}_1^T\boldsymbol{\Sigma}^{-1}\mathbf{m}_1 - \mathbf{m}_2^T\boldsymbol{\Sigma}^{-1}\mathbf{m}_2) - \ln\frac{P(C_1)}{P(C_2)} > 0$$

$$entonces\ \mathbf{x} \in C_1$$

Por una parte

$$((-1,1)-(1,-1))\begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} -0,5 \\ 1 \end{pmatrix} = 9$$

y por otra

$$(-1,1)\begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 6$$

$$(1,-1)\begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 6$$

como además

$$-\ln\frac{P(C_1)}{P(C_2)} = -\ln\frac{0,3}{0,7} = 0,85$$

Tenemos que el valor de la función discriminante es -9+0,85=--8,15 y por tanto **x=(-0,5, 1)** pertenece a la clase negativa cuyo vector de medias es $\mathbf{m_2}$

Ejercicio 2.- (2 puntos) Considere la siguiente matriz de pesos **W**:

$$\begin{pmatrix} 0.0 & 0.1 & -0.1 & -0.1 & -0.1 \\ 0.1 & 0.0 & -0.1 & 0.1 & 0.1 \\ -0.1 & -0.1 & 0.0 & -0.1 & -0.1 \\ -0.1 & 0.1 & -0.1 & 0.0 & 0.1 \\ -0.1 & 0.1 & -0.1 & 0.1 & 0.0 \end{pmatrix}$$

a) (**0,5 puntos**) ¿Qué tipo de red implementa? Muestre el grafo de esta red.
b) (**1 punto**) Comenzando en el estado [-1, 1, 1, 1, -1], calcule el flujo desde este estado al estado estable usando actualizaciones asincrónicas y síncronas.
d) **(0,5 puntos)** ¿Cómo se calcula su función de energía?

Solución.- a) Es una red de Hopfield con 5 nodos, donde los pesos son simétricos wij=wji y donde la matriz de pesos tiene la diagonal principal con valores wii=0

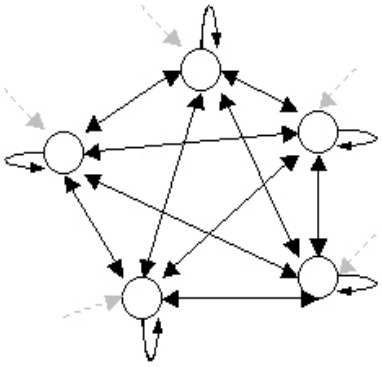


Figura 8: Arquitectura de una Red de Hopfield

b) Síncrona

$$\begin{pmatrix} 0.0 & 0.1 & -0.1 & -0.1 & -0.1 \\ 0.1 & 0.0 & -0.1 & 0.1 & 0.1 \\ -0.1 & -0.1 & 0.0 & -0.1 & -0.1 \\ -0.1 & 0.1 & -0.1 & 0.0 & 0.1 \\ -0.1 & 0.1 & -0.1 & 0.1 & 0.0 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -0,2 \\ 0 \\ 0 \\ 0.2 \end{pmatrix}$$ luego pasa al $$\begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 0.0 & 0.1 & -0.1 & -0.1 & -0.1 \\ 0.1 & 0.0 & -0.1 & 0.1 & 0.1 \\ -0.1 & -0.1 & 0.0 & -0.1 & -0.1 \\ -0.1 & 0.1 & -0.1 & 0.0 & 0.1 \\ -0.1 & 0.1 & -0.1 & 0.1 & 0.0 \end{pmatrix}\begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0,4 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$ luego se queda igual, luego es un **estado estable**, de vector [-1, -1, 1, 1, 1],

b) **Asíncrona**

$$\begin{pmatrix} 0.0 & 0.1 & -0.1 & -0.1 & -0.1 \\ 0.1 & 0.0 & -0.1 & 0.1 & 0.1 \\ -0.1 & -0.1 & 0.0 & -0.1 & -0.1 \\ -0.1 & 0.1 & -0.1 & 0.0 & 0.1 \\ -0.1 & 0.1 & -0.1 & 0.1 & 0.0 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -0,2 \\ 0 \\ 0 \\ 0.2 \end{pmatrix}$$ actualizamos el 5º elemento

$$W \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ -0,2 \\ 0 \\ 0 \\ 0.2 \end{pmatrix}$$ pasa al $$\begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$ Si actualizamos ahora el 3º tenemos

$$W \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0,2 \\ 0 \\ -0,2 \\ 0,2 \\ 0,2 \end{pmatrix}$$ pasa al $$\begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \\ 1 \end{pmatrix}$$ y este **es un estado estable** $$W \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.4 \\ -0,2 \\ 0,4 \\ 0,4 \end{pmatrix}$$ pasa al $$\begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \\ 1 \end{pmatrix}$$

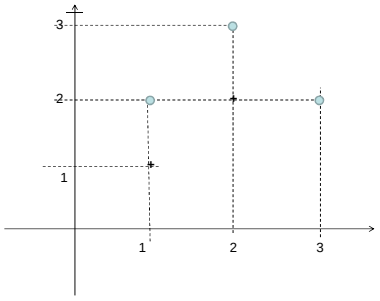
c) Si suponemos que Θ= 0, entonces para el estado de vector [-1, -1, 1, 1, 1], la función de energía es:

$$E(t) = -\frac{1}{2}\sum_{i,j} w_{ij} s_i s_j = -\frac{1}{2}\sum_{i=1}^{4}\sum_{j>i}^{5} w_{ij} s_i s_j = -\frac{1}{2} 2(w_{12}s_1s_2 + w_{13}s_1s_3 + ... + w_{35}s_3s_5 + w_{45}s_4s_5) =$$

$$= -\begin{pmatrix} 0.1(-1)(-1) - 0.1(-1)(1) - 0.1(-1)(1) - 0.1(-1)(1) \\ -0.1(-1)(1) + 0.1(-1)(1) + 0.1(-1)(1) \\ -0.1(1)(1) - 0.1(1)(1) \\ 0.1(1)(1) \end{pmatrix} = -0.2$$

o también $E(t) = -\frac{1}{2}\sum_{i,j} w_{ij} s_i s_j = -\frac{1}{2}\left((-1)(-0,4) + 0 + 0 + 0 + 0\right) = -0,2$

Ejercicio 3.- (2 puntos) Considere los cinco vectores bidimensionales de la siguiente figura, donde las cruces son patrones de la clase positiva y los círculos patrones de la clase negativa. Encuentre el SVM lineal que separa de manera óptima las clases al maximizar el margen utilizando tanto el primal como el dual. Dado el patrón de un conjunto de test de coordenadas (1,5 2) perteneciente a la clase positiva, explique cómo quedaría clasificado según el clasificador anterior.



Solución.- Los vectores soporte son de la clase positiva el (2,2) y de la negativa el (2,3) y el (3,2) tanto para el primal como para el dual, de esta forma consideramos que el (1,2) está mal clasificado en el conjunto de entrenamiento.

$H_0 : \mathbf{w.x} + w_0 = 1$
$H_1 : \mathbf{w.x} + w_0 = -1$

x_1	x_2	*Clase*
2	3	-1
2	2	$+1$

$$w_1x_1 + w_2x_2 + w_0 = -1 \qquad w_1x_1 + w_2x_2 + w_0 = 1$$
$$2w_1 + 3w_2 + w_0 = -1 \qquad 2w_1 + 2w_2 + w_0 = 1$$

de esta dos ecuaciones obtenemos que w_2=-2 y que $2w_1+w_0$=5, pero tenemos dos ecuaciones con tres incógnitas, así que tomamos como vector soporte negativo también el (3,2). Si ponemos las tres ecuaciones

tenemos
$$w_2 = 2$$
$$2w_1 + w_0 = 5$$
$$3w_1 + 2w_2 + w_0 = -1$$

de donde despejando tenemos w_0= 9, w_1=-2 y w_2=-2, y la ecuación del hiperplano es

$H : -2x_1 - 2x_2 + 9 = 0;\ x_2 = -x_1 + 4{,}5$. Esta solución hace que el patrón (1,5, 2) de la clase positiva tome el valor 2 de la función discriminante luego está bien clasificado.

Sol 2.- Si consideramos solo los vectores soporte (2,3) como negativo y (2,2) como positivo, tenemos en el dual, pero esta elección no tiene sentido, dado que no entrenamos bien los patrones (1,29 y (3,2)

$$\max \alpha_1 + \alpha_2 - \frac{1}{2}\left(\alpha_1^2 (s_1)^T.(s_1) - \alpha_1\alpha_2 (s_1)^T.(s_2) - \alpha_2\alpha_1 (s_2)^T.(s_1) + \alpha_2^2 (s_2)^T.(s_2)\right)$$
$$s.a. \quad -\alpha_1+\alpha_2=0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

donde $(s_1) = (2,2)^T$ el positivo *y* $(s_2) = (2,3)^T$ el negativo, por lo que

$$\max\ \alpha_1 + \alpha_2 - \frac{1}{2}\left(8\alpha_1^2 - 10\alpha_1\alpha_2 - 10\alpha_2\alpha_1 + 13\alpha_2^2\right)$$
$$s.a. \quad -\alpha_1 + \alpha_2 = 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0$$

derivando con respecto a α_1, α_2 y λ, tenemos

$$\begin{cases} 1 - 8\alpha_1 + 10\alpha_2 + \lambda = 0 \\ 1 + 10\alpha_1 - 13\alpha_2 - \lambda = 0 \\ \alpha_1 - \alpha_2 = 0 \end{cases}$$

de donde $\alpha_1 = \alpha_2 = 2$ y $\lambda = -5$

$$\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i (s_i) = (2)(-1)\begin{pmatrix}2\\3\end{pmatrix} + (2)(1)\begin{pmatrix}2\\2\end{pmatrix} = \begin{pmatrix}0\\-2\end{pmatrix}$$

$$\hat{w}_0 = 1 - ((s_1)^T \hat{\mathbf{w}}) = 1 - (2,2)\begin{pmatrix}0\\-2\end{pmatrix} = 1 + 4 = 5$$

y la ecuación del hiperplano separador es

$0x_1$-$2x_2$+5=0 , o lo que es igual

$x_2 = 5/2 = 2{,}5$

c) El patrón de coordenadas $(1,5, 2)^T$ al ser el valor de x_2 menor de 2,5 se clasifica en la clase positiva luego se clasifica bien.

Con los patrones de la clase positiva el (2,2) y de la negativa el (2,3) y el (3,2) tenemos las ecuaciones

$$\max\ \alpha_1+\alpha_2+\alpha_3-\ \frac{1}{2}\begin{pmatrix}8\alpha_1^2-10\alpha_1\alpha_2-10\alpha_1\alpha_3-10\alpha_2\alpha_1+13\alpha_2^2\\+12\alpha_2\alpha_3-10\alpha_3\alpha_1+12\alpha_3\alpha_2+13\alpha_3^2\end{pmatrix}$$

s.a. $-\alpha_1+\alpha_2+\alpha_3=0,\ \alpha_1\geq 0,\ \alpha_2\geq 0$

derivando con respecto a $\alpha_1,\ \alpha_2,\ \alpha_3$ y λ, tenemos

$$\begin{cases}1-8\alpha_1+10\alpha_2+10\alpha_2-\lambda=0\\1+10\alpha_1-13\alpha_2-12\alpha_3+\lambda=0\\1+1+\lambda=0\\-\alpha_1+\alpha_2+\alpha_3=0\end{cases}$$

de donde $\alpha_1=4, \alpha_2=\alpha_3=2$ y $\lambda=9$

$$\hat{\mathbf{w}}=\sum_{i\in Sop}\hat{\alpha}_i y_i(s_i)=(4)(1)\begin{pmatrix}2\\2\end{pmatrix}+(2)(-1)\begin{pmatrix}3\\2\end{pmatrix}+(2)(-1)\begin{pmatrix}2\\3\end{pmatrix}=\begin{pmatrix}-2\\-2\end{pmatrix}$$

$$\hat{w}_0=1-(2,2)\begin{pmatrix}-2\\-2\end{pmatrix}=9$$

y la ecuación del hiperplano separador es

$-2x_1-2x_2+9=0$, luego el patrón (1,5 2) pertenece a la clase positiva

y está bien clasificado