INTRODUCCION A LOS MODELOS COMPUTACIONALES 7 septiembre 2018

Alumno/a D...

Cuestiones.- A) (3 puntos)

i) (**1 punto**) Explique brevemente el sentido que tienen los filtros 3x3 y 1x1 en redes neuronales convolucionales. ¿Qué efecto producen?

Solución.- Filtros 3x3

0	**-1**	**0**
-1	**8**	**-1**
0	**-1**	**0**

Permiten arquitecturas más profundas. Tamaño mínimo necesario para aprender conceptos de horizontalidad, verticalidad, binarización de objetos.

Menos parámetros para el mismo campo receptivo

Campo receptivo para una pila de 3 filtros de 3x3 = 1 de 7x7

Si tenemos C entradas y C canales de salida

1 capa de un filtro 3x3 tiene un #parámetros = $3^2 \times C^2 \ aumenta\ como\ O\left(n^2\right)$

3 capas de tres filtros de 3x3, tiene un #parámetros= $3 * (3^2 \times C^2) = 27(C^2)$

1 capa con un filtro de 7x7, tiene un #parámetros= $7^2 \times C^2 = 49(C^2)$

Ventaja agregada: Más ReLU . Regularización

Filtros 1 x 1

Aumentan la no linealidad sin afectar el campo receptivo

Cuando #ip channels == #op channels: Proyección en el espacio de la misma dimensión

Otra perspectiva: Completamente conectado con el peso compartido

Utilizados en Network In Network, NIN, y en GoogLeNET

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

(iii) (**1 punto**) Las redes neuronales RBF son de tipo local o *kernel*. Ponga un ejemplo de red RBF con tres variables de entrada, un nodo en capa oculta y uno de salida. Implemente la fórmula de un clasificador binario con una red RBF como la anterior.

B) (2,5 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.

c) (**0.5 puntos**) Comenzando en el (mismo) estado [-1, 1, 1, 1, -1], calcule el siguiente estado usando

b) (**1 punto**) Comenzando en el estado [-1, 1, 1, 1, -1], calcule el flujo desde este estado al estado estable usando actualizaciones asincrónicas.

actualizaciones síncronas.

d) **(0.5 puntos)** ¿Cómo se calcula su función de energía?

Solución.- b) De forma asíncrona

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

$\begin{pmatrix}-1 & 1 & -1 & 1 & 1\end{pmatrix}$

Actualizamos la 5^a *componente*

$$\begin{pmatrix}-1 & 1 & 1 & 1 & 1\end{pmatrix}\begin{pmatrix}0.0 & -0.1 & 0.1 & -0.1 & -0.1\\ -0.1 & 0.0 & -0.1 & 0.1 & 0.1\\ 0.1 & -0.1 & 0.0 & -0.1 & -0.1\\ -0.1 & 0.1 & -0.1 & 0.0 & 0.1\\ -0.1 & 0.1 & -0.1 & 0.1 & 0.0\end{pmatrix}=\begin{pmatrix}-0.2 & 0.2 & -0.4 & 0.2 & 0.2\end{pmatrix}$$, luego se pasa al

$\begin{pmatrix}-1 & 1 & -1 & 1 & 1\end{pmatrix}$

Actualizamos la 3^acomponente

$$\begin{pmatrix}-1 & 1 & -1 & 1 & 1\end{pmatrix}\begin{pmatrix}0.0 & -0.1 & 0.1 & -0.1 & -0.1\\ -0.1 & 0.0 & -0.1 & 0.1 & 0.1\\ 0.1 & -0.1 & 0.0 & -0.1 & -0.1\\ -0.1 & 0.1 & -0.1 & 0.0 & 0.1\\ -0.1 & 0.1 & -0.1 & 0.1 & 0.0\end{pmatrix}=\begin{pmatrix}-0.4 & 0.4 & -0.4 & 0.4 & 0.4\end{pmatrix}$$, luego se pasa al

$\begin{pmatrix}-1 & 1 & -1 & 1 & 1\end{pmatrix}$ y hemos llegado a un estado estable

c) De forma síncrona tenemos

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

$\begin{pmatrix}-1 & 1 & -1 & 1 & 1\end{pmatrix}$

y de este estado se transita a si mismo, luego es este un estado estable

d) Al no cambiar de signo las componentes del vector

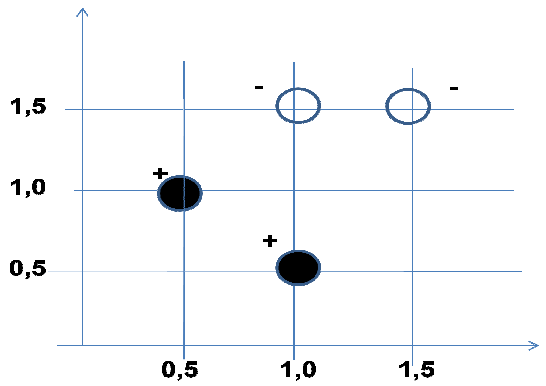
$$E(t)=-\frac{1}{2}\sum_{i,j} w_{ij}s_i s_j+\sum_i s_i\Theta_i$$

El vector x está caracterizado por todos blancos, esto es todas las unidades están activadas. Así, s_i es +1 para todo i y siempre será en este caso $\sum_j w_{ij} s_j > \Theta_i$ entonces s_i sigue siendo +1, para todo i. Si suponemos que Θ=0, entonces

$$E(t) = -\frac{1}{2}\sum_{i,j} w_{ij} s_i s_j = -\frac{1}{2}\sum_{i=1}^{5}\sum_{j=1}^{5} w_{ij} = -\frac{1}{2}(w_{11} + ... + w_{55}) =$$

$$= -\frac{1}{2}(-0.4 + 0.4 - 0.4 + 0.4 + 0.4) = -0.2$$

Ejercicio 2.- (2.5 puntos) Considere los cuatro vectores bidimensionales linealmente separables de la siguiente figura. Encuentre el SVM lineal que separa de manera óptima las clases al maximizar el margen.



Solución Los puntos que son vectores soporte son los positivos (1, 0,5) y el (0,5, 1) y el hiperplano de margen H^+ es la línea que pasa por los dos puntos positivos. El hiperplano de margen H^- es la recta que pasa por el punto negativo (1, 1,5) y es paralela a H^+. La función de decisión es la recta que está entre H^+ y H^-. Esta recta tiene por ecuación $-x_1 + 2 = x_2$.

Tenemos la clase positiva formada por los vectores soporte (0.5 ,1) y (1, 0.5) La clase negativa está formada por el vector soporte (1, 1.5)

La ecuación del hiperplano es $\mathbf{w}^T.\mathbf{x} + w_0 = 0$, o lo que es lo mismo $w_1x_1+w_2x_2+w_0=0$

A) Primal

Si sustituimos los vectores soporte en los hiperplanos positivo y negativo

H +: $\langle w \bullet x^+ \rangle + w_0 = 1$

H ⁻: $\langle w \bullet x^- \rangle + w_0 = -1$

tenemos las ecuaciones

$0.5w_1+w_2+w_0=1$

$w_1+0.5w_2+w_0=1$

$w_1+1.5w_2+w_0=-1$

y operando adecuadamente $w_1=-2$; $w_2=-2$ y $w_0= 4$, por lo que la ecuación del hiperplano separador es

$x_1+x_2-2=0$

B) Dual

$$\max \sum_{i=1} \alpha_i - \frac{1}{2} \sum_{i,j=1} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) = \max \sum_{i=1} \alpha_i - \frac{1}{2} \boldsymbol{\alpha}^{\mathbf{T}} \mathbf{Q} \boldsymbol{\alpha}$$

$$s.a. \ \sum_{i=1} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \text{ i=1,...,n}$$

$$\hat{\mathbf{w}} = \sum_{i=1} \hat{\alpha}_i y_i \mathbf{x_i} = \sum_{i \in Sop} \hat{\alpha}_i y_i \mathbf{x_i}$$

$$\hat{w}_{\mathbf{0}} = 1\text{-} \sum_{j \in Sop} \hat{\alpha}_j y_j (\mathbf{x_j^T x_i}), \quad \text{con } \mathbf{x_i} \in \omega_1 \ y \ \ \hat{\boldsymbol{\alpha}}_{\mathbf{i}} > 0$$

Ahora la función a maximizar es

$$\text{L}(\alpha_1, \alpha_2, \alpha_3, \lambda; \ \mathbf{x_1, x_2, x_3}) =$$

$$= \alpha_1 + \alpha_2 + \alpha_3 - \ \ \frac{1}{2} \left(\begin{array}{l} \alpha_1^2 (\mathbf{x_1^T . x_1}) + \alpha_1 \alpha_2 (\mathbf{x_1^T . x_2}) - \alpha_1 \alpha_3 (\mathbf{x_1^T . x_3}) + \alpha_2 \alpha_1 (\mathbf{x_2^T . x_1}) + \alpha_2^2 (\mathbf{x_2^T . x_2}) - \alpha_2 \alpha_3 (\mathbf{x_2^T . x_3}) + \\ -\alpha_3 \alpha_1 (\mathbf{x_3^T . x_1}) - \alpha_3 \alpha_2 (\mathbf{x_3^T . x_2}) + \alpha_3^2 (\mathbf{x_3^T . x_3}) - \lambda(\alpha_1 + \alpha_2 - \alpha_3) \end{array} \right)$$

sustituyendo los valores tenemos

$$\text{L}(.) = \alpha_1 + \alpha_2 + \alpha_3 - \ \ \frac{1}{2} \left(\begin{array}{l} \alpha_1^2 (\frac{5}{4}) + \alpha_1 \alpha_2 (1) - \alpha_1 \alpha_3 (2) + \alpha_2 \alpha_1 (1) + \alpha_2^2 (\frac{5}{4}) - \alpha_2 \alpha_3 (\frac{7}{4}) + \\ -\alpha_3 \alpha_1 (2) - \alpha_3 \alpha_2 (\frac{7}{4}) + \alpha_3^2 (\frac{13}{4}) - \lambda(\alpha_1 + \alpha_2 - \alpha_3) \end{array} \right)$$

Derivando con respecto a los α_i y λ tenemos

$$\frac{\partial L}{\partial \alpha_1} = 0, \text{ esto es}$$

$$1 - \frac{5}{4} \alpha_1 - \alpha_2 + 2\alpha_3 - \lambda = 0$$

$$\frac{\partial L}{\partial \alpha_2} = 0, \text{ esto es}$$

$$1 - \alpha_1 - \frac{5}{4} \alpha_2 + \frac{7}{4} \alpha_3 - \lambda = 0$$

$$\frac{\partial L}{\partial \alpha_3} = 0, \text{ esto es}$$

$$1 + 2\alpha_1 + \frac{7}{4} \alpha_2 - \frac{13}{4} \alpha_3 + \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = -\alpha_1 - \alpha_2 + \alpha_3 = 0, \text{ luego}$$

$$\alpha_3 = \alpha_1 + \alpha_2$$

Resolviendo el sistema de 4 ecuaciones con 4 incognitas tenemos

$\alpha_1 = 4;\ \alpha_2 = 0$ y $\alpha_3 = 4$

$$\hat{\mathbf{w}} = \sum_{i \in Sop} \hat{\alpha}_i y_i \mathbf{x_i} = (-2,-2)$$

$$\hat{w}_\mathbf{0} = 1\text{-} \sum_{j \in Sop} \hat{\alpha}_j y_j (\mathbf{x_j^T x_i}), \quad \text{con } \mathbf{x_i} \in \omega_1 \ y \ \ \hat{\alpha}_i > 0$$

$$\hat{w}_\mathbf{0} = 4$$

La ecuación es $-2x_1-2x_2+4=0$, esto es $-x_1-x_2+2=0$

Ejercicio 3.- (2 puntos) Dada la información del conjunto de entrenamiento que se muestra en la tabla (Comprar una Computadora).

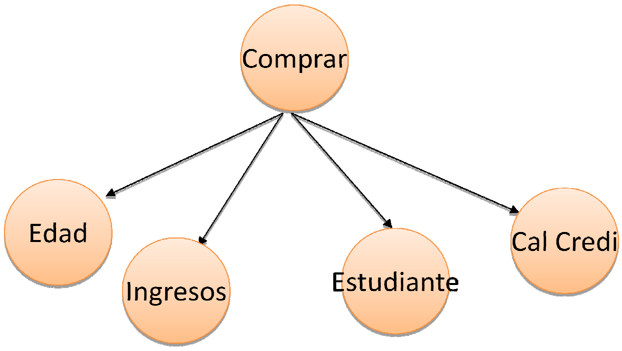
(0,5 puntos) Construir el grafo Bayesiano con los eventos Edad, Ingresos, Profesor, Cal Credi y Clases.

(0,5 puntos) Donde se usan las probabilidades condicionalmente independientes.

(1 punto) Predecir la clase del siguiente ejemplo usando un clasificador Naïve Bayes: edad < = 30, ingresos = medios, profesor = Sí, calificación de crédito= normal

ID	Edad	Ingresos	Profesor	Cal Crédito	Clase/Comprar
1	<=30	altos	No	normal	No
2	<=30	altos	No	excelente	No
3	[31,40]	altos	No	normal	Si
4	>40	medios	No	normal	Si
5	>40	bajos	Si	normal	Si
6	>40	bajos	Si	excelente	No
7	[31,40]	bajos	Si	excelente	No
8	<=30	medios	No	normal	No
9	<=30	bajos	Si	normal	Si
10	>40	medios	Si	normal	Si
11	<=30	medios	Si	excelente	Si
12	[31,40]	medios	No	excelente	Si
13	[31,40]	altos	No	normal	Si
14	>40	medios	No	excelente	No
15	[31,40]	bajos	No	excelente	Si
16	[31,40]	bajos	No	normal	No

Solución.-



a)

b) Los sucesos Edad, Ingresos, Profesor y Calificación de Crédito son condicionalmente independientes dado el suceso Comprar

c) Calculamos la verosimilitud del suceso (Edad<=30, Ingresos = medios, Profesor=Si, Cal Crédito=normal, Comprar=Si)

P(Edad<=30, Ingresos = medios, Profesor=Si, Cal Crédito =normal, Comprar= Si)=
= P(Edad<=30/Comprar=Si)* P(Ingresos = Medios/Comprar=Si)* P(Profesor =Si/Comprar=Si)* P(Cal Crédito =normal/Comprar=Si)* P(Comprar= Si)

Tenemos, utilizando la tabla con los 16 patrones
P(Edad<=30/Comprar=Si)=2/9; P(Ingresos = Medios/Comprar=Si)=4/9
P(Profesor=Si/Comprar=Si)=4/9; P(Cal Crédito =normal/ Comprar=Si)=6/9
P(Comprar=Si)=9/16

Luego P(Edad<=30, Ingresos = medios, Profesor=Si, Cal Crédito =normal, Comprar=Si)=(2/9)(4/9)(4/9)(6/9)(9/16)=12/729=0,0165

Análogamente
P(Edad<=30, Ingresos = medios, Profesor=Si, Cal Crédito =normal, Comprar=No)=
= P(Edad<=30/Comprar=No)* P(Ingresos = Medios/Comprar=No)* P(Profesor=Si/ Comprar=No) * P(Cal Crédito =normal/Comprar=No)* P(Comprar=No)=
=(3/7)(2/7)(2/7)(3/7)(7/16)=0,0065

P(Comprar=Si/(Edad<=30, Ingresos = medios, Profesor=Si, Cal Crédito=normal))=
=0,0165/(0,0165+0,0065)=0,7174
P(Comprar=No/(Edad<=30, Ingresos = medios, Profesor=Si, Cal Crédito=normal))=
=0,0065/(0,0165+0,0065)=0,2826

Se deduce que dados los valores establecidos para el nuevo patrón, la clase de salida es Comprar=Si al tener la mayor probabilidad