Ejercicios de Aprendizaje Automático

9 de diciembre de 2023

1. Evaluación

Ejercicio1. Proporciona una matriz de confusión para el problema de predecir si la inversión de una determinada empresa debe ser *alta*, *baja* o *media* de modo que el Sensibilidad de la clase *alta* sea del 75 %, el número de Falsos Positivos (FP) de la clase *baja* sea 50. Calcula la Especificidad de la clase *media*. NOTA: En la matriz de confusión no dejes ninguna casilla a 0.

Solución Ejercicio 1

Para este ejercicio hay múltiples soluciones, una posible es esta. En negrita están los valores según las restricciones del enunciado, el resto son valores libres para que no haya celdas a cero.

		Real		
		alta	*media*	*baja*
Pred.	*alta*	**75**	3	11
	media	**10**	27	5
	baja	**15**	**35**	16

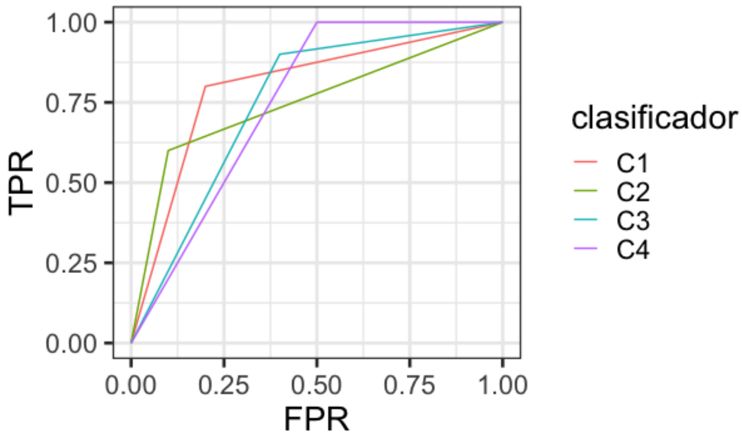
$Especificidad(media) = (75 + 11 + 15 + 16)/(75 + 11 + 15 + 16 + 10 + 5) = 0{,}886$

Ejercicio2. ¿Qué clasificador escogerías y por qué en base a los siguientes resultados?

	C1	C2	C3	C4
Sensibilidad	0.8	0.6	0.9	1.0
Especificidad	0.8	0.9	0.6	0.5

Solución Ejercicio 2

Nos basamos en la curva ROC, el mejor clasificador será el que tenga su coordenada (FPR, TFP) más cerca del punto (0,1) o, de otra forma, el que deje una mayor área bajo su curva. Por tanto, el mejor clasificador es C1.



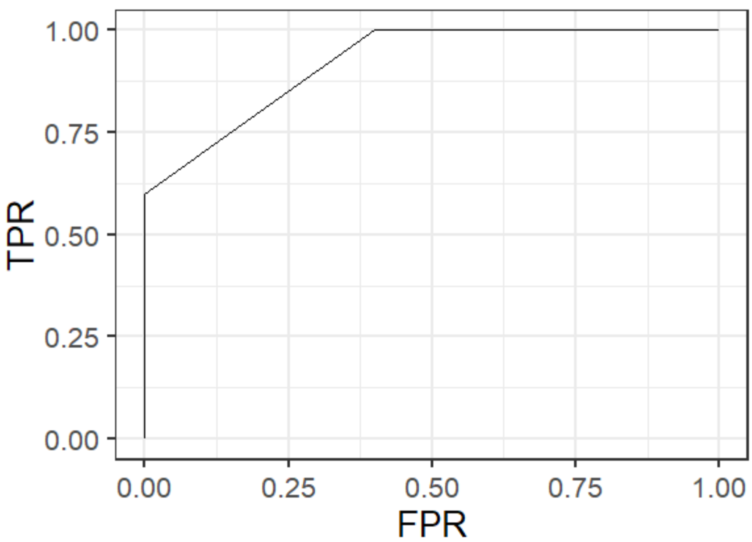
Ejercicio3. Dado un clasificador con las siguientes predicciones:

Real	pos	neg	pos	pos	neg	neg	neg	pos	neg	pos
Score(pos)	0.6	0.1	0.4	0.8	0.6	0.3	0.4	0.7	0.2	0.7

Obtén la curva ROC. ¿Se podría decir que es un buen clasificador?

Solución Ejercicio 3

Coordenadas FPR=(0.0,0.0,0.0,0.2,0.4,0.6,0.8,1.0); TPR=(0.0,0.2,0.6,0.8,1.0,1.0,1.0,1.0)



Ejercicio4.

Supongamos que tenemos 50 libros de minería de datos de un total de 200 libros en una biblioteca. Si un clasificador predice que 10 libros son de minería de datos, pero solo 5 de ellos lo son realmente, ¿Cuál es la Sensibilidad y la Especificidad?

Solución Ejercicio 4

Sensibilidad=5/50=0.1; Especificidad=(150-5)/150=0.9667

2. Árboles de decisión

Ejercicio5. Genera un árbol de decisión para las siguientes funciones booleanas:

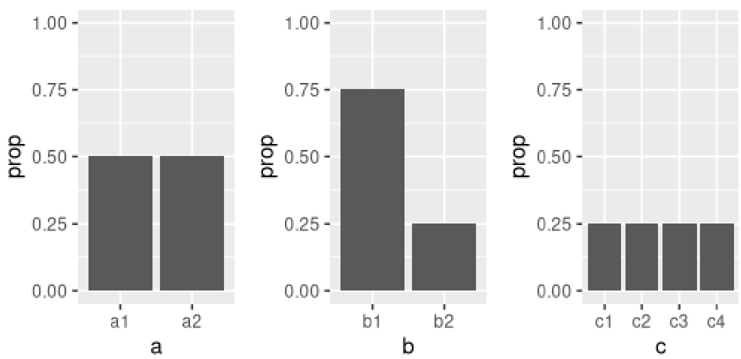
1. $X_1 \wedge \neg X_2$
2. $X_1 \vee (X_2 \wedge X_3)$

Solución Ejercicio 5

En base a las tablas de verdad buscamos la variable más relacionada con el resultado.

1. X_1=F: F
X_1=T
| X_2=F: T
| X_2=T: F
2. X_1=F
| X_2=F: F
| X_2=T
| | X_3=F: F
| | X_3=T: T
X_1=T: T

Ejercicio6. Dadas las variables a, que toma los valores $\{a1, a2\}$, b, que toma los valores $\{b1, b2\}$, y c que toma los valores $\{c1, c2, c3, c4\}$, responde **razonadamente** a la pregunta ¿qué variable tiene la mayor entropía y cual tiene la menor?



Solución Ejercicio 6

La variable con menor entropía es b porque uno de sus valores (v1) es mucho más frecuente que el otro. Las otras dos variables tienen valores equiprobables, pero como c tiene más valores hay más incertidumbre y por tanto esta el la variable con mayor entropía.

Ejercicio7. Dado un conjunto de entrenamiento con 150 ejemplos y tres variables, la variable Color que puede tomar los valores Rojo, Azul o Amarillo, la variable Tamaño que toma los valores Grande o Pequeño y la variable a predecir, COMPRA, que toma los valores Sí o No. Suponiendo que:

- Hay 50 ejemplos donde Color=Rojo, 50 para los que Color=Azul y 50 para los que Color=Amarillo.
- Hay 20 ejemplos para los que Tamaño=Grande y 130 para los que Tamaño=Pequeño.
- Hay 100 ejemplos de la clase COMPRA= Sí y 50 de la clase COMPRA=No
- IG(Color)=IG(Tamaño)

1. **Razona** qué variable selecciona el algoritmo C4.5 como raíz
2. ¿Es el 70 % un porcentaje de acierto aceptable para el método C4.5?

Solución Ejercicio 7

1. El algoritmo C4.5 usa como métrica para decidir que variable se usa como próximo nodo del árbol el Gain Ratio (GR), como esta métrica se basa en dividir la ganancia de información entre SplitInfo y sabiendo que la ganancia de información de las dos variables es iguales, todo se reduce a calcular el valor de SplitInfo.

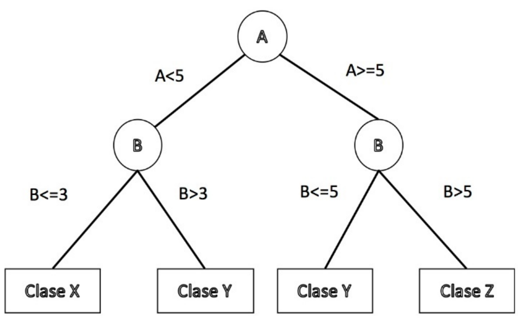
$$\text{SplitInfo(Color)} = (-0{,}33 \cdot \log_2 0{,}33) \cdot 3 = 1{,}58$$

$$\text{SplitInfo(Tamaño)} = -0{,}13 \cdot \log_2 0{,}13 - 0{,}86 \cdot \log_2 0{,}86 = 0{,}57$$

   El algoritmo C4.5 seleccionara la variable Tamaño por delante de la variable Color.

2. La frecuencia de la clase más frecuente es 2 tercios y este es el valor de porcentaje de acierto a batir por cualquier modelo. Un modelo con un 70 % de acierto es mejor que nada, con lo que tendríamos un 67 %, por tanto es aceptable.

Ejercicio8. Dado el siguiente conjuntos de datos y el siguiente árbol de decisión, asumiendo que los ejemplos {EJ1, ...EJ6} forman el conjunto de entrenamiento y {EJ7, ...EJ10} el conjunto de test



Ejemplo	A	B	CLASE
EJ1	3	3	Y
EJ2	2	2	X
EJ3	5	2	X
EJ4	7	6	Y
EJ5	7	7	Y
EJ6	8	6	Z
EJ7	2	4	Y
EJ8	8	9	Z
EJ9	5	5	Z
EJ10	1	4	X

1. ¿Cuál es el error en entrenamiento? ¿Y en test? Utilizando solo los datos de error en entrenamiento y test, responde **razonadamente** si existe sobreajuste o no.
2. Calcula la matriz de confusión asociada al modelo descrito por el árbol de decisión para todos las instancias.
3. Calcula los valores de TP y TN para cada una de las clases.

Solución Ejercicio 8

1. En el conjunto de entrenamiento se acierta en los ejemplos 2 y 6 mientras que se falla en el resto. Esto da una tasa de error de 4/6=67 %. Por otro lado, con los datos de test se acierta en los ejemplos 7 y 8 lo que da una tasa de error de 50 %. Dado que el modelo presenta mejor resultado con los datos de testeo no parece que exista sobreajuste.
2. La matriz de confusión para las 10 instancias es la siguiente:

		Real		
		X	Y	Z
Pred.	X	1	1	0
	Y	2	1	1
	Z	0	2	2

3. TP(X)=1; TN(X)=6; TP(Y)=1 ;TN(Y)=3 ;TP(Z)=2 ;TN(Z)=5

Ejercicio9. Dado un conjunto de datos con 7 variables binarias, Rain, Wind, Summer, Winter, Day y Night que toman el valor “yes” o “no” y Flight_Delay que toma los valores “Delayed” y “not Delayed”, considera la siguiente tabla de datos, que explica cómo se comporta la variable Flight_Delay con respecto a las otras 6.

Feature	Value = yes	Value = no
Rain	Delayed - 30, not Delayed - 10	Delayed - 10, not Delayed - 30
Wind	Delayed - 25, not Delayed - 15	Delayed - 15, not Delayed - 25
Summer	Delayed - 5, not Delayed - 35	Delayed - 35, not Delayed - 5
Winter	Delayed - 20, not Delayed - 10	Delayed - 20, not Delayed - 30
Day	Delayed - 20, not Delayed - 20	Delayed - 20, not Delayed - 20
Night	Delayed - 15, not Delayed - 10	Delayed - 25, not Delayed - 30

Si construimos un árbol de decisión utilizando C4.5, responde **razonadamente** a la pregunta ¿cuál es la variable raíz?

Solución Ejercicio 9

Hay que ver cuál es la variable con mayor valor de GainRatio que es la relación entre IG y SplitInfo. Se puede ver claramente que las variables Rain, Wind, Summer y Day son equiprobables y por lo tanto su valor de SplitInfo será de 1. De estas variables se puede ver de la tabla que la que mejor separa los valores de Delayed es Summer, así que no será necesario hacer los cálculos para las otras tres.

Partimos de las entropías: $H(Delayed) = 1; H(Delayed|Summer) = 0{,}54; H(Delayed|Winter) = 0{,}95; H(Delayed|Night) = 0{,}99$

Los valores de SplitInfo: $\text{SplitInfo}(Winter) = 0{,}95$; $\text{SplitInfo}(Night) = 0{,}90$

Por último calculamos GR:

$$\text{GR}(Delayed|Summer) = \frac{1 - 0{,}54}{1} = 0{,}46$$

$$\text{GR}(Delayed|Winter) = \frac{1 - 0{,}95}{0{,}95} = 0{,}05$$

$$\mathrm{GR}(Delayed|Night) = \frac{1 - 0{,}99}{0{,}90} = 0{,}01$$

La variable seleccionada por C4.5 será Summer.

Ejercicio10. Dado el conjunto de ejemplos descrito a continuación:

Ejemplo	X1	X2	X3	Clase
EJ1	0	250	36	**A**
EJ2	10	150	34	**B**
EJ3	4	20	1	**B**
EJ4	6	78	8	**B**
EJ5	2	90	10	**A**
EJ6	1	170	70	**A**
EJ7	6	200	45	**A**
EJ8	8	160	41	**B**
EJ9	10	180	38	**A**

1. Completa la siguiente tabla de entropías sobre la clase de las particiones asociadas a los atributos $X1$, $X2$, $X3$:

	$H(Clase)$	
	Sí	No
$X1 \leq 5$	0.811	0.971
$X2 \leq 160$	0.722	
$X3 \leq 40$		0.918

2. Sabiendo que $H(Clase) = 0{,}991$, y utilizando los datos de la tabla del apartado a), calcula el atributo más informativo según la **Ganancia de Información** (IG).

Solución Ejercicio 10

1. $H(Clase|X2 > 160) = 0$; $H(Clase|X3 \leq 40) = 1$
2. $$IG(Clase|X1) = 0{,}991 - \left(\frac{4}{9} \cdot 0{,}811 + \frac{5}{9} \cdot 0{,}971\right) = 0{,}091$$
$$IG(Clase|X2) = 0{,}991 - \left(\frac{5}{9} \cdot 0{,}722 + 0\right) = 0{,}590$$
$$IG(Clase|X3) = 0{,}991 - \left(\frac{6}{9} \cdot 1 + \frac{3}{9} \cdot 0{,}918\right) = 0{,}018$$
El atributo más informativo es X2.

Ejercicio11. Dado un conjunto de 15 ejemplos de entrenamiento con una variable X que toma los valores a y b, y una variable a predecir Y con 5 ejemplos de la clase positiva y 10 de la negativa.

1. Calcula la entropía y el índice de Gini de la variable Y.
2. Si a partir de ese conjunto de datos construyes un árbol de decisión sin ningún nodo, ¿qué clase le asignarías para minimizar el error de entrenamiento? Responde razonadamente.

3. Sabiendo que cuando $X = a$ hay 5 ejemplos de la clase positiva y 5 de la negativa, y que cuando $X = b$ todos los ejemplos son de la clase negativa, utiliza esta variable para construir un árbol de decisión. ¿Es este árbol mejor en términos de error que el obtenido en el apartado 2? Responde razonadamente.

Solución Ejercicio 11

1.
$$Gini(Y) = 1 - (5/15)^2 - (10/15)^2$$

$$H(Y) = -5/15 \log_2(5/15) - 10/15 \log_2(10/15)$$

2. Se asignaría la clase mayoritaria, esto es la negativa. El error es 5/15
3. No es mejor dado que el error obtenido es el mismo, sigue clasificando mal 5 de los 15 ejemplos

3. Vecinos más cercanos

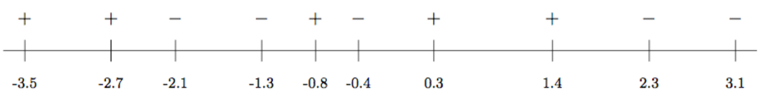
Ejercicio12. Dado el conjunto de ejemplos del ejercicio 10 considera el conjunto $\{EJ1, EJ2, \ldots, EJ8\}$ como entrenamiento, realiza los cálculos para clasificar el ejemplo $EJ9$ utilizando 3NN con la distancia euclídea. ¿Es un acierto?

Solución Ejercicio 12

$$dist(EJ9, EJ1) = \sqrt{10^2 + 70^2 + 2^2} = 70{,}74$$
$$dist(EJ9, EJ2) = 30{,}27$$
$$dist(EJ9, EJ3) = 164{,}33$$
$$dist(EJ9, EJ4) = 106{,}4$$
$$dist(EJ9, EJ5) = 95{,}54$$
$$dist(EJ9, EJ6) = 34{,}71$$
$$dist(EJ9, EJ7) = 21{,}56$$
$$dist(EJ9, EJ8) = 20{,}32$$

Con estas distancias, los tres vecinos más cercanos son $EJ2$, $EJ7$ y $EJ8$. Como dos de ellos tienen clase B esa sería la predicción y sería un fallo.

Ejercicio13. Dado el siguiente conjunto de ejemplos con un atributo real y una variable objetivo que toma dos valores (+,-)



x	-3.5	-2.7	-2.1	-1.3	-0.8	-0.4	0.3	1.4	2.3	3.1
y	+	+	−	−	+	−	+	+	−	−

1. Utilizando como conjunto de test los ejemplos (-3.5, +), (-0.8, +) y (-0.4,-) y como conjunto de entrenamiento el resto, calcula **razonadamente** el porcentaje de error que se obtiene con K-NN, K=3.
2. Suponiendo que entrenamos este clasificador con las 10 instancias usando validación cruzada de 5 cajas, **da una posible partición (solo una)** y calcula el error cometido en **una** iteración de la validación cruzada.

Solución Ejercicio 13

1. Para -3.5 los vecinos son -2.7, -2.1 y -1.3, la predicción es '-' y es un error. Para -0.8 los vecinos son -1.3, -2.1 y 0.3, la predicción es '-' y es un error. Para -0.4 los vecinos son -1.3, 0.3 y -2.1, la predicción es '-' y es un acierto.

   Esto da una tasa de error del 66 %.

2. Hay tanta soluciones como combinaciones de 10 ejemplos tomados de 2 en 2. Así que elijo 1.

   Caja1: (-3.5,+), (-2.7,+)

   Caja2: (-2.1,-), (-1.3,-)

   Caja3: (-0.8,+), (-0.4,-)

   Caja4: (0.3,+), (1.4,+)

   Caja5: (2.3,-), (3.1,-)

   Seleccionamos caja 1 vara validar. Entrenamos con los ejemplos de las cajas 2, 3, 4 y 5

	VECINO1	VECINO2	VECINO3	PREDICCIÓN	
(-3.5,+)	(-2.1,-)	(-1.3,-)	(-0.8,+)	-	FALLO
(-2.7,+)	(-2.1,-)	(-1.3,-)	(-0.8,+)	-	FALLO

Seleccionamos caja 2 vara validar. Entrenamos con los ejemplos de las cajas 1, 3, 4 y 5

	VECINO1	VECINO2	VECINO3	PREDICCIÓN	
(-2.1,-)	(-3.5,+)	(-2.7,+)	(-0.8,+)	+	FALLO
(-1.3,-)	(-0.8,+)	(-0.4,-)	(0.3, +)	+	FALLO

Seleccionamos caja 3 vara validar. Entrenamos con los ejemplos de las cajas 1, 2, 4 y 5

	VECINO1	VECINO2	VECINO3	PREDICCIÓN	
(-0.8,+)	(-1.3,-)	(0.3, +)	(-2.1,-)	-	FALLO
(-0.4, +)	(-1.3,-)	(0.3,+)	(-2.1, +)	+	ACIERTO

Seleccionamos caja 4 vara validar. Entrenamos con los ejemplos de las cajas 1, 2, 3 y 5

	VECINO1	VECINO2	VECINO3	PREDICCIÓN	
(0.3,+)	(-0.4, +)	(-0.8,+)	(-1.3,-)	+	ACIERTO
(1.4, +)	(2.3,-)	(-0.4,-)	(3.1,-)	-	FALLO

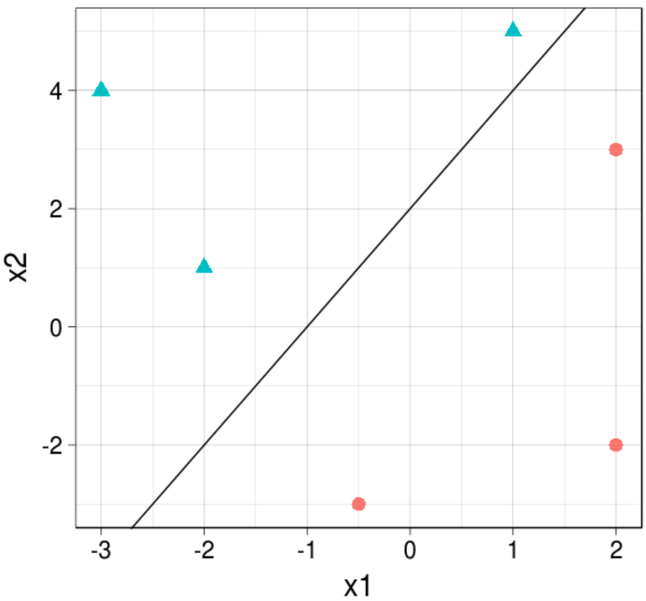
Seleccionamos caja 5 vara validar. Entrenamos con los ejemplos de las cajas 1, 2, 3 y 4

	VECINO1	VECINO2	VECINO3	PREDICCIÓN	
(2.3,-)	(-0.4, -)	(0.3,+)	(1.4,+)	+	FALLO
(3.1, -)	(-0.4, -)	(0.3,+)	(1.4,+)	+	FALLO

Como se puede observar solo acierta en 2 de los 10 ejemplos.

4. Redes neuronales

Ejercicio14. Calcula los valores de los pesos w_0, w_1, w_2 para el perceptrón cuya frontera de decisión se ilustra en la siguiente figura.



Solución Ejercicio 14

Buscamos la recta que muestra la figura y esta es $x2 = 2x1 + 2$. Dado que el valor umbral del perceptrón es 0, dejamos la ecuación de la recta con un 0 a un lado del igual, por ejemplo $2x1 - x2 + 2 = 0$ y de aquí tenemos los valores de los pesos asociados a cada variable: $w_0 = 2$; $w_1 = 2$; $w_2 = -1$. Otra alternativa con -2x1+x2-2: $w_0 = -2$; $w_1 = -2$; $w_2 = 1$

Ejercicio15. Diseña un perceptrón para la función booleana $A \wedge \neg B$

Solución Ejercicio 15

En vez de usar el algoritmo de aprendizaje del perceptrón podemos ver en la tabla de verdad de la función si valores de A y B implican una salida alta (1) o baja (-1) para saber si los sus pesos deben ser altos o bajos. Otra opción es hacer una gráfica con los cuatro puntos y resolver el problema como el ejercicio anterior. En cualquier caso una solución puede ser $w_0 = -0{,}5$; $w_1 = 1$; $w_2 = -1$

Ejercicio16. Construye a mano una red de perceptrones para la función booleana XOR.

Solución Ejercicio 16

Hay que tener en cuenta que la función XOR se puede expresar como $(A \wedge \neg B) \vee (\neg A \wedge B)$, así vemos que podemos generar la red con tres perceptrones dos que hagan los resultados entre paréntesis, a partir del ejercicio anterior, y la salida de estos dos se combina con un último perceptrón que codifique la función OR.

Ejercicio17. Considera una red neuronal con **función de activación lineal**. Tiene dos entradas, una neurona oculta y una unidad de salida. Habrá un total de cinco pesos (w_{ca}, w_{cb}, w_{c0}, w_{dc}, w_{d0}) que vamos a inicializar todos a 0,1. Calcula la actualización de los pesos utilizando el algoritmo de propagación hacía atrás con una tasa de aprendizaje $\eta = 0{,}3$ y los siguientes ejemplos de entrenamiento:

a	b	d
1	0	1

NOTA: Solo es necesario ejecutar una pasada del algoritmo para esta instancia, no se pide llegar a unos valores de convergencia. Al usar una función de activación lineal hay que tener en cuenta las expresiones de δ_k y δ_h adecuadas a este caso, que son:

$$\delta_k = (y_k - o_k)$$

$$\delta_h = \sum_k \delta_k w_{\text{kh}}$$

Solución Ejercicio 17

Presentamos el ejemplo (1,0)

$$o_c = 1 \cdot w_{\text{ca}} + 0 \cdot w_{\text{cb}} + 1 \cdot w_{c0} = 0{,}1 + 0{,}1 = 0{,}2$$

$$o = o_c \cdot w_{\text{dc}} + 1 \cdot w_{d0} = 0{,}2 \cdot 0{,}1 + 0{,}1 = 0{,}12$$

$$w_{\text{dc}} = w_{\text{dc}} + \eta\ (d - o) \cdot o_c = 0{,}1 + 0{,}3(1 - 0{,}12) \cdot 0{,}2 = 0{,}1528$$

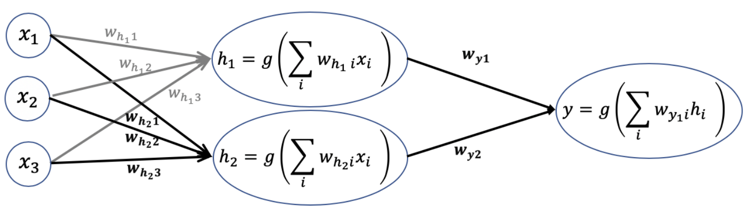
$$w_{d0} = w_{d0} + \eta\ (d - o) \cdot 1 = 0{,}1 + 0{,}3(1 - 0{,}12) \cdot 1 = 0{,}364$$

$$w_{\text{ca}} = w_{\text{ca}} + \eta\ \cdot w_{\text{dc}} \cdot (d - o)x^a = 0{,}1 + 0{,}3 \cdot 0{,}1 \cdot (1 - 0{,}12) \cdot 1 = 0,126$$

$$w_{\text{cb}} = w_{\text{cb}} + \eta\ \cdot w_{\text{dc}} \cdot (d - o)x^b = 0{,}1 + 0{,}3 \cdot 0{,}1 \cdot (1 - 0{,}12) \cdot 0 = 0,13$$

$$w_{c0} = w_{c0} + \eta\ \cdot w_{\text{dc}} \cdot (d - o)1 = 0{,}1 + 0{,}3 \cdot 0{,}1 \cdot (1 - 0{,}12) \cdot 1 = 0,126$$

Ejercicio18. Considera la siguiente red neuronal, con $\mathbf{g}(\mathbf{x}) = \dfrac{\mathbf{1}}{\mathbf{1} - \mathbf{e}^{-\mathbf{x}}}$. Considera el ejemplo de entrenamiento $(x_1, x_2, x_3,\ y) = (0, 1, 1, 2)$, de tal modo que cuando se presenta este ejemplo a la red se obtiene una salida con valor 1.1. Calcula la actualización del peso $\mathbf{w_{h_1 2}}$utilizando el algoritmo de propagación hacia atrás. El valor de la tasa de aprendizaje y la inicialización de pesos aparecen debajo de la red neuronal.



$w_{h_1 j} = 0.1,\quad w_{h_2 j} = 0.2,\quad w_{y j} = 0.3,\quad \eta = 0.2$

Solución Ejercicio 18

En primer lugar tenemos que calcular la salida de las neuronas:

$$o_{h_1} = g(w_{h_1 1}x_1 + w_{h_1 2}x_2 + w_{h_1 3}x_3) = g(0{,}1 * 0 + 0{,}1 * 1 + 0{,}1 * 1) = g(0{,}2) = \frac{1}{1 - e^{-0{,}2}} = 5{,}52$$

$$o_{h_2} = g(w_{h_2 1}x_1 + w_{h_2 2}x_2 + w_{h_2 3}x_3) = g(0{,}2 * 0 + 0{,}2 * 1 + 0{,}2 * 1) = g(0{,}4) = \frac{1}{1 - e^{-0{,}4}} = 3{,}03$$

$$o_y = g(w_{y1}o_{h_1} + w_{y1}o_{h_2}) = g(0{,}3 \cdot 5{,}52 + 0{,}3 \cdot 3{,}03) = g(2{,}565) = 1{,}1$$

Después los valores de delta:
$\delta_y = o_y(1 - o_y)(y - o_y) = 1{,}1(1 - 1{,}1)(2 - 1{,}1) = -0{,}099$
$\delta_{h_1} = o_{h_1}(1 - o_{h_1})w_{y_1}\delta y = 5{,}52 * (1 - 5{,}52) * 0{,}3 * (-0{,}099) = 0{,}74$
Por tanto,
$w_{h_1 2} = w_{h_1 2} + \eta x_2 \delta_{h_1}$ = 0,1+0.2*0.74*1=0.248

Ejercicio19. Considera un perceptrón y dos entradas x_1 y x_2 donde el valor de la entrada de sesgo es 1 y los pesos asignados a cada entrada son $(w_0 = 2, w_1 = 1, w_2 = -1)$. La función de activación de la neurona es la función signo, de manera que los ejemplos se clasifican como positivos si el valor en el hiperplano es positivo o cero y como negativo en otro caso.

1. Escribe la ecuación del hiperplano separador inicial e indica cual es la clasificación que daría para el punto (-1, 2).

2. Supongamos que el punto (-1, 0) es de la clase negativa. Dados los pesos anteriores aplica la regla del perceptrón para que se clasifique correctamente y da la ecuación del hiperplano resultante usando una tasa de aprendizaje de 0,5.

Solución Ejercicio 19

La ecuación del hiperplano de separación inicial viene dada por $h(x_1, x_2) = x_1 - x_2 + 2$

1. Evaluamos esta ecuación en el punto (-1,2), obteniendo $h(-1, 2) = -1 - 2 + 2 < 0$, así que el ejemplo se clasifica como negativo.

2. $h(-1, 0) = -1 + 2 = 1 > 0$. Luego el perceptrón diría que es de la clase positiva.

   Actualizamos pesos (el superíndice 1 indica los pesos de la iteración anterior)

$$w_1^2 = w_1^1 + \eta \cdot (y - h(-1, 0)) \cdot x_1 = 1 + 0{,}5 \cdot (-2) \cdot (-1) = 2$$

$$w_2^2 = -1 + 0{,}5 \cdot (-2) \cdot 0 = -1$$

$$w_0^2 = 2 + 0{,}5 \cdot (-2) \cdot 1 = 2 - 1 = 1$$

Ejercicio20. Considera una red neuronal con **función de activación lineal**. Tiene dos entradas (a, b), una neurona oculta (c) y una unidad de salida (d). Habrá un total de cinco pesos $(w_{ca}, w_{cb}, w_{c0}, w_{dc}, w_{d0})$ que vamos a inicializar todos a 0,1 y con una tasa de aprendizaje $\eta = 0{,}3$. Calcula el valor de los pesos utilizando el algoritmo de propagación hacía atrás para la siguiente instancia:

a	b	d
1	1	0

Recuerda que en este caso: $\delta_k = (y_k - o_k)$ y $\delta_h = \sum_k \delta_k w_{kh}$

Solución Ejercicio 20

$o_c = 0{,}1 \cdot 1 + 0{,}1 \cdot 1 + 0{,}1 \cdot 1 = 0{,}3$; $o_d = 0{,}1 \cdot 1 + 0{,}1 \cdot 0{,}3 = 0{,}13$
$\delta_d = 0 - 0{,}13 = -0{,}13$; $\delta_c = -0{,}13 \cdot 0{,}1 = -0{,}013$

$$w_{d0} = 0{,}1 + 0{,}3 \cdot (-0{,}13) \cdot 1 = 0{,}061$$

$$w_{dc} = 0{,}1 + 0{,}3 \cdot (-0{,}13) \cdot 0{,}3 = 0{,}0883$$

$$w_{c0} = 0{,}1 + 0{,}3 \cdot (-0{,}013) \cdot 1 = 0{,}0961$$

$$w_{ca} = 0{,}1 + 0{,}3 \cdot (-0{,}013) \cdot 1 = 0{,}0961$$

$$w_{cb} = 0{,}1 + 0{,}3 \cdot (-0{,}013) \cdot 1 = 0{,}0961$$