

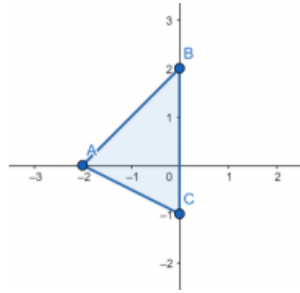
Tarea 9 Reconocimiento de Patrones

Yamil Ernesto Morfa Avalos.

1 de junio de 2021

B. Preguntas cortas

1. Construiremos a mano una red con neuronas cuya función de activación sea la función signo. Con el objetivo de identificar si un punto del plano (x_1, x_2) está o no en el triángulo:



Para estos construimos los datos y su respectivas etiquetas ($y = 1$ si está en el triángulo y $y = 0$ si no lo está). Estos se muestran en la figura siguiente.

Arquitectura:

Construiremos una red con una sola capa oculta con n neuronas, paso a paso funcionaria de la siguiente forma:

$$z^{[1](i)} = W^{[1]}x^{(i)} + b^{[1](i)}$$

donde $x^{(i)} \in \mathbb{R}^{2 \times 1}$ es una observación i de los datos, $W^{[1]} \in \mathbb{R}^{n \times 2}$, $b^{[1](i)} \in \mathbb{R}^{n \times 1}$ los parámetros a optimizar en esta capa. Luego calculamos la función de activación para estos valores:

$$a^{[1](i)} = \text{sign}(z^{[1](i)})$$

donde

$$\text{sign}(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

notemos que $\frac{d \text{sign}}{dx}$ no está definida para $x = 0$:

$$\lim_{t \rightarrow 0^-} \frac{\text{sign}(t) - \text{sign}(0)}{t} = +\infty; \quad \lim_{t \rightarrow 0^+} \frac{\text{sign}(t) - \text{sign}(0)}{t} = 0$$

Una manera de abordar este problema podría ser tomar $\frac{d \text{sign}}{dx} = \begin{cases} 0 & x \neq 0 \\ c \neq 0 & x = 0 \end{cases}$.

Análogamente la capa de salida entonces será:

$$z^{[2](i)} = W^{[2]}a^{[1](i)} + b^{[2](i)}$$

$$a^{[2](i)} = \text{sign} \left(z^{[2](i)} \right)$$

$$\hat{y}^{(i)} = a^{[2](i)}$$

Con la función de costo entropía cruzada:

$$J = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log \left(\hat{y}^{(i)} \right) + \left(1 - y^{(i)} \right) \log \left(1 - \hat{y}^{(i)} \right) \right]$$

Gradiente:

$$\frac{\partial J}{\partial \hat{y}^{(i)}} = -\frac{1}{m} \left[y^{(i)} \frac{1}{\hat{y}^{(i)}} - \left(1 - y^{(i)} \right) \frac{1}{1 - \hat{y}^{(i)}} \right]; \quad \frac{\partial \hat{y}^{(i)}}{\partial a^{[2](i)}} = 1; \quad \frac{\partial a^{[2](i)}}{\partial z^{[2](i)}} = \frac{\partial \text{sign}}{\partial z^{[2](i)}} \left(\partial z^{[2](i)} \right) = \begin{cases} 0 & z^{[2](i)} \neq 0 \\ c & z^{[2](i)} = 0 \end{cases}$$

definimos la derivada de *sign* de esta manera, dado que para $x = 0$ esta no es diferenciable. Luego:

$$\frac{\partial J}{\partial z^{[2](i)}} = \frac{\partial J}{\partial \hat{y}^{(i)}} \frac{\partial \text{sign}}{\partial z^{[2](i)}} \left(\partial z^{[2](i)} \right) = \begin{cases} 0 & z^{[2](i)} \neq 0 \\ -\frac{c}{m} \left[y^{(i)} \frac{1}{\hat{y}^{(i)}} - \left(1 - y^{(i)} \right) \frac{1}{1 - \hat{y}^{(i)}} \right] & z^{[2](i)} = 0 \end{cases}$$

Luego:

$$\frac{\partial J}{\partial W^{[2]}} = \frac{\partial J}{\partial z^{[2](i)}} \frac{\partial z^{[2](i)}}{\partial W^{[2]}} = \frac{\partial J}{\partial z^{[2](i)}} a^{[1](i)}$$

$$\frac{\partial J}{\partial b^{[2](i)}} = \frac{\partial J}{\partial z^{[2](i)}} \frac{\partial z^{[2](i)}}{\partial b^{[2](i)}} = \frac{\partial J}{\partial z^{[2](i)}}$$

Análogamente para la capa de salida:

$$\begin{aligned} \frac{\partial J}{\partial z^{[1](i)}} &= \frac{\partial J}{\partial z^{[2](i)}} \frac{\partial J}{\partial a^{[1](i)}} \frac{\partial a^{[1](i)}}{\partial z^{[1](i)}} = \frac{\partial J}{\partial z^{[2](i)}} W^{[2]} \frac{\partial J}{\partial z^{[1](i)}} = \\ &= W^{[2]} \left[\begin{cases} 0 & z^{[2](i)} \neq 0 \\ -\frac{c}{m} \left[y^{(i)} \frac{1}{\hat{y}^{(i)}} - \left(1 - y^{(i)} \right) \frac{1}{1 - \hat{y}^{(i)}} \right] & z^{[2](i)} = 0 \end{cases} \right] \left[\begin{cases} 0 & z^{[1](i)} \neq 0 \\ c & z^{[1](i)} = 0 \end{cases} \right] = \end{aligned}$$

$$= \begin{cases} 0 & z^{[2](i)} \neq 0 \vee z^{[1](i)} \neq 0 \\ -\frac{c^2}{m} \left[y^{(i)} \frac{1}{\hat{y}^{(i)}} - (1 - y^{(i)}) \frac{1}{1 - \hat{y}^{(i)}} \right] W^{[2]} & z^{[1](i)} = 0 \wedge z^{[2](i)} = 0 \end{cases}$$

$$\frac{\partial J}{\partial W^{[1]}} = \frac{\partial J}{\partial z^{[1](i)}} x^{(i)} \quad \frac{\partial J}{\partial b^{[1]}} = \frac{\partial J}{\partial z^{[1](i)}}$$

Ahora dado la tasa de aprendizaje α calculamos:

$$W^{[2]} = W^{[2]} - \alpha \frac{\partial J}{\partial W^{[2]}}$$

$$b^{[2]} = b^{[2]} - \alpha \frac{\partial J}{\partial b^{[2]}}$$

$$W^{[1]} = W^{[1]} - \alpha \frac{\partial J}{\partial W^{[1]}}$$

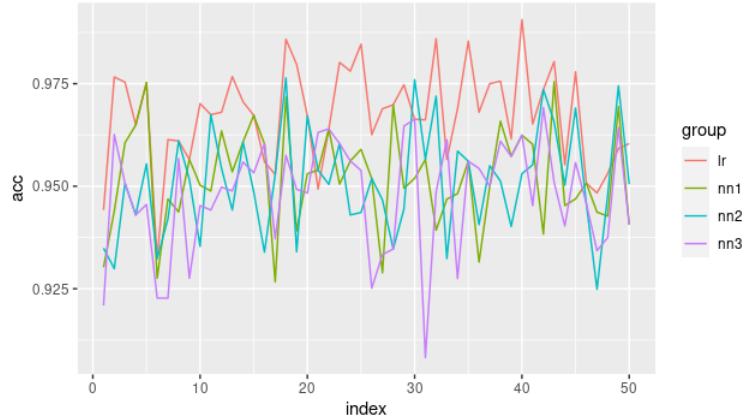
$$b^{[1]} = b^{[1]} - \alpha \frac{\partial J}{\partial b^{[1]}}$$

Notemos que en la mayoría de los casos estos gradientes serán cero y por tanto no se actualizarán los parámetros, lo cual es equivalente a que el modelo no aprenda.

C. Análisis de datos

1. En este problema se trata de construir un clasificador para los datos Wisconsin breast cancer sobre tumores, la clasificación es en tumores benignos (1) y malignos (0). La base consta de 683 observaciones de 9 variables que ofrecen información sobre un tumor determinado, y la variable Class que agrupa los tumores.

Figura 1: Comparación de modelos



En la figura vemos el comportamiento de los valores de precisión para 4 modelos: *nn1*, *nn2*, *nn3*, *lr*. Una NN con una capa oculta de 30 neuronas, una NN con 3 capas de 30 neuronas, una NN con 3 capas de 30 neuronas y una regresión logística respectivamente. Para estos obtuvimos medias de predicción de 0,952, 0,948, 0,947, 0,968 respectivamente. El modelo que mejor predicciones mostró es el de regresión, sin embargo todos se comparten muy bien y la diferencia entre estos es pequeña.

2. Partiendo de los datos de la Secretaria de Salud de COVID-19. Nos centraremos en los datos correspondientes a CDMX. Estos consisten en 394143 observaciones de pacientes, con información almacenada en 40 variables. En este problema se quiere estudiar si existe alguna relación entre la presencia de enfermedades como la diabetes y hipertensión, y hábitos como el tabaquismo, y el tipo de paciente, esto se refiere a si tipo de atención que recibió el paciente en la unidad. Se denomina como ambulatorio si regresó a su casa o se denomina como hospitalizado si fue ingresado a hospitalización.

Para esto nos restringimos a las siguientes variables de interés que usaremos como descriptores: {"SEXO", "NEUMONIA", "EDAD", "DIABETES", "EPOC", "ASMA", "INMUSUPR", "HIPERTENSION", "CARDIOVASCULAR", "OBESIDAD", "RENAL_CRONICA", "TABAQUISMO"}

Nuestro Target será la variable TIPO_PACIENTE".

Figura 2:

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.506e+00  2.837e-02 -88.317 < 2e-16 ***
SEXO         4.353e-01  8.446e-03  51.535 < 2e-16 ***
NEUMONIA     -1.010e+00  9.816e-03 -102.929 < 2e-16 ***
EDAD         5.315e-02  2.806e-04 189.438 < 2e-16 ***
DIABETES     -3.770e-03  1.707e-03  -2.209 0.02716 *
EPOC         3.978e-03  2.183e-03   1.822 0.06851 .
ASMA         5.446e-08  2.455e-03   0.000 0.99998
INMUSUPR     2.974e-03  2.557e-03   1.163 0.24476
HIPERTENSION 2.016e-03  1.911e-03   1.055 0.29150
CARDIOVASCULAR 9.984e-03  2.388e-03   4.180 2.91e-05 ***
OBESIDAD     1.540e-02  1.315e-03  11.711 < 2e-16 ***
RENAL_CRONICA -1.773e-02  2.815e-03  -6.298 3.01e-10 ***
TABAQUISMO   5.486e-03  1.821e-03   3.013 0.00259 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Como se muestra en el modelo, de la figura anterior las variables referentes a "DIABETES", "EPOC", "ASMA", "INMUSUPR", "HIPERTENSION" no parecen ser significantes para esta pregunta en particular. Eliminando estas variables obtenemos

Figura 3:

```

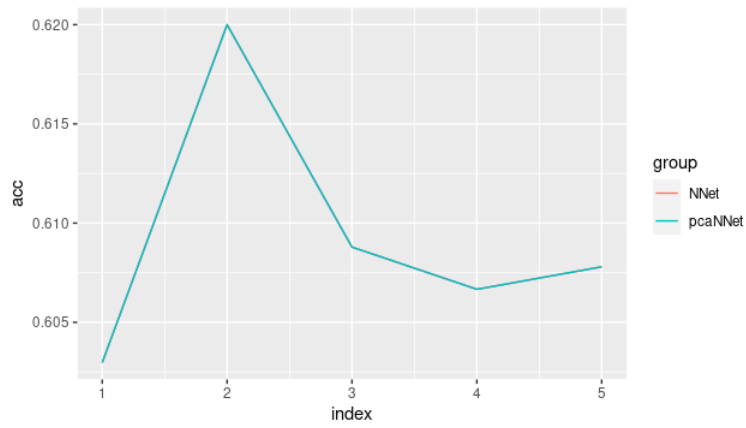
Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.5051393  0.0283653 -88.317 < 2e-16 ***
SEXO         0.4353509  0.0084456  51.548 < 2e-16 ***
NEUMONIA     -1.0102878  0.0098158 -102.925 < 2e-16 ***
EDAD         0.0531514  0.0002803 189.634 < 2e-16 ***
CARDIOVASCULAR 0.0114482  0.0022005   5.202 1.97e-07 ***
OBESIDAD     0.0157575  0.0012971  12.149 < 2e-16 ***
RENAL_CRONICA -0.0155664  0.0025268  -6.160 7.26e-10 ***
TABAQUISMO   0.0062893  0.0017413   3.612 0.000304 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Para esos modelos obtuvimos una precisión de 0,846 y 0,851 respectivamente. Dada la insignificancia del segundo modelo y el valor obtenido de precisión podemos suponer que existe una relación entre las variables y el tipo de paciente. Y para nuevos datos, basados en estas podemos decir con un 85 % de seguridad que tipo de paciente este será.

3. En este problema trabajamos con los datos de spam vistos en la tarea anterior. La idea es construir un clasificador usando redes neuronales para identificar si un correo es spam o no. Se utilizó los modelos de la librería caret: nnet y pcaNNet y los resultados obtenidos se muestran en la siguiente tabla.

Figura 4:



como podemos ver no se muestra diferencia alguna entre estos modelos, además de ser los valores de precisión obtenidos sumamente bajos.