Boletín 4: redes neuronales

Para la realización de las prácticas correspondientes a este boletín se utilizará scikit-learn en el CESGA. Utilizaremos un **SEED_VALUE=1**.

1. Se tiene una red neuronal con una capa oculta, s_l={2, 3, 1}, $W^{(1)}$= (-2, 1; 1, -1; 3, -1), $W^{(2)}$= (2, 3, 1), $b^{(1)}$= (0; -1; 1), y $b^{(2)}$= (0). La red utiliza la función de activación sigmoide en todas las neuronas salvo la de salida, cuya función de activación es la identidad. Además, la función de coste es el error cuadrático. Dado el ejemplo (1, 1, 1):

    a. Calcula $z_i^{(l)}$ y $a_i^{(l)}$ para todas las neuronas.

    b. Usando el algoritmo de retro-propagación del error, calcula $\delta_i^{(l)}$ para todas las neuronas.

    c. Usando el algoritmo de retro-propagación del error, determina los valores finales de cada peso ($W_{ij}^{(l)}$) y bias ($b_i^{(l)}$) de la red neuronal tras finalizar la primera iteración del algoritmo, asumiendo un valor de λ=1, y una tasa de aprendizaje α=0,5.

    Nota: este ejercicio debe hacerse sin utilizar ninguna función de scikit-learn. Para realizar los cálculos (multiplicación de matrices, etc.) se recomienda utilizar Python.

2. Dado el problema de clasificación Blood Transfusion Service Center:

    a. La clase que implementa el Perceptrón multicapa (MLP) en problemas de clasificación en scikit-learn es *sklearn.neural_network.MLPClassifier*. Revisa los parámetros y métodos que tiene.

    b. Divide los datos en entrenamiento (80%) y test (20%).

    c. Realiza la experimentación con *MLPClassifier* usando los valores por defecto de los parámetros, excepto para *activation='tanh* y *solver='lbfgs'*, y *max_iter=2000.* Además, utiliza los siguientes hiper-parámetros:

        i. *hidden_layer_sizes*: prueba entre 1 y 3 capas ocultas, y varía el número de neuronas en dichas capas (el número de neuronas por capa debe ser el mismo en todas las capas ocultas para reducir el número de combinaciones de hiper-parámetros).

        ii. *alpha*: parámetro de regularización. Prueba potencias enteras de 10 multiplicadas por 1 y 5 (…; 0,05; 0,1; 0,5; 1; 5; 10; 50; …).

        Muestra las gráficas del error de entrenamiento con validación cruzada (5-CV) frente al valor de los hiper-parámetros, y justifica la elección de los valores más apropiados. Para cada combinación de valor del número de capas ocultas y parámetro de regularización, que sea de interés, se debe generar una gráfica donde se represente en el eje horizontal el número de neuronas en la capa oculta.

        ¿Cuál es el menor error de validación cruzada, su desviación estándar y el valor de los hiper-parámetros para el que se consigue?

d. Muestra la gráfica del error de test frente al valor de los hiper-parámetros (siguiendo el mismo esquema que en el apartado anterior), y valora si la gráfica del error de entrenamiento con validación cruzada ha hecho una buena estimación del error de test.

   ¿Cuál es el error de test para el valor de los hiper-parámetros seleccionados por la validación cruzada?

3. Repite el ejercicio 2 pero para el problema de regresión Energy Efficiency con la variable de salida *cooling load* mediante la función *sklearn.neural_network.MLPRegressor*.

Entregable

Se debe entregar un único fichero comprimido con el nombre *PrimerApellido_SegundoApellido.zip* (también son válidos los formatos .rar y .7z), que contenga dos archivos:

- El primer archivo debe ser de tipo pdf, y contendrá exclusivamente las respuestas a los ejercicios (incluyendo las gráficas necesarias para justificar dichas respuestas). No se incluirá en este archivo ningún otro tipo de texto.

- El segundo archivo será de tipo ipynb, y permitirá reproducir toda la experimentación realizada en el boletín.

Ejercicio 1

I. **Calcula $x_i^{(l)}$ y $a_i^{(l)}$ para todas las neuronas.**

$$z^{(2)} = \begin{pmatrix} -1 \\ -1 \\ 3 \end{pmatrix}; \quad a^{(2)} \approx \begin{pmatrix} 0{,}2689 \\ 0{,}2689 \\ 0{,}9526 \end{pmatrix}; \quad z^{(3)} \approx 2{,}2973; \quad a^{(3)} = z^{(3)} \approx 2{,}2973 \tag{1}$$

II. **Usando el algoritmo de retropropagación del error, calcula $\delta_i^{(l)}$ para todas las neuronas.**

$$\delta^{(2)} \approx \begin{pmatrix} 0{,}5101 \\ 0{,}7652 \\ 0{,}0586 \end{pmatrix}; \quad \delta^{(3)} \approx 1{,}2973 \tag{2}$$

III. **Usando el algoritmo de retropropagación del error, determina los valores finales de cada peso ($W_{ij}^{(l)}$) y *bias* ($b_i^{(l)}$) de la red neuronal tras finalizar la primera iteración del algoritmo, asumiendo un valor de $\lambda = 1$ y una tasa de aprendizaje $\alpha = 0{,}5$.**

$$W^{(1)} \approx \begin{pmatrix} -1{,}2551 & 0{,}2449 \\ 0{,}1174 & -0{,}8826 \\ 1{,}4707 & -0{,}5293 \end{pmatrix}; \quad b^{(1)} \approx \begin{pmatrix} -0{,}2551 \\ -1{,}3826 \\ 0{,}9707 \end{pmatrix} \tag{3}$$

$$W^{(2)} \approx \begin{pmatrix} 0{,}8256 & 1{,}3256 & -0{,}1179 \end{pmatrix}; \quad b^{(2)} \approx -0{,}6487 \tag{4}$$

Ejercicio 2

- **Menor error de validación cruzada, su desviación estándar y valor de los hiperparámetros:** $\Delta = 0{,}194006$, $\sigma = 0{,}029822$, $\text{param\_hidden\_layer\_sizes} = (3,)$, $\text{param\_alpha} = 0{,}5$.

- **Error de test para los hiperparámetros de validación cruzada:** $\Delta = 0{,}2$, $\text{param\_hidden\_layer\_sizes} = (3,)$, $\text{param\_alpha} = 0{,}5$.

Ejercicio 3

- **Menor error de validación cruzada, su desviación estándar y valor de los hiperparámetros:** $\text{MSE} = 0{,}479828$, $\sigma = 0{,}122977$, $\text{param\_hidden\_layer\_sizes} = (32, 32, 32)$, $\text{param\_alpha} = 0{,}1$.

- **Error de test para los hiperparámetros de validación cruzada:** $\text{MSE} = 0{,}493655$, $\text{param\_hidden\_layer\_sizes} = (32, 32, 32)$, $\text{param\_alpha} = 0{,}1$.

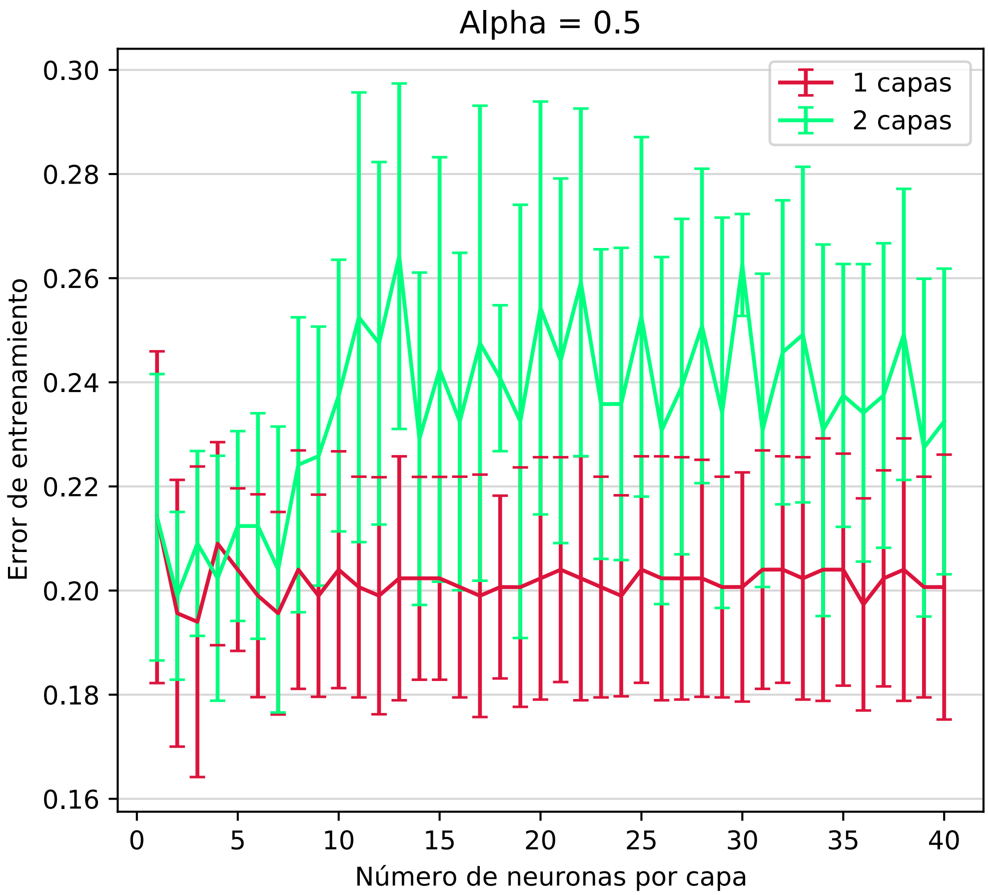


Figura I: Ejercicio 2: error de validación cruzada en datos de validación.

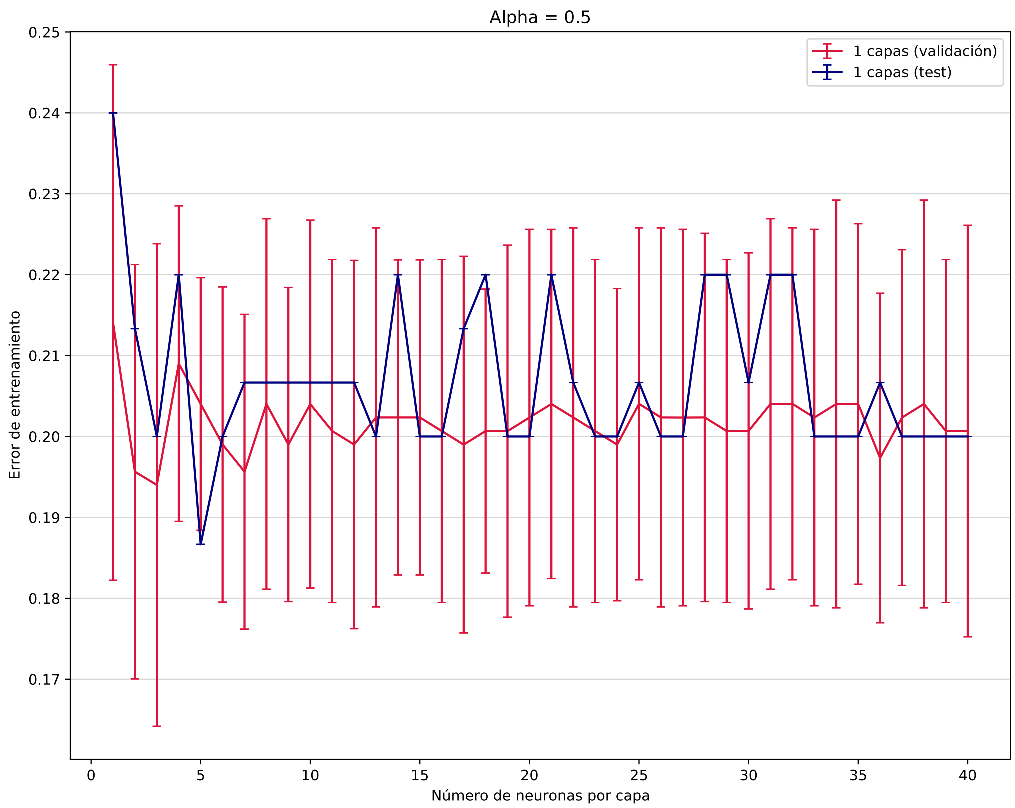


Figura II: Ejercicio 2: error de validación cruzada en datos de validación frente al mismo error en datos de test.

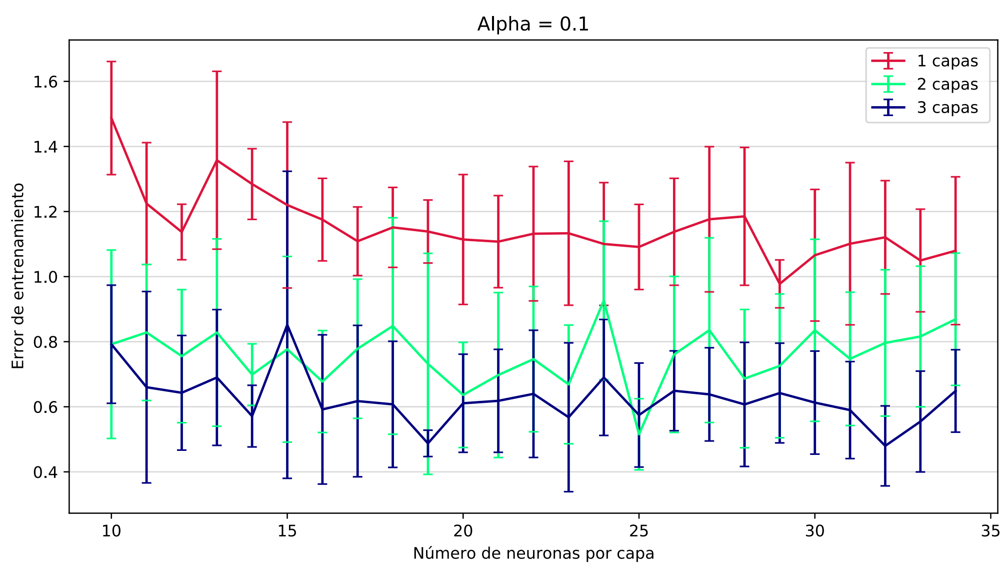


Figura III: Ejercicio 3: error de validación cruzada en datos de validación.

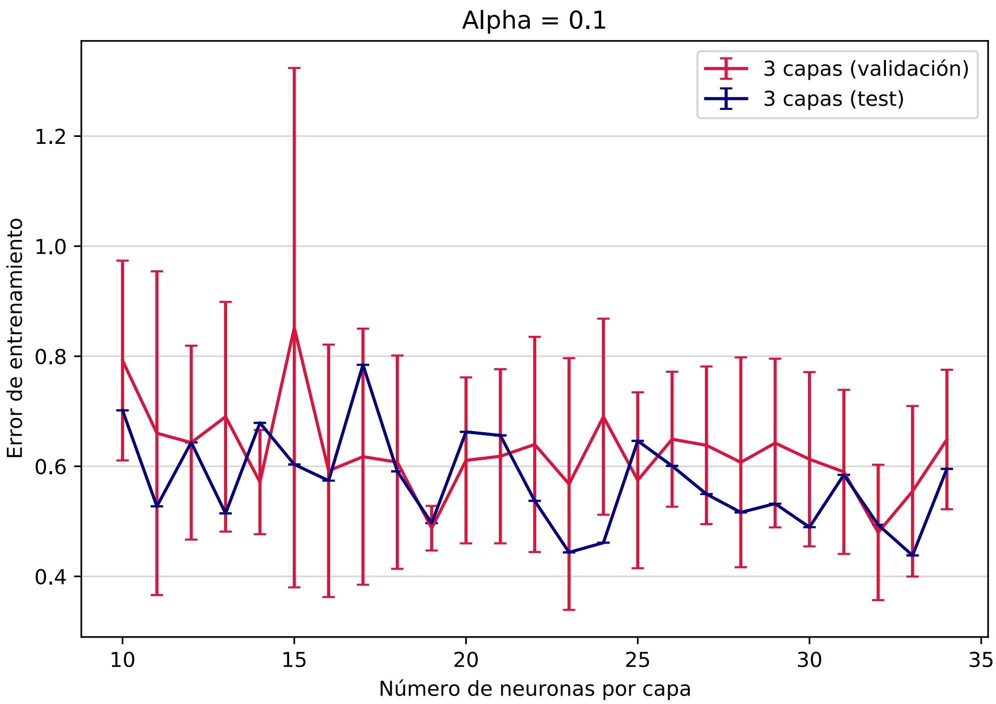
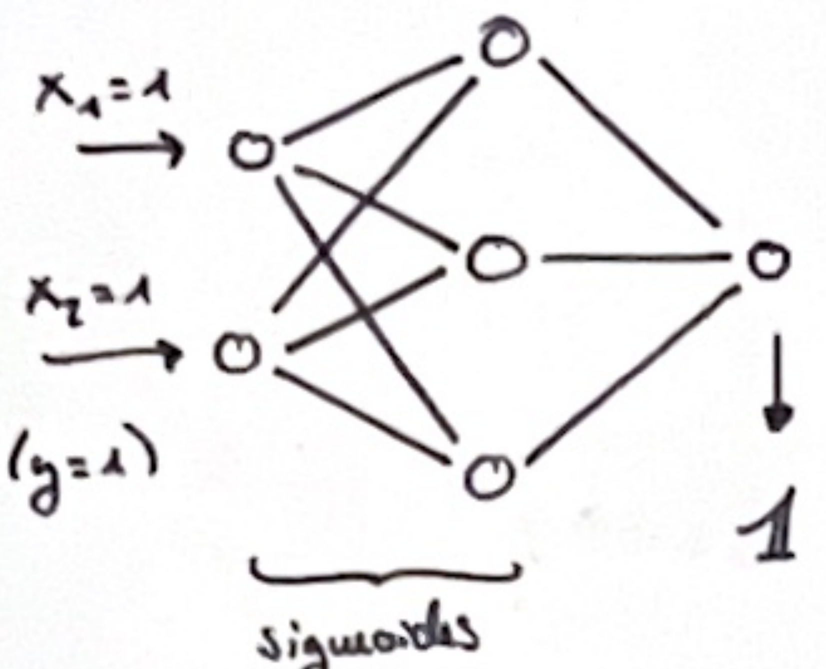


Figura IV: Ejercicio 3: error de validación cruzada en datos de validación frente al mismo error en datos de test.

Boletín 4. Redes neuronales

Ejercicio 1.



$$W^{(1)}=\begin{pmatrix}-2 & 1\\ 1 & -1\\ 3 & -1\end{pmatrix};\quad b^{(1)}=\begin{pmatrix}0\\-1\\1\end{pmatrix}$$

$$W^{(2)}=\begin{pmatrix}2 & 3 & 1\end{pmatrix};\quad b^{(2)}=(0)$$

$$f(z)=\frac{1}{1+e^{-z}}$$

a) $z^{(l+1)}=W^{(l)}a^{(l)}+b^{(l)}$; $a^{(l+1)}=f(z^{(l+1)})$

$$z^{(2)}=W^{(1)}\underset{x}{\underset{\parallel}{a^{(1)}}}+b^{(1)}=\begin{pmatrix}-2 & 1\\ 1 & -1\\ 3 & -1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}+\begin{pmatrix}0\\-1\\1\end{pmatrix}=\begin{pmatrix}-1\\0\\2\end{pmatrix}+\begin{pmatrix}0\\-1\\1\end{pmatrix}=\boxed{\begin{pmatrix}-1\\-1\\3\end{pmatrix}}$$

$$a^{(2)}=f(z^{(2)})=\begin{pmatrix}f(-1)\\f(-1)\\f(3)\end{pmatrix}\underset{\text{sigmoide}}{=}\begin{pmatrix}\frac{1}{1+e}\\ \frac{1}{1+e}\\ \frac{1}{1+e^{-3}}\end{pmatrix}\approx\boxed{\begin{pmatrix}0.2689\\0.2689\\0.9526\end{pmatrix}}$$

$$z^{(3)}=W^{(2)}a^{(2)}+b^{(2)}=\begin{pmatrix}2 & 3 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{1+e}\\ \frac{1}{1+e}\\ \frac{1}{1+e^{-3}}\end{pmatrix}+(0)=\frac{5}{1+e}+\frac{1}{1+e^{-3}}\approx\boxed{2.2973}$$

$$a^{(3)}=f(z^{(3)})\overset{1}{=}z^{(3)}\approx\boxed{2.2973}$$

b) $\delta^{(nl)}=(a^{(nl)}-y)\circ f'(z^{(nl)})$ / $f'(z)=\begin{cases}1 & \text{si } f(z)=1\\ \frac{e^{-z}}{(1+e^{-z})^2} & \text{si } f(z)=\frac{1}{1+e^{-z}}\end{cases}$

Salida: $\delta^{(3)}=(a^{(3)}-y)\circ \overset{1}{\cancel{f'(z^{(3)})}}=\frac{5}{1+e}+\frac{1}{1+e^{-3}}-1\approx\boxed{1.2973}$

Capa oculta: $\delta^{(2)}=[(W^{(2)})^T\delta^{(3)}]\circ f'(z^{(2)})=\begin{pmatrix}2\left(\frac{5}{1+e}+\frac{1}{1+e^{-3}}-1\right)\\ 3\left(\frac{5}{1+e}+\frac{1}{1+e^{-3}}-1\right)\\ \frac{5}{1+e}+\frac{1}{1+e^{-3}}-1\end{pmatrix}\circ\begin{pmatrix}\frac{e}{(1+e)^2}\\ \frac{e}{(1+e)^2}\\ \frac{e^{-3}}{(1+e^{-3})^2}\end{pmatrix}=$

$$=\begin{pmatrix}2\left(\frac{5}{1+e}+\frac{1}{1+e^{-3}}-1\right)\frac{e}{(1+e)^2}\\ 3\left(\frac{5}{1+e}+\frac{1}{1+e^{-3}}-1\right)\frac{e}{(1+e)^2}\\ \left(\frac{5}{1+e}+\frac{1}{1+e^{-3}}-1\right)\frac{e^{-3}}{(1+e^{-3})^2}\end{pmatrix}\approx\boxed{\begin{pmatrix}0.5101\\0.7652\\0.0586\end{pmatrix}}$$

c) Calculamos primero los gradientes $\nabla_{W^{(l)}} J = \delta^{(l+1)} (a^{(l)})^T$; $\nabla_{b^{(l)}} J = \delta^{(l+1)}$

$$\boxed{\nabla_{W^{(1)}} J = \delta^{(2)} (a^{(1)})^T \approx \begin{pmatrix} 0.5101 \\ 0.7652 \\ 0.0586 \end{pmatrix} (1 \quad 1) = \begin{pmatrix} 0.5101 & 0.5101 \\ 0.7652 & 0.7652 \\ 0.0586 & 0.0586 \end{pmatrix}}$$

$$\boxed{\nabla_{b^{(1)}} J = \delta^{(2)} \approx \begin{pmatrix} 0.5101 \\ 0.7652 \\ 0.0586 \end{pmatrix}}$$

$$\boxed{\nabla_{W^{(2)}} J = \delta^{(3)} (a^{(2)})^T \approx 1.2973 \, (0.2689 \quad 0.2689 \quad 0.9526) \approx (0.3488 \quad 0.3488 \quad 1.2358)}$$

$$\boxed{\nabla_{b^{(2)}} J = \delta^{(3)} \approx 1.2973}$$

Fijamos $\Delta W^{(l)} = \Delta b^{(l)} = 0$ y $\begin{cases} \Delta W^{(l)} := \cancel{\Delta W^{(l)}}^{\,0} + \nabla_{W^{(l)}} J \\ \Delta b^{(l)} := \Delta b^{(l)} + \nabla_{b^{(l)}} J \end{cases}$

Actualizamos los pesos y bias con $\alpha = 1/2$, $\lambda = 1$ y $m = 1$ (un batch)

$$W^{(l)} = W^{(l)} - \alpha \left[\frac{1}{m} \Delta W^{(l)} + \lambda W^{(l)} \right] = W^{(l)} - \frac{1}{2} \left[\nabla_{W^{(l)}} J + W^{(l)} \right]$$

$$b^{(l)} = b^{(l)} - \alpha \left[\frac{1}{m} \Delta b^{(l)} \right] = b^{(l)} - \frac{1}{2} \nabla_{b^{(l)}} J$$

$$\boxed{W^{(1)}} \approx \begin{pmatrix} -2 & 1 \\ 1 & -1 \\ 3 & -1 \end{pmatrix} - \frac{1}{2} \left[\begin{pmatrix} 0.5101 & 0.5101 \\ 0.7652 & 0.7652 \\ 0.0586 & 0.0586 \end{pmatrix} + \begin{pmatrix} -2 & 1 \\ 1 & -1 \\ 3 & -1 \end{pmatrix} \right] \approx \boxed{\begin{pmatrix} -1.2551 & 0.2449 \\ 0.1174 & -0.8826 \\ 1.4707 & -0.5293 \end{pmatrix}}$$

$$\boxed{b^{(1)}} \approx \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 0.5101 \\ 0.7652 \\ 0.0586 \end{pmatrix} = \boxed{\begin{pmatrix} -0.2551 \\ -1.3826 \\ 0.9707 \end{pmatrix}}$$

$$\boxed{W^{(2)}} \approx (2 \quad 3 \quad 1) - \frac{1}{2} \left[(0.3488 \quad 0.3488 \quad 1.2358) + (2 \quad 3 \quad 1) \right] \approx \boxed{(0.8256 \quad 1.3256 \quad -0.1179)}$$

$$\boxed{b^{(2)}} \approx 0 - \frac{1}{2} \cdot 1.2973 \approx \boxed{-0.6487}$$