

María Viviana Lara – 202122798

Catalina Gómez - 202122560

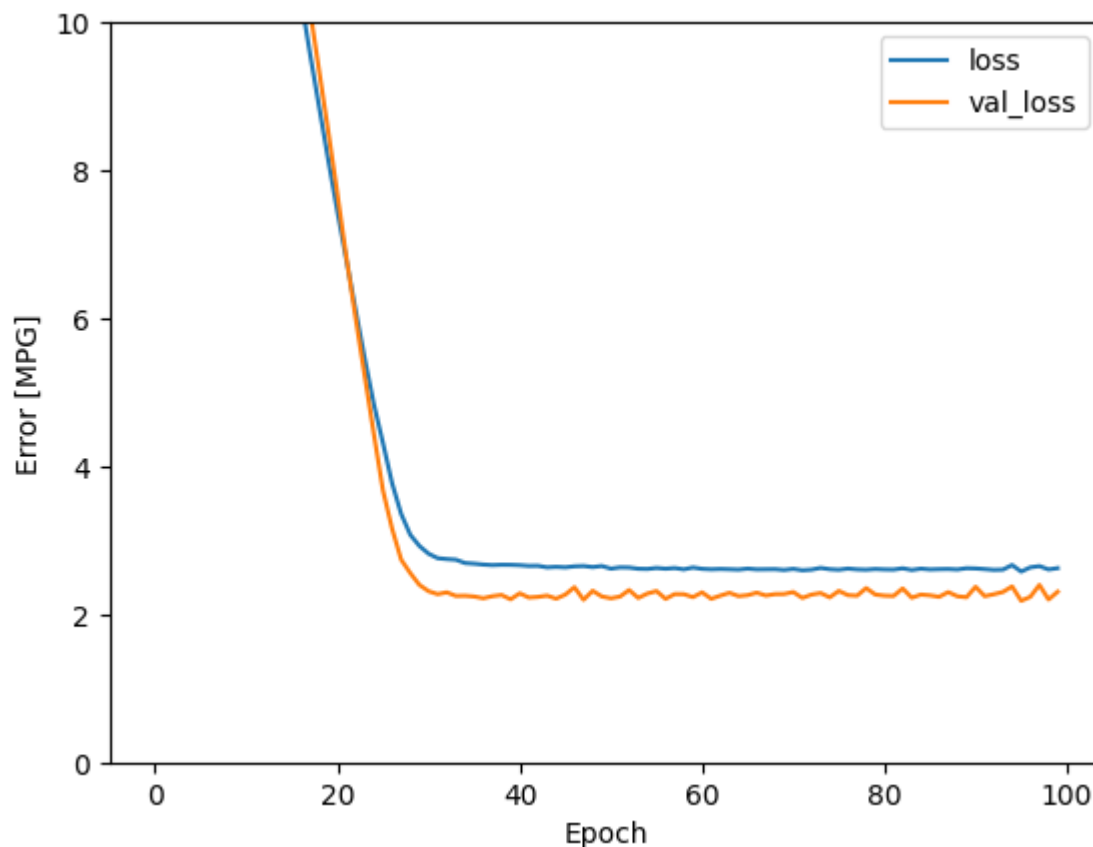
Juan David Rios - 201515312

Taller 6 – Intro. a Redes Neuronales

2. Entrene este nuevo modelo (éste será nuestro modelo base), grafique el historial de perdida de entrenamiento y validación, y comente sus resultados en su reporte

R/ El MAE para nuestro modelo base es de 2.767 evaluándolo en nuestros datos de prueba. El modelo entrenado tiene una capa oculta con una neurona y 6 variables explicativas: Cylinders, Displacement, Horsepower, Weight, Acceleration, Model Year

El modelo base se entrenó con un optimizador Adam con un learning rate de 0.1

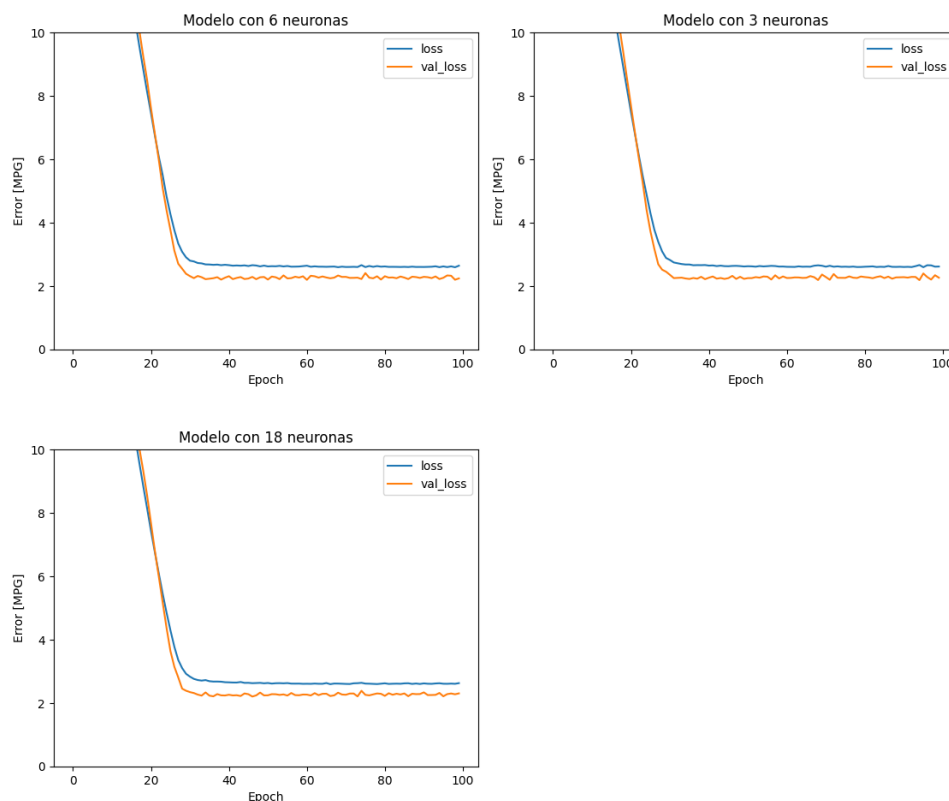


En general, el historial de perdida muestra una mejora significativa alrededor de la época 20 y después una mejora casi imperceptible en las épocas siguientes. También se puede observar que existe un nivel bajo de overfitting, ya que los datos de entrenamiento y de validación del entrenamiento muestran valores de perdida similares.

3. Modifique el número de neuronas (3 casos diferentes al modelo base), entrene los modelos resultantes y compare. Use siempre los mismos datos de entrenamiento y prueba. Comente sus resultados en su reporte

R/ Se generaron 3 modelos diferentes, manteniendo todos los parámetros constantes, únicamente modificando el número de neuronas. Se utilizaron 3, 6 y 18 neuronas en cada uno de estos

```
{'modelo_base': 2.7678310871124268,  
'modelo 6 neuronas': 2.7776596546173096,  
'modelo 3 neuronas': 2.773944854736328}
```



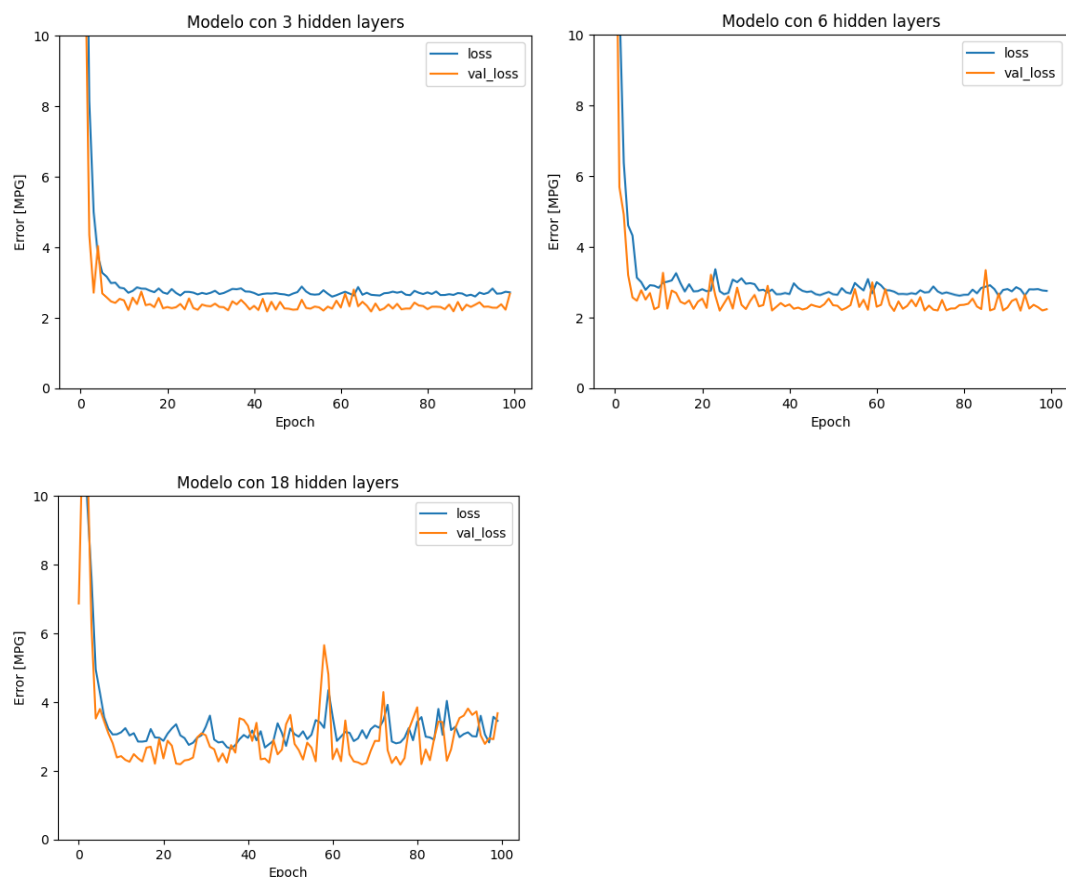
El comportamiento del historial de perdida en los 3 modelos es similar, sin embargo, revisando los datos de prueba, vemos que logra un valor menor del MAE, comparado con el modelo base y los 2 modelos adicionales. Es curioso que el modelo con 6 neuronas genere una MAE más alto que el modelo con 18 neuronas.

4. Modifique el número de capas (3 casos diferentes al modelo base), entrene los modelos resultantes y compare. Use siempre los mismos datos de entrenamiento y prueba.

R/ Regresamos a las características iniciales del modelo base y modificamos únicamente el número de capas del modelo, se utilizaron 3, 6 y 18 capas en cada uno de estos modelos obteniendo un valor del MAE como se observa a continuación:

```
'modelo 3 hidden layers': 2.8330602645874023,
'modelo 6 hidden layers': 2.968151807785034,
'modelo 18 hidden layers': 3.423922061920166}
```

Observamos que por los valores del MAE estos modelos son mayores que el modelo base. Sin embargo, valores medios de 3 capas ocultas pueden ayudar a estar cerca del mejor modelo hasta el momento (modelo base).



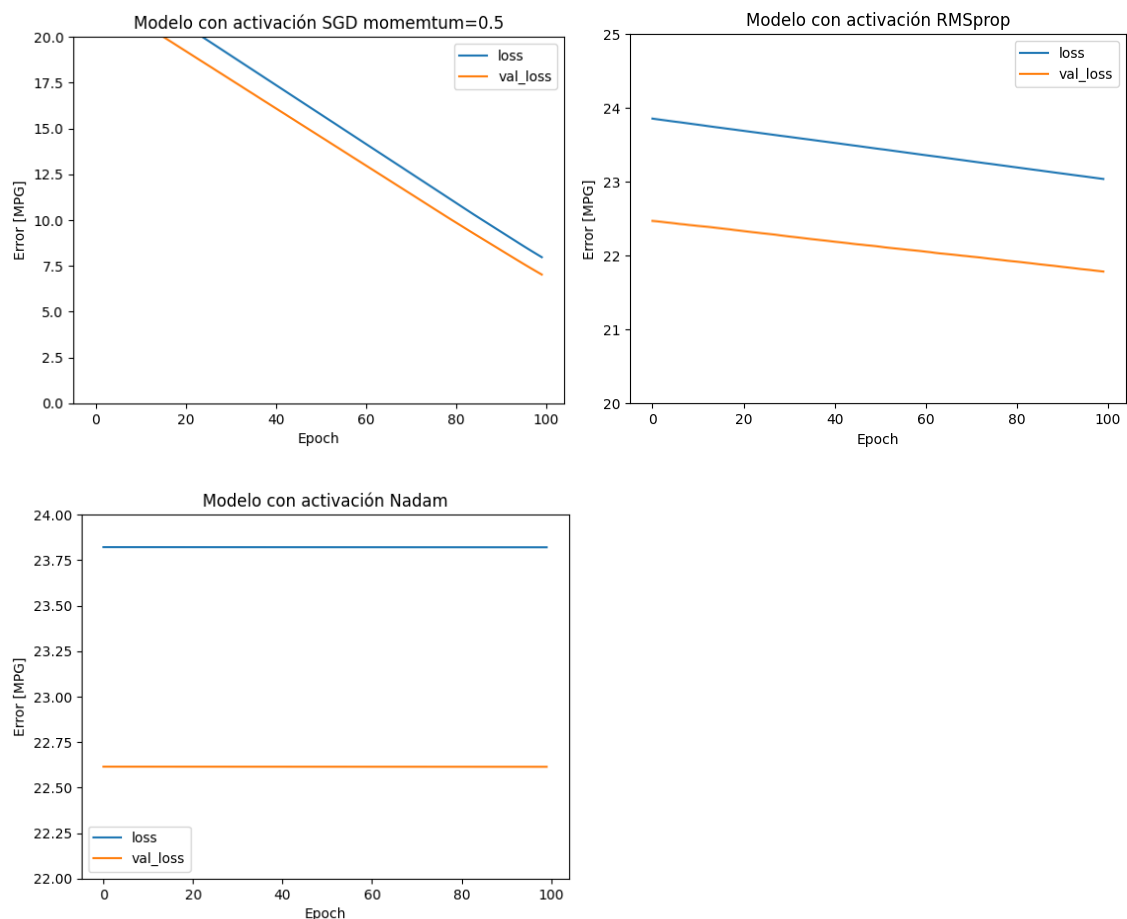
También, se observa que a medida que aumenta el número de hidden layers, la búsqueda de un óptimo en la función de pérdida se vuelve más errática.

5. Modifique las funciones de activación (3 casos diferentes al modelo base), entrene los modelos resultantes y compare. Use siempre los mismos datos de entrenamiento y prueba

Se construyeron 3 modelos únicamente modificando la función de activación, se probaron las funciones de activación SGD (learning_rate=0.01, momentum=0.5), RMSprop (learning_rate=0.001, rho=0.01) y Nadam (learning_rate=0.00001).

```
'modelo con SGD': 7.385102272033691,
'modelo con RMSprop': 22.06812858581543,
'modelo con Nadam': 22.839012145996094}
```

Con estas funciones de activación vemos que estos modelos se alejan del MAE del modelo base y los convierte en peores modelos.



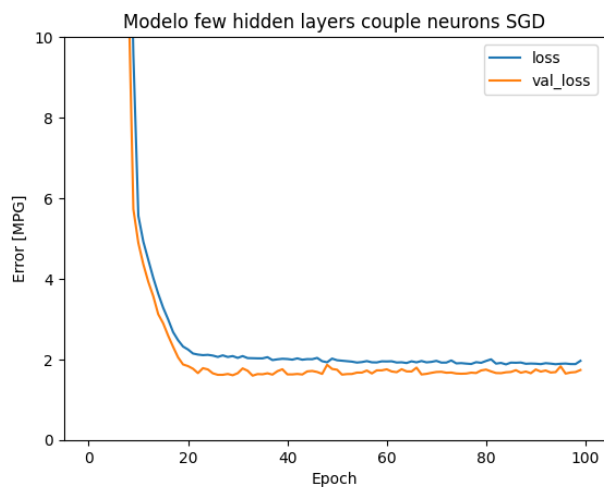
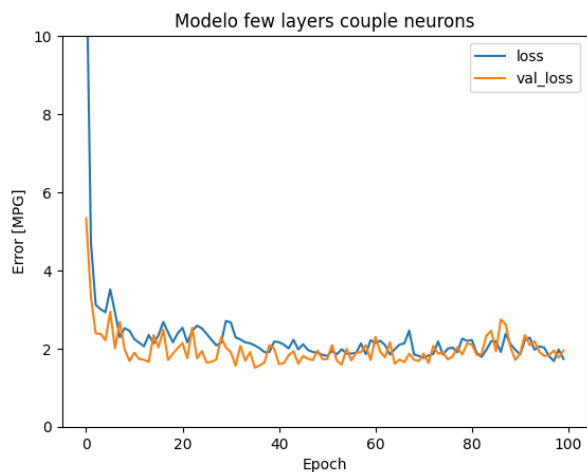
En estos casos podemos observar que estas funciones cuentan con un nivel mayor de overfitting y no logran capturar el comportamiento de la variable de respuesta.

6. A partir de los resultados anteriores proponga y evalúe 2 modelos diferentes al modelo base y compárelos.

Para este punto se crearon dos modelos con las siguientes características:

Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 6)	13
dense_34 (Dense)	(None, 24)	168
dense_35 (Dense)	(None, 24)	600
dense_36 (Dense)	(None, 3)	75

Las dos primeras hidden layers tienen una función de activación Relu en ambos modelos, sin embargo, la tercera capa tiene una función de activación Adam/SGD.



```
'modelo few hidden layers couple neurons': 1.89825439453125,
'modelo few hidden layers couple neurons SGD': 2.1854135990142822}
```

Entre las dos opciones planteadas vemos que el modelo utilizando la función de activación de SGD, presenta mejores resultados entre estos dos modelos.

```
{'modelo_base': 2.7678310871124268,
 'modelo 6 neuronas': 2.7776596546173096,
 'modelo 3 neuronas': 2.773944854736328,
 'modelo 18 neuronas': 2.7520625591278076,
 'modelo 3 hidden layers': 2.8330602645874023,
 'modelo 6 hidden layers': 2.968151807785034,
 'modelo 18 hidden layers': 3.423922061920166,
 'modelo con SGD': 7.385102272033691,
 'modelo con RMSprop': 22.06812858581543,
 'modelo con Nadam': 22.839012145996094,
 'modelo few hidden layers couple neurons': 1.89825439453125,
 'modelo few hidden layers couple neurons SGD': 2.1854135990142822}
```

Comparándolo con todos los modelos construidos hasta el momento y con el modelo base, podemos ver que el modelo con 3 capas, donde las primeras dos capas tienen 24 neuronas y

una activación relu, y la tercera capa tiene 3 neuronas, donde el modelo se entrenó con un optimizador Adam con un learning rate de 0.1, es el mejor modelo ya que logra optimizar el MAE con un valor en los datos de prueba de ~ 1.898 .