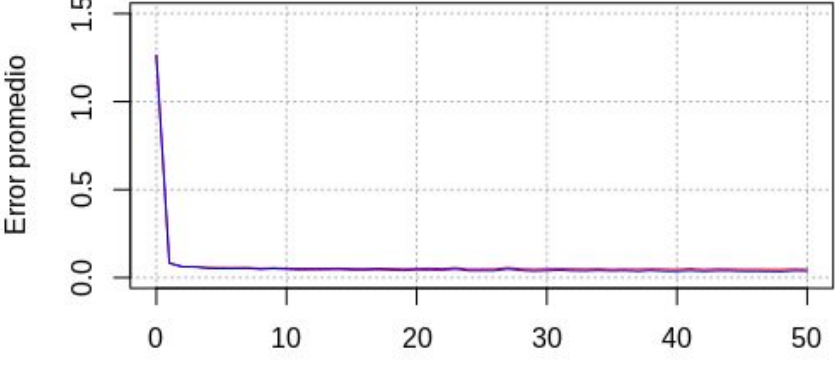
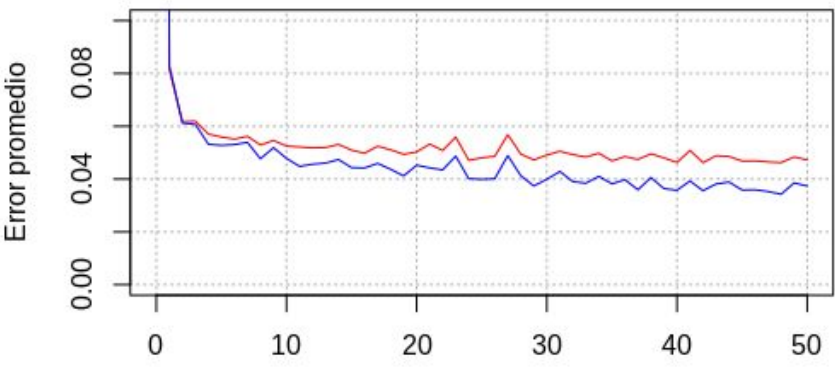


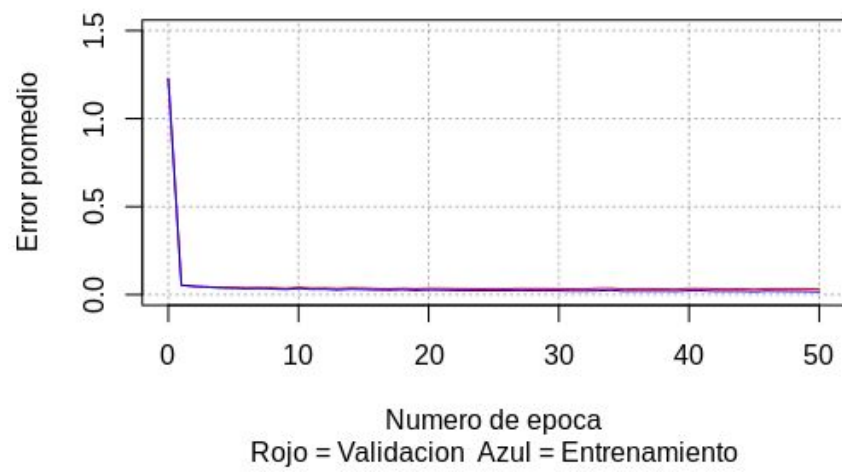
Tarea MLP

2)

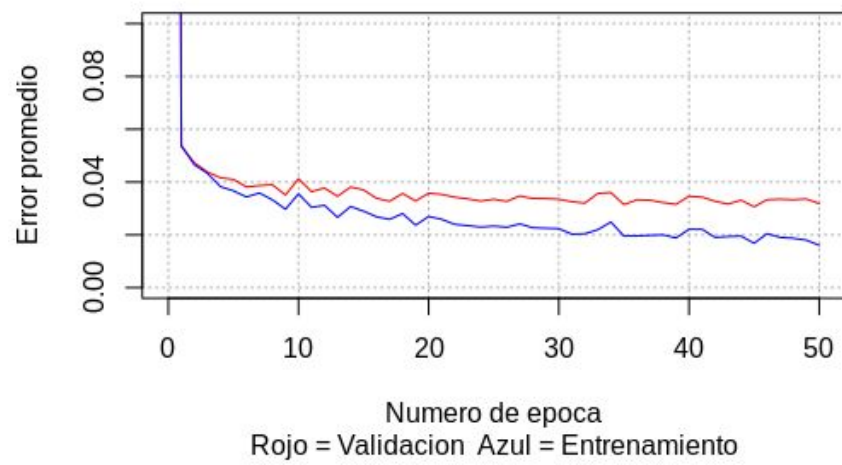
Número de neuronas en la capa escondida	Gráfica error promedio vs épocas
20	<p style="text-align: center;">Error por epoca</p>  <p style="text-align: center;">Numero de epoca Rojo = Validacion Azul = Entrenamiento</p> <p style="text-align: center;">Error por epoca</p>  <p style="text-align: center;">Numero de epoca Rojo = Validacion Azul = Entrenamiento</p>

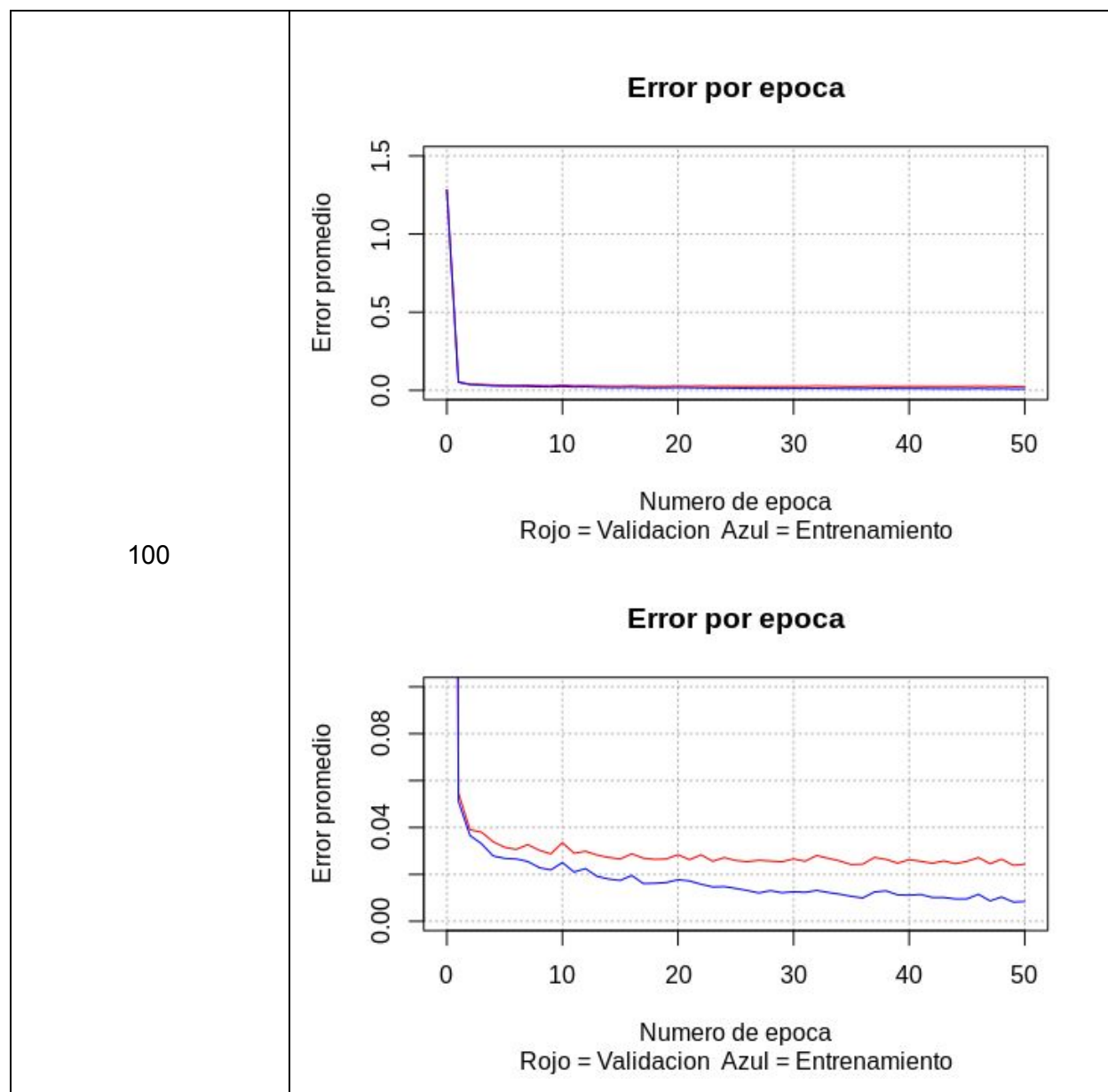
50

Error por epoca



Error por epoca





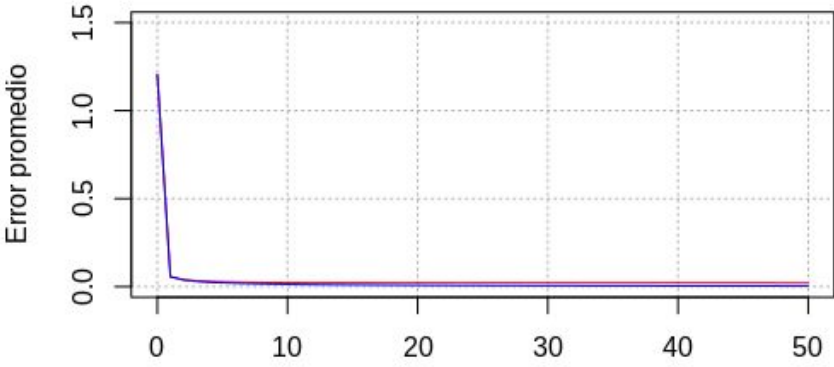
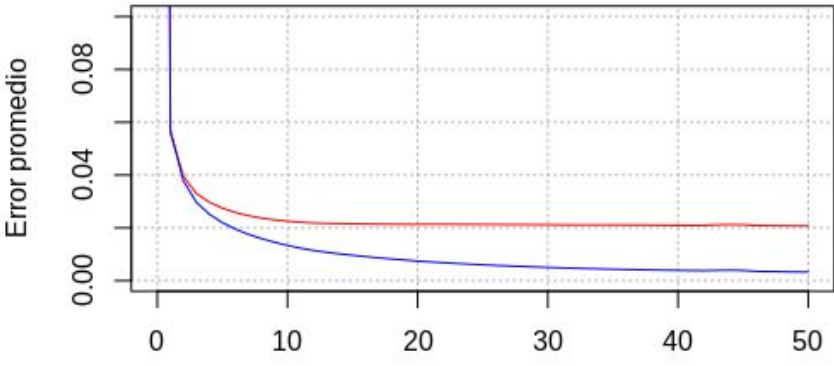
Número de neuronas en la capa escondida	Porcentaje de clasificados correctamente	
	Criterio viejo	Criterio nuevo
20	94.37%	87.76%
50	96.55%	91.84%
100	97.59%	93.83%

Cuando vemos las gráficas de error promedio desde lejos parecieran ser idénticas las 3, pero cuando nos acercamos nos damos cuenta que a medida que colocamos más neuronas en la capa escondida, el error baja tanto con los datos de entrenamiento como los de validación. Si comparamos en cada caso el error de entrenamiento con el de validación, podemos ver que tienen un comportamiento muy parecido solo que el de validación está un poco desplazado hacia arriba. Aunque el error disminuye a medida de que aumentan las neuronas en la capa escondida, hay dos cosas importantes que decir: mientras más neuronas haya, más espacio y tiempo consume el algoritmo en entrenar, en nuestro caso el de 100 neuronas tardaba hora y media, el de 50 tomó cuarenta minutos y el de 20 tardó veinte minutos. Y también que aunque el error baje, esto no significa que los resultados sean mejores, así que para comprobar esto veamos ahora el porcentaje de clasificaciones.

Mientras más neuronas hay, mejor es el porcentaje con el criterio nuevo y el viejo (tomar el máximo como 1 y el resto como 0). Llegando hasta un 97.59% con el criterio viejo y 93.83% con el nuevo. Tiene sentido que el viejo de mejor que el nuevo, ya que en los casos en los que el nuevo clasifica bien, el viejo también lo hace, pero el viejo también lo puede hacer en casos en los que el nuevo no puede.

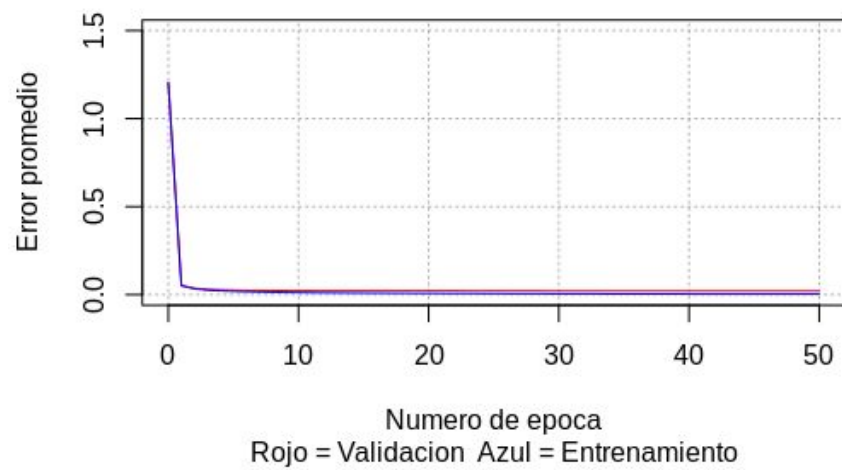
Comparando con las tareas anteriores, los resultados son mucho mejores, con el perceptrón simple lo mejor que logramos fue 75.35% y con el adaline 85.92%, por lo cual mejoramos en un 22.24% y 11.67% respectivamente, esto representa una mejora muy grande. Aunque el perceptrón multicapas tome mucho más tiempo y espacio, la mejora es tan grande que lo vale.

3)

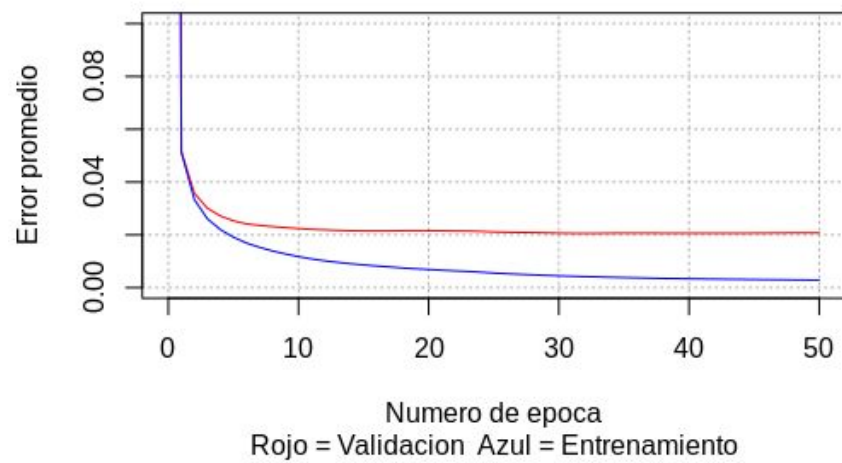
alpha	Gráfica error promedio vs épocas
0	<p data-bbox="878 407 1127 443">Error por epoca</p>  <p data-bbox="760 869 1247 936">Numero de epoca Rojo = Validacion Azul = Entrenamiento</p> <p data-bbox="878 1003 1127 1039">Error por epoca</p>  <p data-bbox="760 1465 1247 1533">Numero de epoca Rojo = Validacion Azul = Entrenamiento</p>

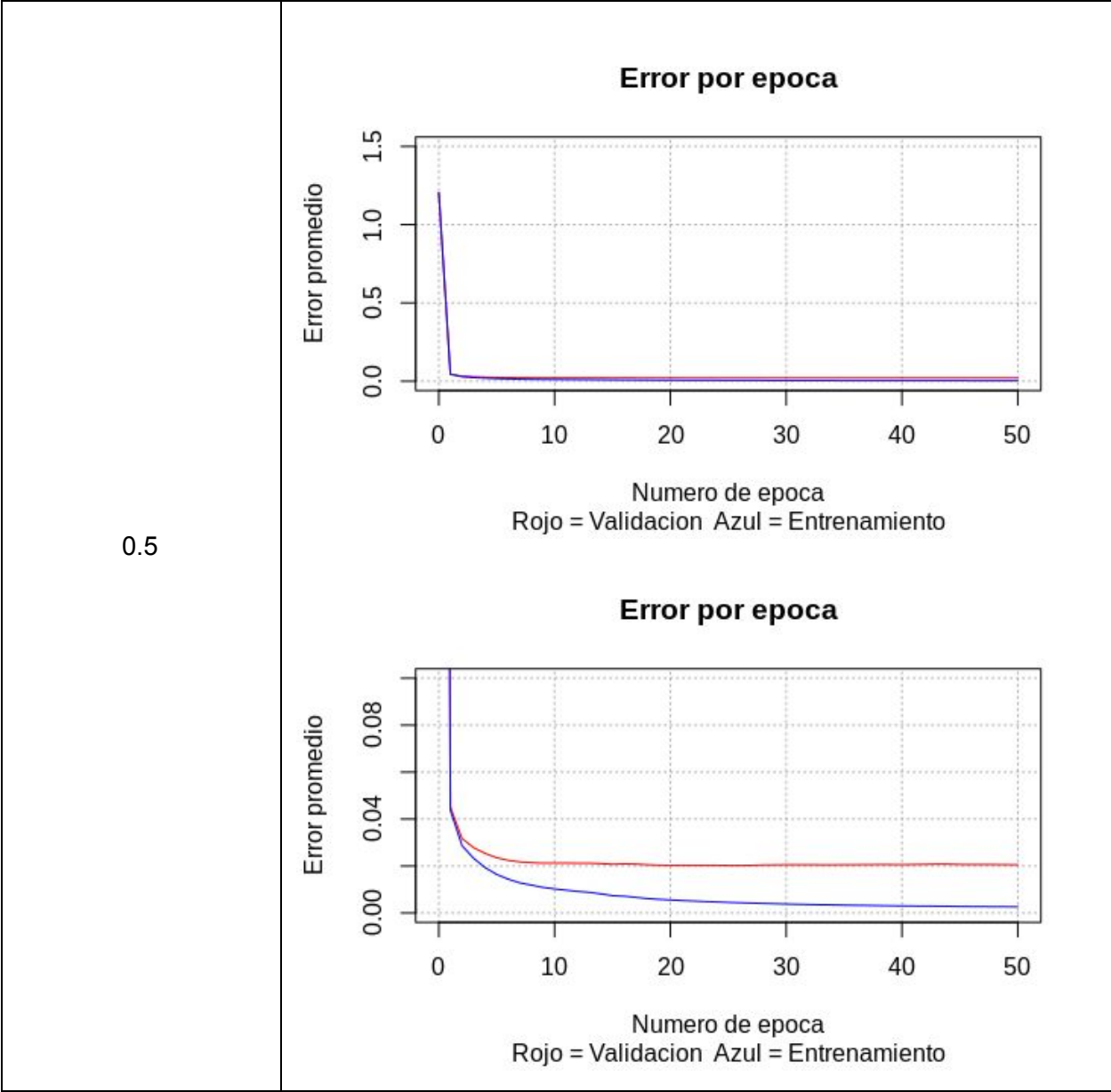
0.25

Error por epoca



Error por epoca





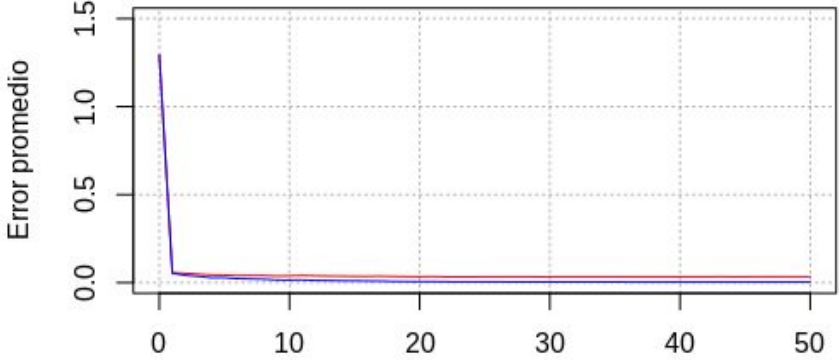
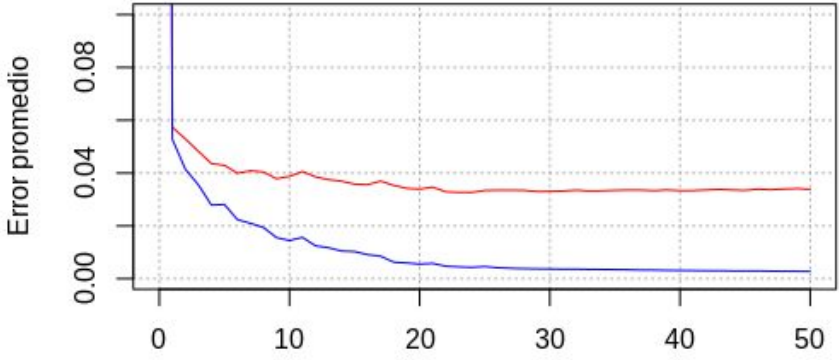
Para comparar con el experimento 1, tomaremos el MLP con 100 neuronas en la capa oculta, ya que este fue el que dio mejor y además tiene la misma cantidad de neuronas que los de este experimento.

Si vemos las gráficas de error notamos que en las de este experimento la curva no tiene cambios drásticos en la pendiente, pareciera estar continuamente disminuyendo, como una gráfica de $\frac{1}{x^2}$, mientras que en el experimento 1 tiende a bajar pero tiene algunos picos que suben. Es difícil ver ya que los valores son sumamente pequeños pero el error de entrenamiento en este experimento al final da entre 0.003 y 0.0025 dependiendo del alpha, y en el experimento 1 da 0.008, por lo cual en estos casos es mejor el error con los datos de entrenamiento, y con los de validación nos da 0.02 en este experimento y 0.024 en el anterior, por lo que aquí también mejora. Aunque como dijimos antes esto no significa que sea mejor, ahora veamos los porcentajes.

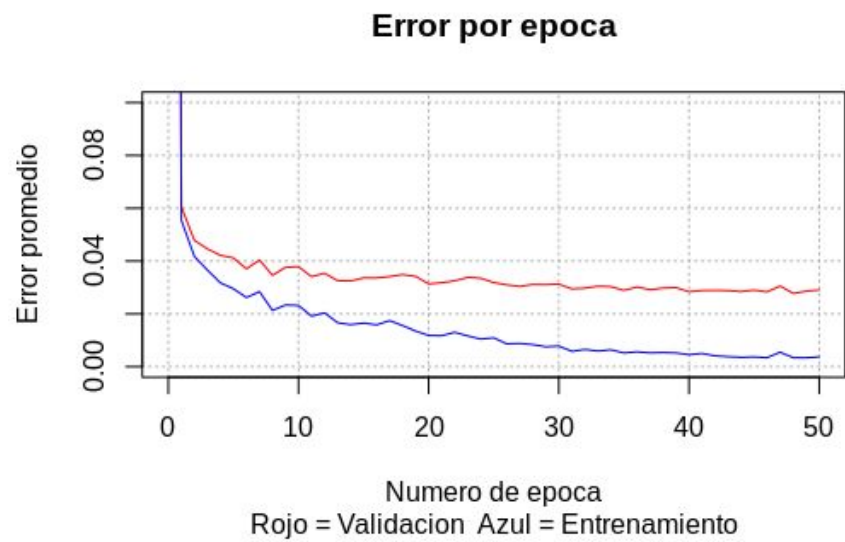
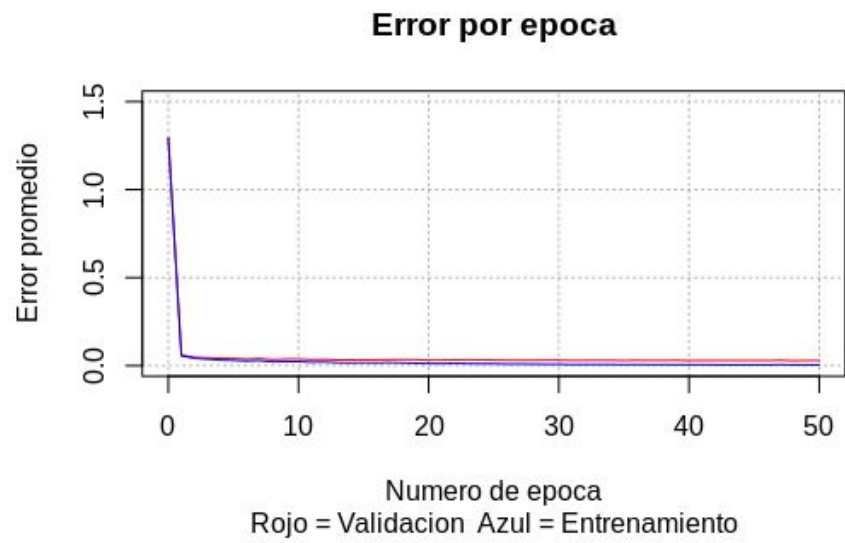
Con el criterio viejo en los 3 casos dio un resultado sumamente parecido, varía en un 0.04%, la diferencia entre estos y el experimento 1 es de menos de 0.2%, por lo cual hay una mejora pero no es lo suficientemente grande como para concluir que este experimento fue mejor. Con el criterio nuevo el mejor fue alpha = 0.5 con 92.21%, el porcentaje aumentó a medida que el alpha aumento, en el experimento 1 obtuvimos un 93.83%, lo que significa una mejora de 1.6%, por lo cual el experimento 1 fue mejor. Se puede observar que con el criterio nuevo, a medida de que el alpha aumenta el porcentaje aumenta, es importante recalcar que lo recomendado para el alpha es un valor entre 0 y 1. Esto nos ayuda a ver la importancia de la constante de momentum en práctica.

Usamos 50 épocas en ambos casos para hacer una comparación más justa, no probamos con distintos números de épocas ya que el algoritmo toma un tiempo considerable en correr. No notamos nada especial con respecto al número de épocas por esto.

No vemos indicios de sobre-entrenamiento ya que con 50 épocas el error de validación no crece, siempre decrece aunque cada vez decrece mucho menos.

Tamaño del conjunto de entrenamiento	Gráfica error promedio vs épocas
1/4	<p data-bbox="878 401 1127 436">Error por epoca</p>  <p data-bbox="760 867 1247 930">Numero de epoca Rojo = Validacion Azul = Entrenamiento</p> <p data-bbox="878 999 1127 1035">Error por epoca</p>  <p data-bbox="760 1465 1247 1528">Numero de epoca Rojo = Validacion Azul = Entrenamiento</p>

1/2



Tamaño del conjunto de entrenamiento	Porcentaje de clasificados correctamente	
	Criterio viejo	Criterio nuevo
1/4	96.02%	90.33%
1/2	96.84%	92.69%

Nuevamente del experimento 1 tomaremos a la red con 100 neuronas, para hacer una comparación más justa.

Para asegurarnos de tener datos de las tres clases ordenamos los datos iniciales aleatoriamente y tomamos el primer cuarto y la primera mitad de los elementos respectivamente.

Al comparar los resultados podemos ver que con ambos criterios mientras menos datos de entrenamiento hay, el porcentaje es menor, como es de esperar. Lo interesante es que con menos datos el porcentaje no baja tanto, como podemos ver en la siguiente tabla.

Tamano del conjunto		1/4	1/2	1
Porcentaje	Criterio viejo	96.02%	96.84%	97.59%
	Criterio nuevo	90.33%	92.69%	93.83%

Con el criterio viejo la diferencia más grande fue de menos de 1% ($\text{pct}(1) - \text{pct}(\frac{1}{2})$ y $\text{pct}(\frac{1}{2}) - \text{pct}(\frac{1}{4})$) y con el criterio nuevo la diferencia más grande fue de 2.36%, aquí es más grande pero sigue siendo pequeña.

Obviamente es mejor entrenarlo con el conjunto completo pero es muy bueno este experimento ya que nos permite ver que disminuyendo el tamaño del conjunto de entrenamiento considerablemente el porcentaje baja muy poco, lo que nos permite tener resultados muy buenos y que el algoritmo sea mucho más rápido porque tiene un conjunto más pequeño.