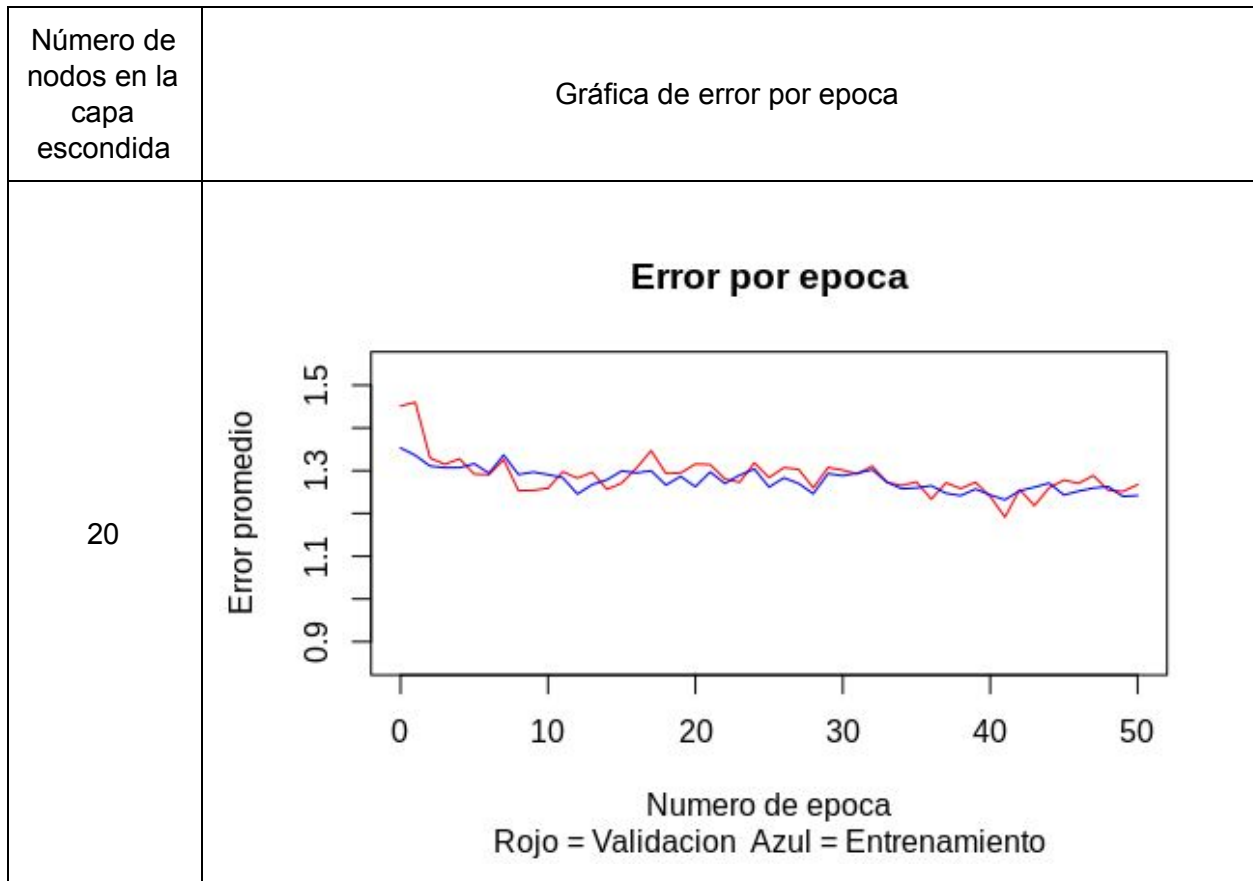
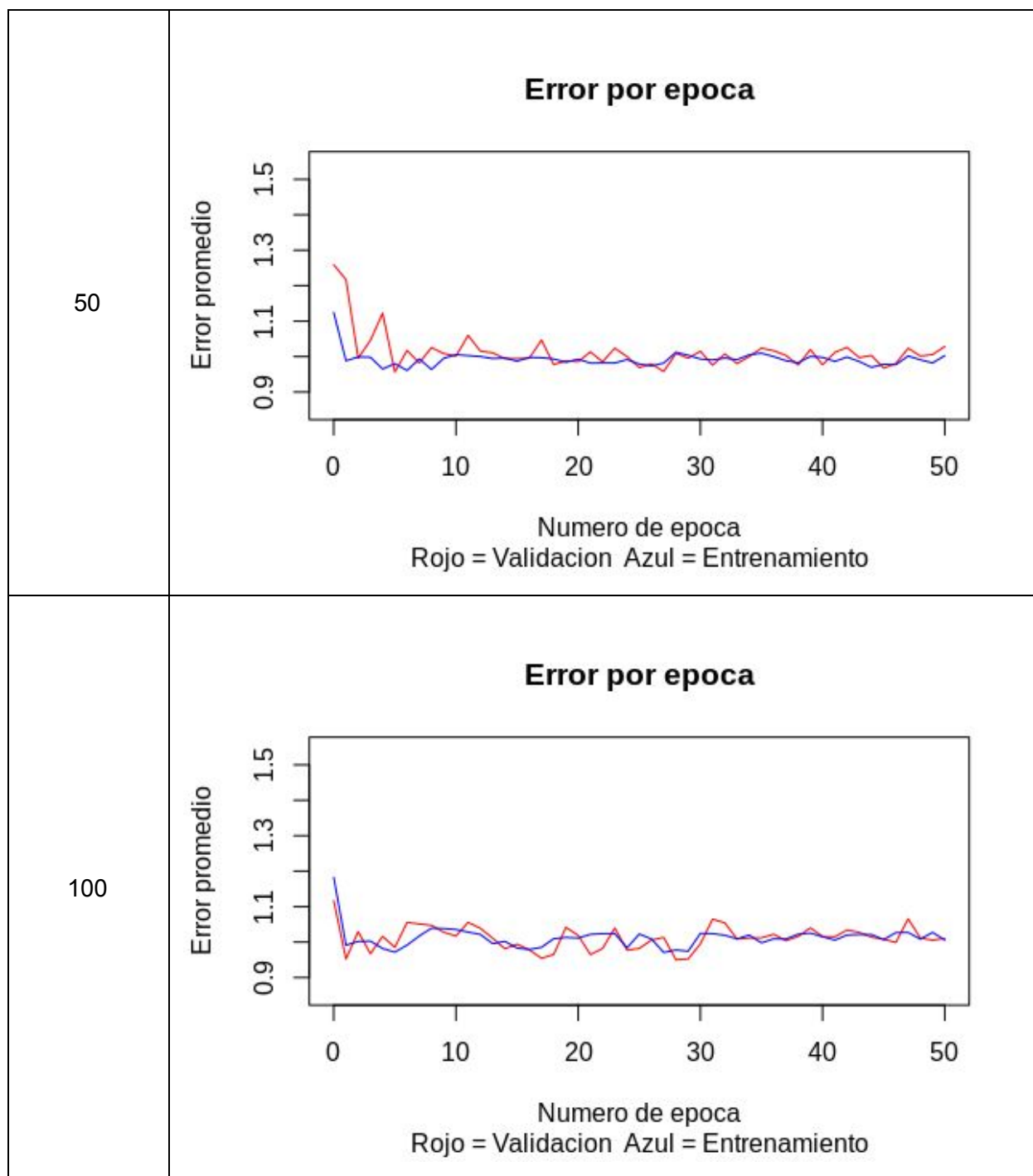


Tarea 4

2)





Analizar el error en la epoca numero 0 no tiene mucho sentido ya que ahi solo depende del azar de los pesos asignados, pero sirve para ver la evolución del algoritmo. Durante todas las épocas relevantes el error de la red con 20 neuronas escondidas fue mayor que el de las otras. Mientras que las de 50 y 100 tienen graficas casi idénticas. En todas las redes el error con los datos de entrenamiento y de validación es muy parecido esto nos ayuda a pensar que no ocurre un sobreentrenamiento o “overfitting”. De estas observaciones podríamos concluir

que 20 es un número bajo, y que a partir de 50 es un número mejor para este problema, no probamos con tamaños mayores que 100 ya que correr el algoritmo con 50 épocas tarda mucho tiempo. También es importante tener en cuenta que el error no lo es todo en estas redes, puede que redes con mayor error funcionen mejor que otras con menor error.

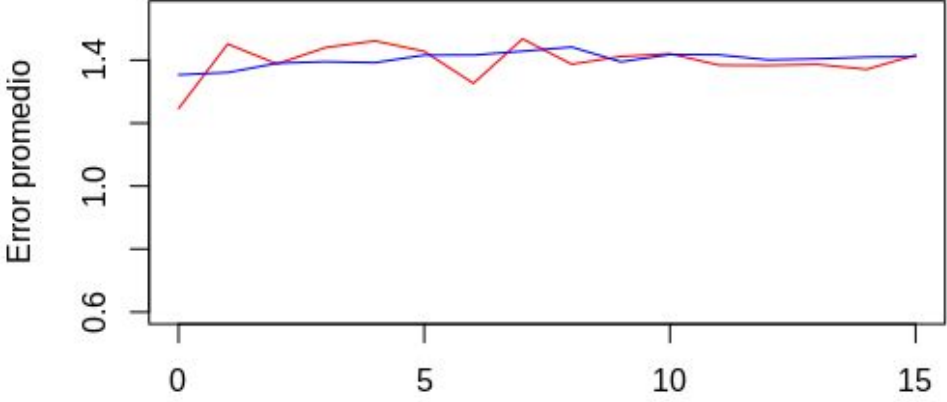
Ahora comparemos con nuestro perceptrón simple y con el adaline.

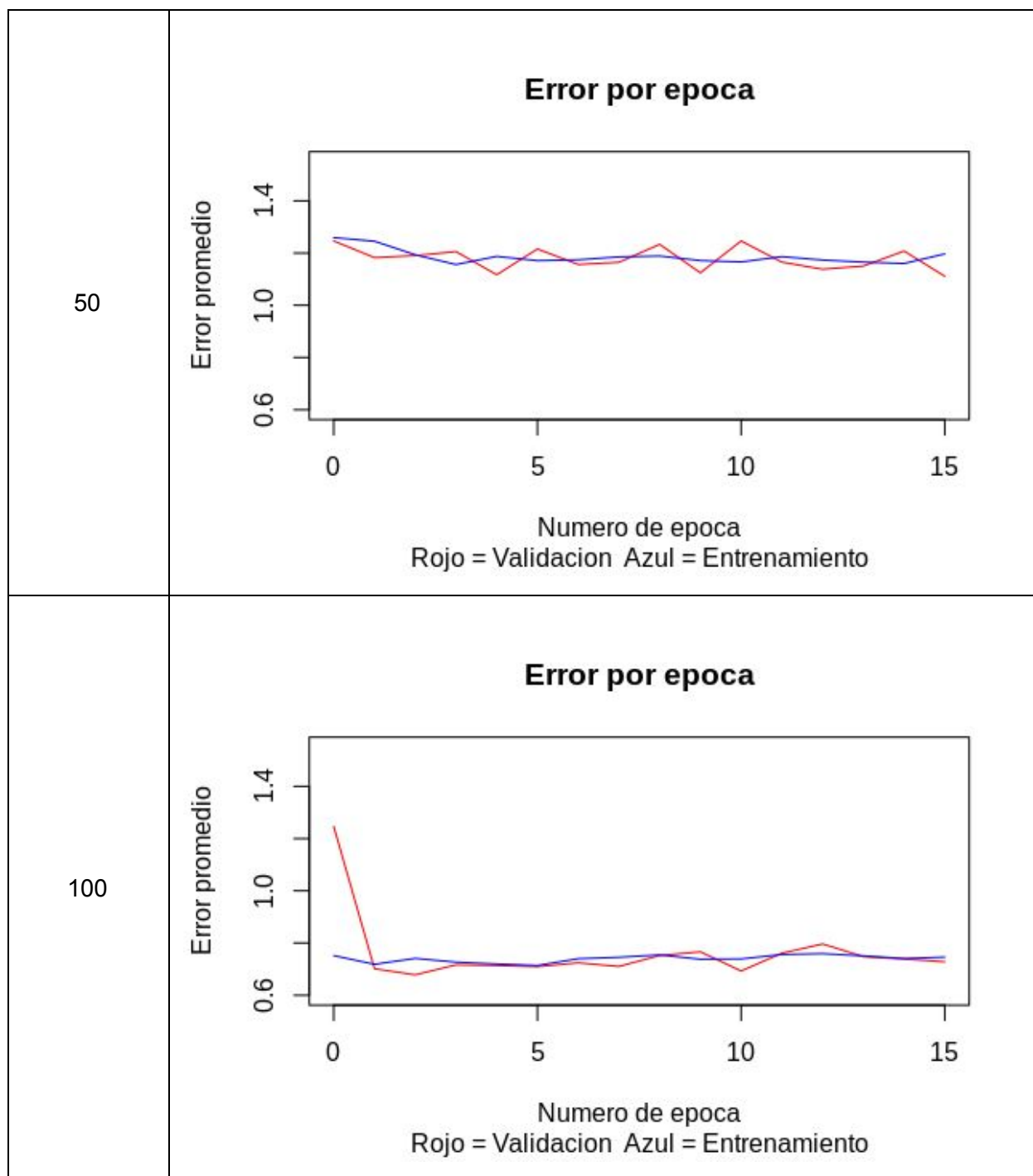
Número de nodos en la capa escondida	Porcentaje de clasificados correctamente	
	Criterio nuevo	Criterio viejo
20	88.84%	94.56%
50	91.90%	96.63%
100	93.57%	97.44%

Podemos observar que con ambos criterios es mucho mejor que las anteriores, sobretodo el de 100 neuronas, esto confirma lo dicho anteriormente de que el error no lo es todo, ya que la de 100 y la de 50 tienen un error idéntico pero la de 100 es mejor. El criterio viejo llegó a clasificar correctamente hasta un 97.44% que es un porcentaje bastante alto, es normal que el viejo de más que el nuevo, ya que en los casos en los que el nuevo da bien el viejo también lo hace y también lo puede hacer en otros que el nuevo no puede. En conclusión el perceptrón multicapas es mucho mejor en esta tarea que el perceptrón simple y el adaline, aunque hay que tener en cuenta que consume más memoria y toma más tiempo en entrenar y producir resultados.

Antes de la seguir con la parte 3, tuvimos que tomar una decisión, el algoritmo con 50 épocas y 100 neuronas en la capa escondida tardaba unas 4 horas en correr, por lo que decidimos solo usarlo para la comparación con las tareas anteriores, los siguientes experimentos nos piden comparar con el primero, por lo cual lo que haremos es volver a correr el primer experimento con 15 épocas y realizar los siguientes con 15 también para que haya una comparación justa.

Los resultados del primer experimento con 15 épocas fueron:

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

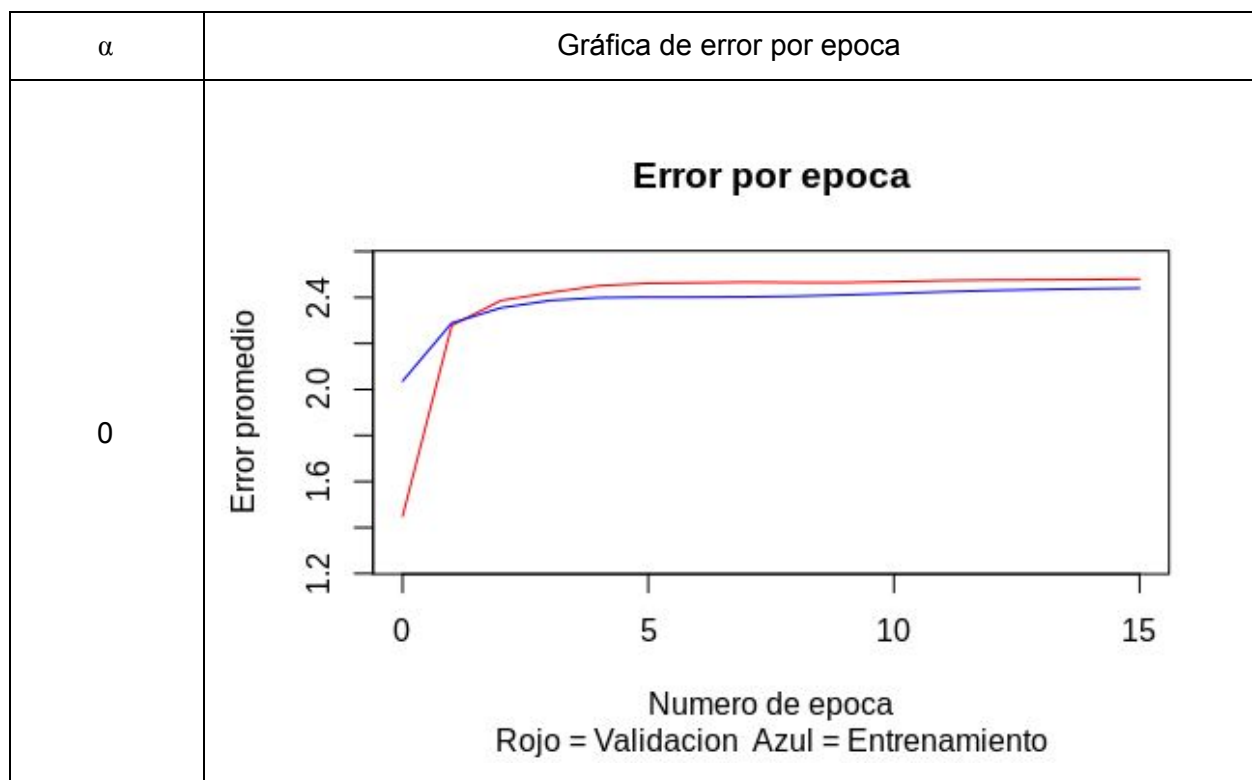


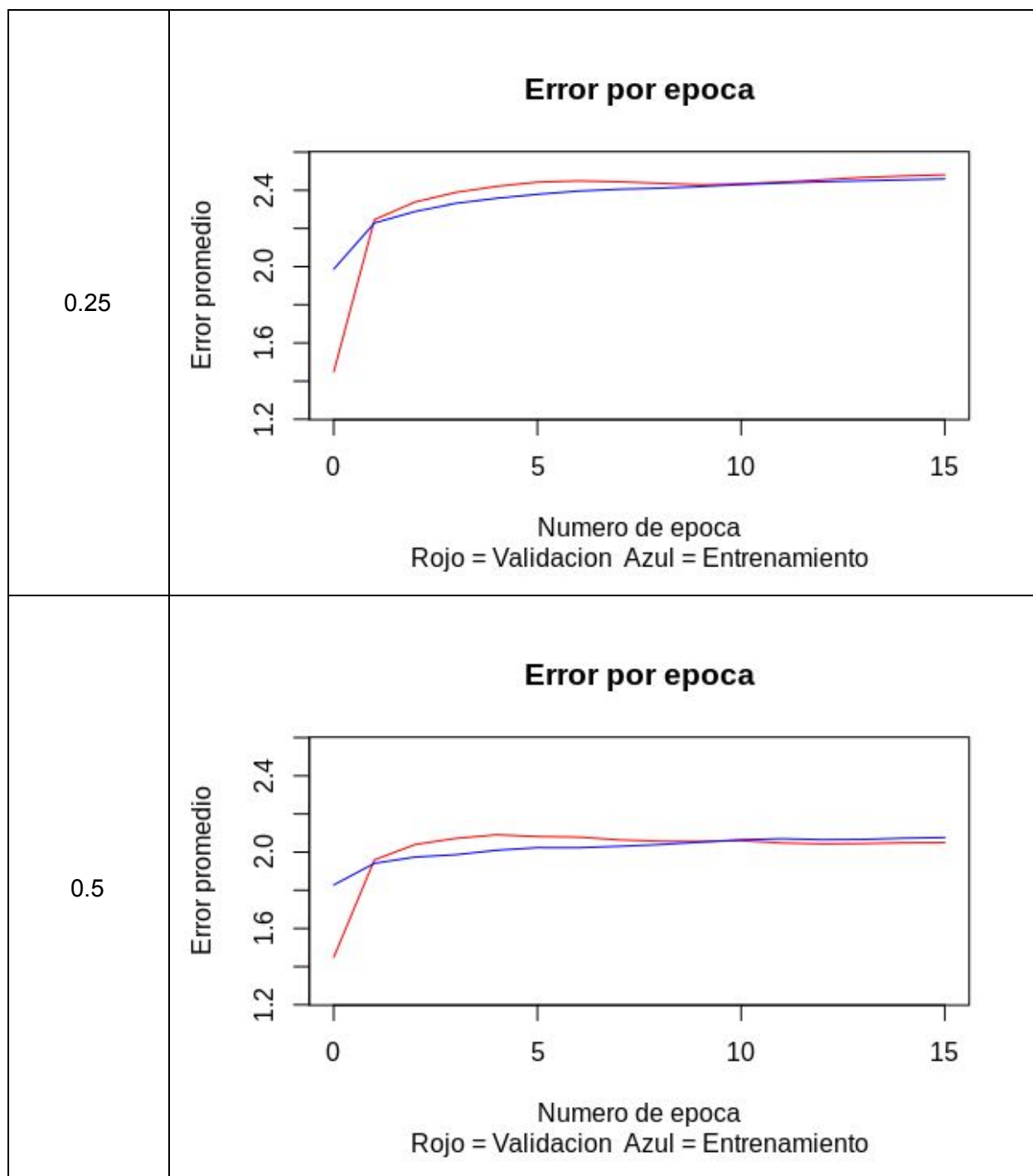
En este caso el error con 50 neuronas subi3 mientras que el de 100 neuronas bajo. El de 20 se mantuvo igual.

Número de nodos en la capa escondida	Porcentaje de clasificados correctamente	
	Criterio nuevo	Criterio viejo
20	83.11%	93.33%
50	89.15%	95.83%
100	91.77%	96.90%

Los porcentajes bajaron un poco como era de esperarse pero siguen siendo buenos.

3)





Claramente el error en estas es mucho más grande que en el experimento 1. Además es interesante notar que las 3 gráficas tienen una forma super parecida. Aquí el 0 tuvo el mayor error, luego el 0.25 y finalmente el 0.5. Pero como dijimos antes el error no lo es todo, podría que teniendo mayor error funcionen mejor que las del experimento 1.

Con 15 épocas no vemos señal de sobreentrenamiento, ya que el error para datos de validación y de entrenamiento es prácticamente igual.

Ahora veamos los porcentajes:

α	Porcentaje de clasificados correctamente	
	Criterio nuevo	Criterio viejo
0	89.50%	97.84%
0.25	89.65%	97.75%
0.5	90.52%	97.79%

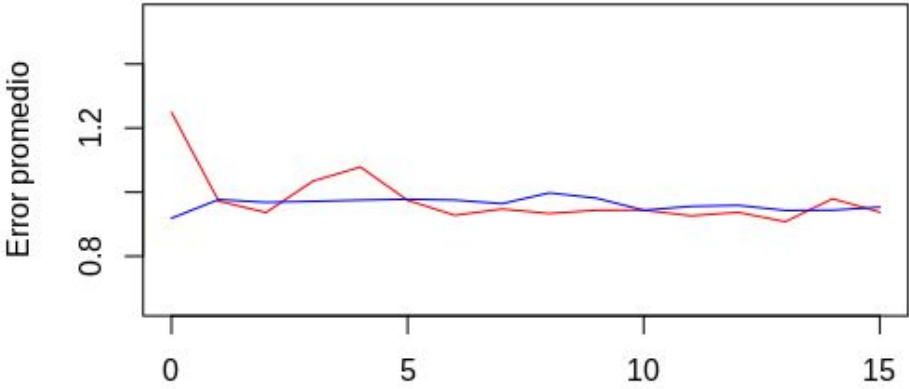
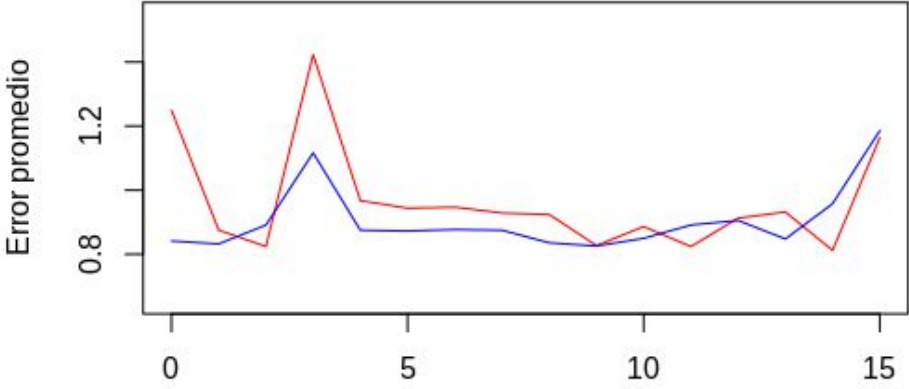
Con el criterio nuevo da un poco peor que el mejor del experimento 1, pero con el viejo da un poco mejor. Tienen el mismo número de neuronas en la capa oculta y misma tasa de aprendizaje que el experimento anterior, pero el anterior tiene $\alpha = 0.9$.

El resultado que mejor dio con el criterio viejo en este experimento fue con la constante de momentum igual a 0, esto significa que no tiene efecto, con un 97.84% que es un porcentaje sumamente alto para 15 épocas. La diferencia entre este y el mejor del experimento 1 es de casi 1%, que no parece tanto, pero al estar tan cerca del 100% hace una gran diferencia. Con la constante igual a 0.25 y 0.5 dieron valores idénticos, también mejores que el experimento anterior, de hecho casi idénticas al valor con la constante igual a 0.

Con el criterio nuevo el experimento anterior dio mejor por poco, una diferencia de casi 2% en el peor caso y 1% en el mejor, esto nos dice que con el criterio nuevo usar una constante igual a 0.9 puede ser mejor.

Es probable que el efecto de la constante de momentum se viera más con un número mayor de épocas, pero como dijimos anteriormente, tuvimos que reducirlas por el tiempo que tomaba.

4)

Tamaño del conjunto de entrenamiento	Gráfica de error por epoca
1/4	<p data-bbox="834 485 1101 520">Error por epoca</p>  <p data-bbox="446 632 479 829">Error promedio</p> <p data-bbox="602 919 1339 955">0 0.8 1.2 15</p> <p data-bbox="846 993 1084 1024">Numero de epoca</p> <p data-bbox="703 1026 1230 1058">Rojo = Validacion Azul = Entrenamiento</p>
1/2	<p data-bbox="834 1165 1101 1201">Error por epoca</p>  <p data-bbox="446 1312 479 1509">Error promedio</p> <p data-bbox="602 1600 1339 1635">0 0.8 1.2 15</p> <p data-bbox="846 1673 1084 1705">Numero de epoca</p> <p data-bbox="703 1707 1230 1738">Rojo = Validacion Azul = Entrenamiento</p>

Al comparar los errores de entrenamiento y validación, vemos que se comportan de una forma similar. El error da un poco mayor al mejor caso del experimento 1 y mucho menos que el peor del 1.

Ahora veamos los porcentajes de clasificación:

Tamano del conjunto	Porcentaje de clasificados correctamente	
	Criterio nuevo	Criterio viejo
1/4	89.87%	96.46%
1/2	90.21%	96.64%

Con el criterio nuevo, este experimento en el mejor caso da un 1,5% peor, y con el criterio viejo da prácticamente lo mismo, solo varia un 0.26%. Estos resultados son muy buenos ya que estamos eliminando una gran parte del conjunto de entrenamiento y aun así estamos obteniendo porcentajes casi idénticos, sobretodo con el criterio viejo. Claramente el experimento 1 da resultados mejores, pero este nos permite hacer el entrenamiento con menos datos, lo que nos permite ahorrar recursos y tiempo, y aún así obtener resultados muy buenos.

Como conclusión final podemos decir que el MLP es muy bueno para resolver este problema, mejora al perceptrón simple en más de 25% y al adaline en más de 10%. Hay que tener en cuenta que toma mucho tiempo y recursos, pero gracias a la parte 4 vimos que con un subconjunto de los datos de entrenamiento podemos obtener resultados muy buenos igualmente. Nuestro mejor resultado fue de 97.84%, pero solo con 15 épocas, quizás con más se podría llegar al 98%, que sería un porcentaje casi excelente.