Exercise 2.d

In figures 3,4,5 are reported the accuracies obtained from training the model for 15 epochs, 3 times for each learning rate and each optimizer, as well as the mean of the results from the 3 runs. Figures 6,7,8 show the difference between optimizers in the evolutions of losses and accuracies, averaged among the 3 trainings. From all this data we can see how the best learning rate is 10^-2, leading to high accuracies with all optimizers. On the other hand, a learning rate of 0.5 performs badly on all except for SGD, where we achieve the highest averaged test accuracy. Differently from other learning rates, here the loss decreases with all optimizers, also with those that get a low accuracy. This might be due to overfitting, since the learning is so high that it immediately converges to the local minima of the training data, creating a model which is not robust to variance at all. Finally, the learning rate of 10^-5 splits the optimizers into two parts: adaptive methods perform well, whereas SGD and SGD with momentum diverge.

Between the two, adaptive methods seem to be slightly more robust to different learning rates, they converge faster, and they achieve higher accuracies.

	1st run		2nd run		3rd run		Mean	
	Training	Test	Training	Test	Training	Test	Training	Test
SGD	0.9999	0.9844	1.0000	0.9849	1.0000	0.9844	1.0000	0.9846
SGDw/Momentum	0.1999	0.2052	0.1043	0.1028	0.1031	0.1135	0.1358	0.1404
RMSProp	0.1010	0.0975	0.1006	0.0982	0.1003	0.0958	0.1006	0.0972
AMSGrad	0.1039	0.1011	0.1297	0.1369	0.1062	0.0892	0.1123	0.1091

Figure 3: learning rate 0.5

	1st run		2nd run		3rd run		Mean	
	Training	Test	Training	Test	Training	Test	Training	Test
SGD	0.9378	0.9395	0.9374	0.9396	0.9362	0.9362	0.9371	0.9384
SGDw/Momentum	0.9945	0.9793	0.9939	0.9792	0.9941	0.9773	0.9942	0.9786
RMSProp	0.9774	0.9675	0.9774	0.9577	0.9800	0.9712	0.9783	0.9655
AMSGrad	0.9956	0.9787	0.9953	0.9780	0.9965	0.9777	0.9958	0.9781

Figure 4: learning rate 10^-2

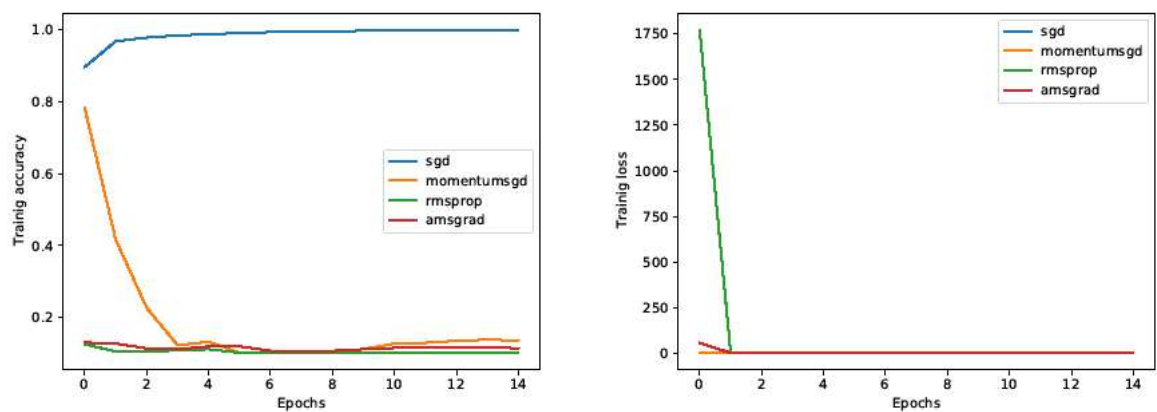|	1st run		2nd run		3rd run		Mean	
	Training	Test	Training	Test	Training	Test	Training	Test
SGD	0.1301	0.1295	0.0931	0.1015	0.1206	0.1222	0.1146	0.1177
SGDw/Momentum	0.3406	0.3480	0.2136	0.2214	0.2672	0.2740	0.2738	0.2811
RMSProp	0.9130	0.9169	0.9134	0.9156	0.9121	0.9147	0.9128	0.9157
AMSGrad	0.9151	0.9168	0.9146	0.9168	0.9150	0.9168	0.9149	0.9168

Figure 5: learning rate 10^-5

Figure 6: average of the 3 runs' losses (right) and accuracies (left) of different optimizers over 15 epochs using a learning rate of 0.5
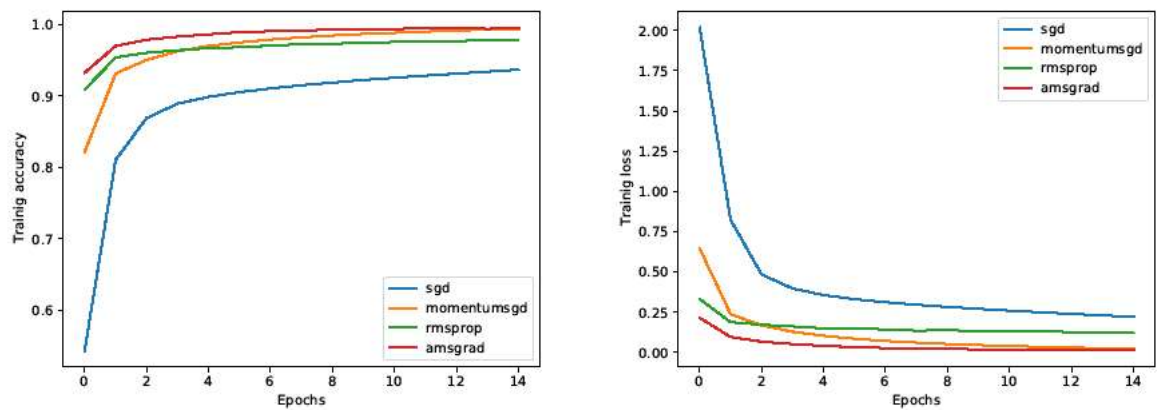

Figure 7: average of the 3 runs' losses (right) and accuracies (left) of different optimizers over 15 epochs using a learning rate of 10^-2
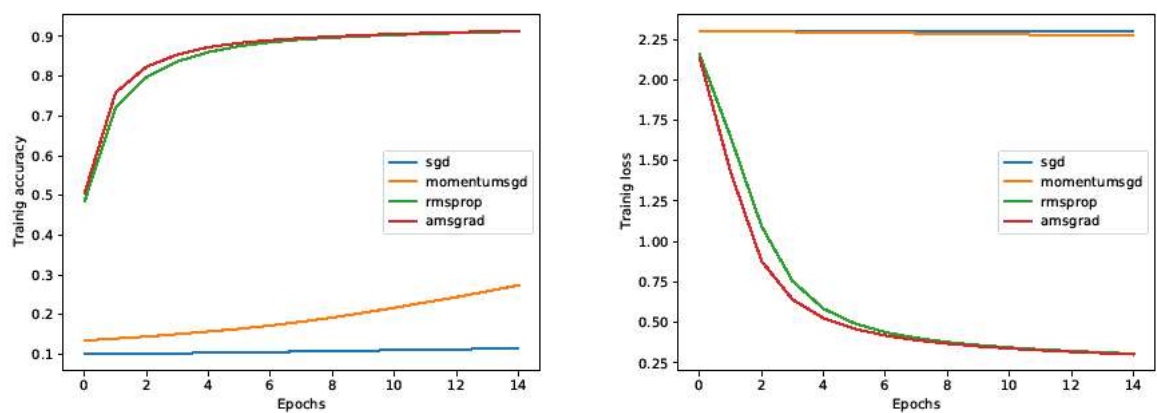

Figure 8: average of the 3 runs' losses (right) and accuracies (left) of different optimizers over 15 epochs using a learning rate of 10^-5