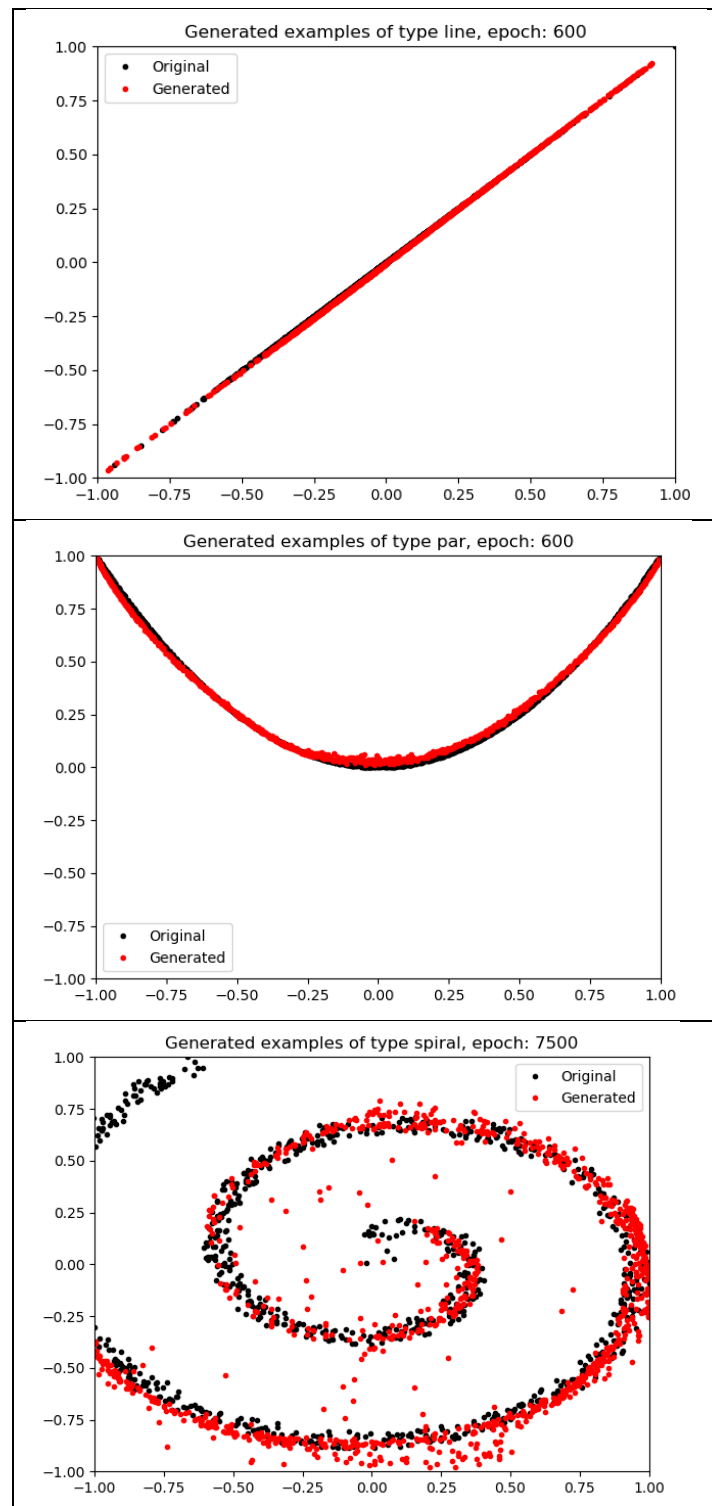# Advanced methods in Machine Learning – exercise 3

Hodaya Koslowsky

The Model:

The model I chose after several non-successful tries is the following:

Generator: A FC network, with 3 hidden layers, and two types of activation functions. The dimensions of the layers are as following:

	Line	Parabolic	Spiral
Input-noise-size	2	64	64
Hidden layer1	128	128	256
Hidden layer2	254	254	512
Hidden layer3	512	512	1024
Output layer	2	2	2

For all three conditions there have been dropout layer and activation functions after each layer. The exact structure is thus:

$$input\ layer \rightarrow hidden1 \rightarrow ReLU \rightarrow hidden2 \rightarrow ReLU \rightarrow dropout \rightarrow hidden3 \rightarrow ReLU \rightarrow dropout \rightarrow output\ layer \rightarrow tanh$$

The architecture and the dimensions of the Discriminator were the same as the 'line' model of the generator, except there were no dropout layers.

Optimizers: both Adam, and with a weight decay of steps that I altered between the three conditions. Generally, the line and the parabolic distributions converged after around 500 epochs, so the weight decay was adjusted to that. For the spiral form, it took many more, so the weight decay happened in steps of roughly 1000 epochs.

Loss function: binary cross entropy.

Difficulties:

In the beginning I used layers sizes that are much smaller: around 5-30 in size. I tuned many types of parameters including different number of layers and layer sizes, activation functions, loss functions, dropout, number of discriminator and generator optimization steps for each epoch, weight decay etc. Eventually the big step that helped me achieved good results is increasing the layer sizes significantly. After this change, the line and parabolic distributions converged very quickly, but the spiral distribution still took many tries to converge.

Plots of the progress over epochs:

Line	Parabole	Spiral