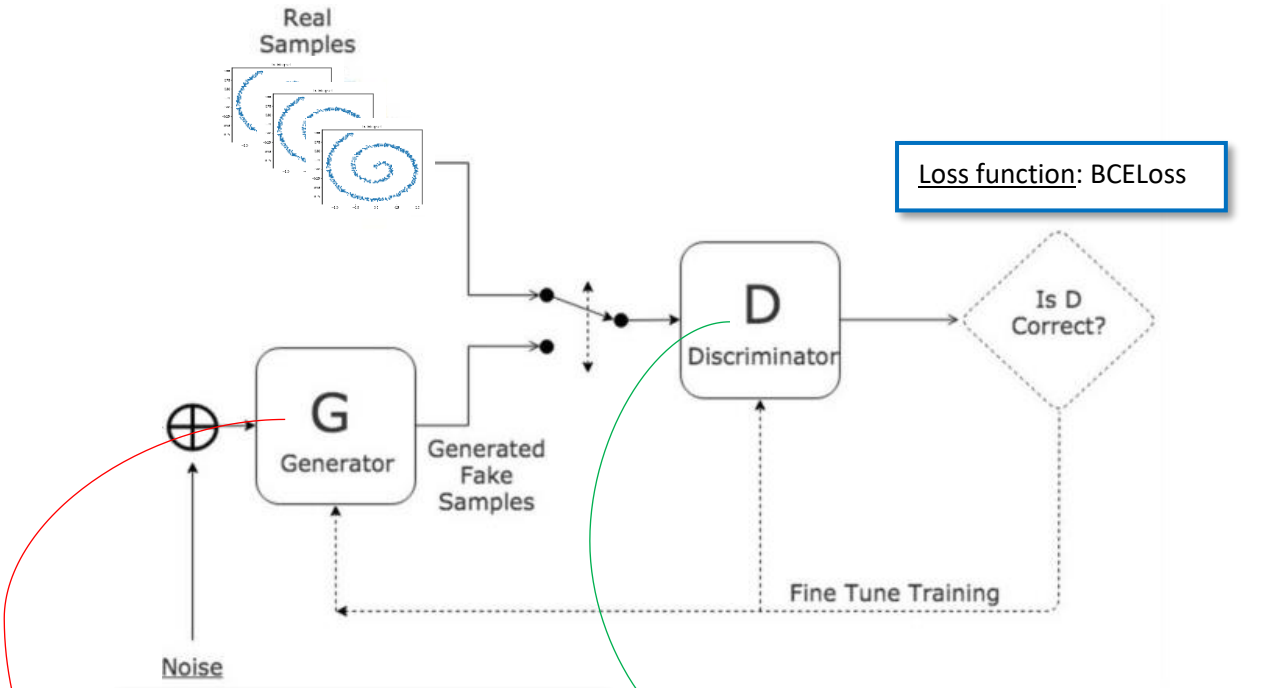
Advanced ML - Ex3

Daniel Juravski 206082323

Eyal Orbach 015369317

Model Architecture:



Feed-Forward Network:

- Layer1 size: 2->20
- Layer2 size: 20->30
- Layer3 size: 30->20
- Layer4 size: 20->2
- Activation Function: relu (last layer linear only)
- Batch normalization before each activation

Optimizer: Adam

Learning Rate: 0.0003

Feed-Forward Network:

- Layer1 size: 2->20
- Layer2 size: 20->30
- Layer3 size: 30->20
- Layer4 size: 20->2
- Activation Function: relu (last layer sigmoid)
- Batch normalization before each activation

Optimizer: Adam

Learning Rate: 0.0001

On that experiment it was all about the tunings. We noticed that the GANs requires a lot of hyper-parameters tuning work. The 'line' data was quite easy for learning, but the 'parabola' and the 'spiral' required some more attentions. That expressed with the Discriminator vs. the Generator conflicts. The 'spiral' also required us to stop using tanh over the last layer, to reach the desired scale.

We tried the next approaches:

- Using different learning rates for each of the networks.
- Using different optimizers (wasn't stable).
- Loop k times over the real and the fake data when training the discriminator (slowed the performance and didn't produce improvement).
- We found out that using the sigmoid over the layer of the discriminator gave good results.
- We found out that tanh helped reach a somewhat similar distribution but didn't reached the desired scale. We switched it eventually to a linear layer with no activation.
- We tried to switch the discriminator loss to its negative form (was not required eventually).
- We tried different types of regularization methods like dropout, L2 and BarchNorm. BarchNorm proved most successful.

Results:

Target (given)	Trained (our)