**ML HW6**

1. Describe the difference between WGAN\* and GAN\*\*, list at least two differences
2. While GAN uses JD divergence as the loss between two probability distribution

WGAN use Wasserstein distance to determine loss

1. The network of WGAN does not use sigmoid activation function
2. Noise is added to cost function in WGAN to prevent gradient vanishing and to stabilize the model
3. Please plot the “Gradient norm” result.

Original statistics are listed below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| layer num | c =0.01 | c = 0.001 | c = 0.1 | gp |
| 1 | -1.341778 | -5.050287 | -1.278818 | -0.564038 |
| 2 | -1.64724 | -6.791865 | -0.157347 | -0.651738 |
| 3 | -2.239508 | -8.784249 | 0.324475 | -0.719529 |
| 4 | -3.124343 | -9.694891 | 0.526916 | -0.970853 |