1. Describe the difference between WGAN* and GAN**, list at least two differences
Ans:

Technically, the difference from (DC−)GAN to WGAN is to remove the last sigmoid function, loss function, weight, and RMSProp as optimizer.

Theoretically, (DC−)GAN is a simple image−generating network and WGAN is using a new distance measurement method based on GAN. To be more specific, its name is Earth Mover distance. This is the reason why changing the loss function and RMSProp optimizer in the implementation.

Moreover, the change from GAN to WGAN is expected to solve GAN's problems. For example, it has no need to "maintain a careful balance in training of the discriminator and the generator or require a careful design of the network architecture neither."

Diff 01 — WGAN sue the Earth Mover distance, which is different from GAN.
Diff 02 — GAN has "mode dropping phenomenon", but WGAN solve this problem.

2. Please plot the "Gradient norm" result.

(a) Use training dataset, set the number of discriminator layer to 4 (minimum requirement)

(b) Plot two setting:

weight clipping
gradient penalty

(c) Y−axis: gradient norm(log scale), X−axis: discriminator layer number (from low to high)

Ans: