Lecture 3 - Exercise sheet

Master in Deep Learning - Generative Models

Pablo Miralles-González, Javier Huertas-Tato

January 16, 2025

Problem 1.

What is the mathematical intuition behind the loss function of the generator $\max_D V(D, G)$?

Solution: As seen in class, it is mathematically equivalent to the **Jensen-Shannon divergence** between the probability distribution of the real data and the probability distribution produced by our generator model. We do not need to understand JS divergence, but just know that it measures how different two probability distributions are in some sense.

Thus, the generator wants to minimize this quantity and make the two probability distributions as similar as possible, so that real data is difficult to distinguish from generated data.

Of course, we need to remember that this is not the real loss function we minimize during training! We alternate between the maximization and minimization problem, not necessarily reaching this $\max_D V(D, G)$ at any point.

Problem 2.

Consider the vanilla GAN and the WGAN. In terms of the mathematical intuition of the loss function (see previous the problem), what is the difference between the two?

Solution: It changes the Jensen-Shannon divergence from GANs between the probability distribution of the real data and the probability distribution produced by our generator model to another distance function between probability distributions: the **Earth Mover's Distance**.

Problem 3.

Could you explain, in your own words, the mode collapse problem?

Solution: Mode collapse occurs when the generator restricts its output to a limited set of possibilities. In the case of MNIST, this could mean producing images of only a select few digits.

Problem 4.

Could you explain, in your own words, the vanishing gradient problem?

Solution: If the discriminator is "too good", then the updates to the generator can become very small, creating a vicious cycle in which the generator does not improve and the discriminator gets even better.

Problem 5.

Why do we add the gradient penalty term to the discriminator loss of WGANs?

Solution: As seen in class, the discriminator must be a 1-Lipschitz function, and the loss gradient penalty term tries to enforce that.

Problem 6.

What are the differences between the discriminators of GANs and WGANs?

Solution:

- **GAN discriminator**:
  - Its output represents the probability of being a real image (normally the real-valued logit before applying the sigmoid).
  - No restrictions on the discriminator are required.
- **WGAN discriminator**:
  - Its output is a real value with no specific meaning, only that greater values mean that the image is more real.
  - The discriminator must be 1-Lipschitz.

Problem 7.

How would you modify the generator and discriminator to create a Conditional GAN? You should be able to generate images of a specific class from a noise vector and the given class.

Solution: We will publish after the assignment, as it is very related!

Problem 8.

Does a Conditional GAN help with mode collapse? Why?

Solution: Yes, it can help, because you force the generator to at least produce images for each one of the classes. Of course, it could still have mode collapse inside each class. For example, in MNIST, we make sure that the generator can produce every digit, but it may always generate the same 1 digit.