Homework 3 – Generative AI (CS/DS 552, Whitehill, Spring 2025)

Collaboration policy: You may complete this assignment with a partner, if you choose. In that case, both partners should sign up on Canvas in a **pre-made group** as a team, and only one of you should submit the assignment. You are permitted to use ChatGPT (or another AI tool) without restriction to help save you time typing boilerplate code, making complicated visualizations, and overcoming tedious syntactic issues. However, *you must fully understand all the work that you submit.*

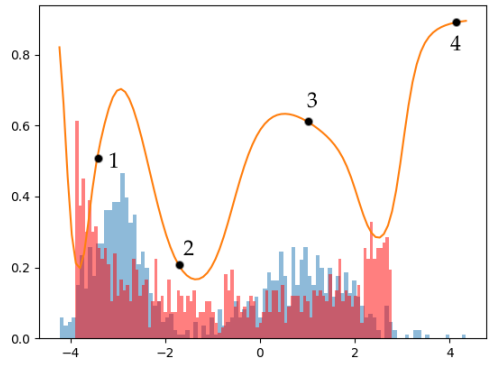
1. **Generative Adversarial Networks (GANs) for Face Generation [32 pts]**

   (a) **GAN from Scratch [24 pts]**: Using the same dataset as in Homework 2 part 1, train a GAN to generate face images. Show a collage of 20 generated faces after training. Report any patterns you observe, e.g., do you see evidence of mode collapse?

   (b) **VAE As a Starting Point [8 pts]**: Now, instead of training the GAN from scratch, use the *decoder* network from the VAE you trained in Homework 2 as a starting point for G. Since $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ for the VAE, the most natural architecture to use for the GAN generator to accept a noise vector from this same distribution. After training, show 20 sampled face images from the GAN, show 20 images for the VAE you trained in Homework 2, and compare them qualitatively. Also compare the images you generate to when you trained the GAN from scratch.

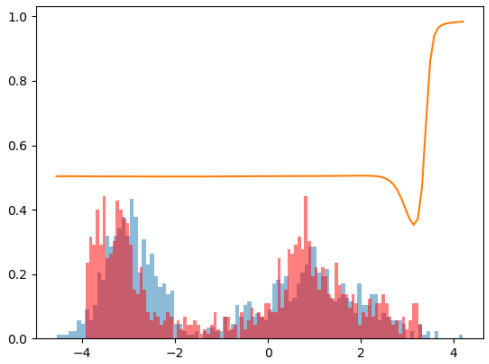
2. **Generative Adversarial Networks (GANs) for 1-d Data [28 pts]**

   (a) **Explain D's Outputs [4 pts]**:
   Suppose we train a GAN to generate 1-d data to match a given data distribution $P_{\text{data}}(x)$. The figure below shows two histograms (red and blue) as well as an orange curve. The blue histogram is the true data distribution $P_{\text{data}}(x)$. The red histogram is the generated data distribution $P(G(z))$ of a trained GAN generator G whose input $z \sim \mathcal{U}(0, 1)$. The orange curve shows the output $D(G(z))$ of the trained discriminator D.



   For each of the four labeled points $1, 2, 3, 4$ in the figure, explain in qualitative terms why the corresponding outputs of the discriminator $D(x_1), D(x_2), D(x_3), D(x_4)$ equal their respective values 0.5, 0.2, 0.6, and 0.9.

   (b) **What Will the GAN Learn [4 pts]**:
   The figure below is similar to the previous exercise but represents the GAN at a later stage of its training. As shown in the histograms, G never produces an output less than about -3.9, nor does it produce an output bigger than about 3.4.



i. For which region ($x < -3.9$ or $x > 3.4$) is it *more likely* that G, through continued training, will learn to produce data similar to the true data distribution, and why?

ii. What is undesirable about the flatness of the output of D for $x < 2.5$?

(c) **Show How a GAN Learns to Approximate $P_{\text{data}}(x)$ [20 pts]**

Similar in spirit to Homework 2, part 2, this is an open-ended exercise: Choose some true data distribution $P_{\text{data}}(x)$ (where, for simplicity and ease of visualization, $x \in \mathbb{R}$), choose a GAN architecture, and show how, over the course of GAN training, the generator can learn to approximate $P_{\text{data}}(x)$ more and more closely. In addition to your code, you should submit a sequence of 10 graphs (one for every few epochs of training) containing the same 2 histograms and orange curve as in the previous exercises, but tailored to your $P_{\text{data}}(x)$, $P(G(z))$, and D.

Submission: Create a Zip file containing all your Python code (which can be in multiple files) and your PDF file, and then submit on Canvas. If you are working as part of a group, then only **one** member of your group should submit (but make sure you have already signed up in a pre-allocated team for the homework on Canvas).