CSCI 1051 Homework 4

February 1, 2023

Submission Instructions

Please upload your solutions by **5pm Wednesday February 1, 2023.** Remember you have 24 hours no-questions-asked *combined* lateness across all assignments.

- You are encouraged to discuss ideas and work with your classmates. However, you **must acknowledge** your collaborators at the top of each solution on which you collaborated with others and you **must write** your solutions independently.
- Your solutions to theory questions must be typeset in LaTeX or markdown. I strongly recommend uploading the source LaTeX (found on the homepage of the course website) to Overleaf for editing.
- Your solutions to coding questions must be written in a Jupyter notebook. I strongly suggest working with colab as we do in the demos.
- You should submit your solutions as a **single PDF** via the assignment on Canvas.

Problem 1 (from January 30)

In class, we motivated diffusion by arguing that GANs can "cheat to win". Why do diffusion models not suffer from mode collapse like GANs do?

Solution to Problem 1

GANs can "cheat to win" because if the generator produces a very good output and fools the discriminator, the generator can learn to only produce that output. Diffusion models do not suffer from mode collapse like GANs do because diffusion gradually adds noise to an image until it is completely random noise. Then, it uses neural networks to recover an image by training the model to predict the noise at each timestep. Therefore, diffusion does not suffer from mode collapse because the structure of the data was retained.

Problem 2 (from January 31)

Refer to the notebook here to solve this problem.

Part 1

In class, we showed that we could find weights with minimum ℓ_2 norm which satisfied the standard linear regression loss, if we followed gradient descent from the right initialization. Plot the optimal weights with minimum ℓ_2 norm and argue that the plot is evidence of our claim.

Part 2

In class, we did a strange re-parameterization of the standard linear regression loss. I told you that we could find weights with minimum ℓ_1 norm which satisfied the modified linear regression loss, if we followed gradient descent from the right initialization. Copying and modifying the code in the notebook, plot the gradient and comment on whether the plot is evidence of our claim.

Solution to Problem 2

Part 1

The point on the plot is evidence of the curve because of the following:

$$w = X^\top (XX^\top)^{-1} y$$

Since X is a 1x2 matrix [2 4], when it is multiplied by $X^\top$, we get 20. Then, taking the inverse, 1/20.

$$w = \begin{bmatrix} 2 \\ 4 \end{bmatrix} (\frac{1}{20})(20) = \begin{bmatrix} 2 \\ 4 \end{bmatrix} = X^\top$$

Therefore, this is evidence of our claim because we solved for the solution to the loss function with the minimum ℓ_2 norm.

Part 2

The plot is evidence of the claim that we can find weights with minimum ℓ_1 norm that satisfies the linear regression loss through following gradient descent because the plot shows that from the initialization point at (3,0), the ℓ_1 norm is minimized along the curve.