Note: All coding problems to be submitted with GitHub Link. Do not Upload the files/folder. Use git commands only.

Note: this is the distribution of questions:

1. Question 1 to Question 3: Required for everyone.
2. Question 4 to Question 5: Bonus question for both Graduate Students and Undergraduate Students

# Problem 1 (10 points)

For each of the following norms, explain what properties will they favor when used in reconstruction error: *L*0, *L*1, and *L*2

* Lp – yields ||x||p = i|p p > 0
* L0 hence is ||x||0 making the summation zero
* L1 ||x||1 = i|. Favors
* L2 ||x||2 =|2, ||x||2 =
* L1 and L2 adds a penalty to the cost, L1 the penalty is equal to the absolute value of the magnitude of coefficient. L2, the penalty is equal to the square of the magnitude of coefficients. L1 provides sparse solutions.

# Problem 2 (30 points)

Given a set of contrast images with sharp geometric edges (e.g., photo-lithography masks for microprocessor manufacturing) write down a formulation for recon- struction error that would work best. Justify your choice.

L2 Loss, otherwise known as squared value, given outliers the model will try to fit the data points better, the errors are squared by the outliers, leading to the model really wanting to get these .

L2 = ||W||^2, J(w|X,y) = J(w|X,y) + α(1/2)w^Tw. The corresponding gradient yielding J(w|X,y) = αw + (J(w|X,y). Updating the gradient w <- w - €(αw + J(w|X,y)), w <- (1-€)w - €(J(w|X,y)).

# Problem 3 (20 points)

Given a set of images of wildlife taken in their natural habitat write down a formulation for reconstruction error that would work best. Justify your choice.

L1 Loss, otherwise known as absolute value loss, does not go out of its way to change the outlier values, it treats them just as individual data points so as not to compromise the model. This might lead to poor predictions every now; however, we would not want to completely change the output of the images as we want to represent the wildlife images as they are.

||w||1 = , J(w|X,y) = α||w||1 + J(w|X,y), taking the gradient yields J(W|X,y) = sign(w) + J(X|y|w), sign(w) is simply the sign of w applied element-wise.

A quadratic cost function represented through Taylor series.

J(w) = H(w-w\*), with H being the Hessian matrix of J w/respect to w evaluated at w\*. With Hessian being diagonal, H = diag(|H1,1,…,Hn,n) where each Hi,I > 0.

The quadratic approximation of L1 decomposes into a sum over the parameters:

J(w|X,y) = J(w\*|X,y) + 2 +

Wi = sign(wi\*)max{(|wi\*| - ,0})

# Problem 4 (40 points)

Given distributions p and q. If q is parameterized by *θ*, how would you choose the value for *θ* to make q closest to p among all possible q’s.

1. Write down formulation of how would you measure the closeness of *q* to *p*.
2. Explain what you would do to maximize this closeness (i.e. make *q* and *p*

maximally close, or minimally different or divergent)

Write a report on one of the following topics related to GANS:

1. InfoGAN https://arxiv.org/abs/1606.03657
2. CycleGAN https://arxiv.org/pdf/1703.10593.pdf