Intro Deep Learning: Homework 5

Daniel Novikov

April 20th, 2022

Problem 1

For each of the following norms, explain what properties will they favor when used in reconstruction error: L_0, L_1, and L_2

Answer:

- L_0 norm counts the number of nonzero weights, favoring sparser reconstructions.
- L_1 norm is the sum of absolute values of the weights, which penalizes solutions with large weight values. Makes the optimization criterion convex.
- L_2 norm is the euclidean length of the weights, which penalizes solutions with large weights even more than L_1, makes the optimization criterion convex.

Problem 2

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

Answer:

I would use Mean Square Error with L_0 regularization (favoring sparse reconstructions). The contrast images with sharp geometric edges are themselves sparse images, which is why L_0 norm will help the model reconstruct the images sparsely with clear edges rather than smoothing over edges and creating blurry regions.

$$J(w) = \frac{1}{m} \sum_{x} [(x - f(x; w))^2] + L_0(w)$$

Problem 3

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

Answer:

I would use Mean Square Error with L_1 regularization (favoring small weight values). Wildlife in their natural habitat may blend in with their surroundings, so our reconstruction loss should favor solutions which preserve boundaries between objects without the over-smoothing present with L_2 regularization, while not pushing sparsity too much either as with L_0, because the entire image should have color rather than just the edges.

$$J(w) = \frac{1}{m}\sum_{x}[(x - f(x; w))^2] + L_1(w)$$

Problem 4

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. (a) Write down formulation of how would you measure the closeness of q to p. (b) Explain what you would do to maximize this closeness (i.e. make q and p maximally close, or minimally different or divergent)

Answer:

Find the Q that minimizes KL Divergence:

$$D_{KL}(P||Q) = \sum_{x\in\chi} P(x)log(\frac{P(x)}{Q(x)})$$

KL Divergence is minimized when p and q are maximally close: the ratio of probabilities of event x occurring in distributions P and Q is 1, the logarithm evaluates to 0 and the KL divergence is 0.