ECE/CS 559: Neural Networks **Fall 2024**

Mock Final

Wednesday December 11, 2024 — There are **7 questions**; the actual exam will have 5.

Note: ***Please write your name on every page!*** *You are expected to work on your own on this exam. This exam is* ***closed notes*** *and no calculators are allowed. Other electronic devices or communication with others are also* ***not*** *allowed. Include your reasoning, not just the final answer. Be clear and concise. Watch your time. If stuck, move on then come back later.* ***Good luck!***

NAME:

— You may use this page and backs of pages for scratch work. —

1. Consider the following data points in $\mathbb{R}^2$: $\begin{bmatrix} -0.01 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0.01 \\ -1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$.

   (a) (3 pts) Perform, step by step, Lloyd's algorithm for k-means with $k = 2$ until convergence, starting with the second and third points as initial centers. Express the final centers clearly. Draw the initial and final Voronoi cells (along with the data points) on the same plot.

   (b) (2 pts) Suggest a different initialization that will not converge to the same final centers.

2. Consider the following data points in $\mathbb{R}^2$: $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$.

   (a) (1 pt) State the competitive (winner-take-all) learning rule in clear mathematical notation. What "kind" of rule is this?

   (b) (4 pts) Perform this learning rule, with one difference: *after every update, round weights to the nearest integer*. Process the data in the given order, for 2 epochs. Use $\eta = \frac{1}{2}$ and initialize with weights $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Give the clustering interpretation for the final weights.

3. Consider a simple single-layer CNN, which takes as input a 2×2 image and produces two 2×2 channels, by applying kernels of width 1 with bias.

   (a) (2 pts) Describe the parameters of this CNN and write the forward equations from the input to the two channels.

   (b) (3 pts) Write the backpropagation equations for this CNN layer, starting with the gradients of the loss at the output. (You don't need to know the loss.)

4. Consider an autoencoder, with input x, embedding z, and output $\hat{x}$. Model the prior of the embedding to be a standard Gaussian vector. Let the encoder be a neural network parametrized by W producing an output $g(x; W)$ plus a standard Gaussian noise. Let the decoder be a neural network parametrized by V producing an output $f(z; W)$ plus a Gaussian with a *learnable covariance matrix* Σ.

   (a) (2.5 pts) Write the general form of the ELBO loss and specialize it to this autoencoder, simplifying it in the same way as the Gaussian autoencoder in class.

   (b) (2.5 pts) In your homework, you concatenated z and the reconstruction error $x - \hat{x}$. Explain why this won't work here. Then, use the fact that you can write $\Sigma^{-1} = M^\intercal M$ to suggest a simple modification that would work.

5. Consider a GAN to generate a real-valued random variable X. Say the population of X is distributed in $[0,1]$ with density $f_X(x) = 2 - 2x$. Use a noise Z uniformly distributed in $[0,1]$. Use the idealized GAN loss we considered in class:

$$\mathop{\mathbf{E}}_{X,Z}[-\log D(X) - \log(1 - D(G(Z)))]$$

   (a) (2.5 pts) Say the initial generator is just the identity function $G(z) = z$. What is the optimal discriminator in this case?

   (b) (2.5 pts) What is the optimal generator G? That is, find G such that $G(Z)$ has the same distribution as X.

6. Say we would like to build an ML model that takes as input Arabic return in Latin script, and outputs the same text in Arabic script. When written in Latin script, many Arabic letters require multiple characters to represent,

   (a) (1 pt) What kind of sequence model is most appropriate for this problem? Explain why.

   (b) (4 pts) Sketch an RNN architecture that you could use for this model. Draw a diagram with embeddings, units, inputs, and outputs clearly marked. Write a simple input-state-output equation for each unit.

7. Say you have a graph $\mathcal{G} = (V, E)$ consisting of vertices $v \in V$ and edges $E \subset V^2$, the subset of pairs of vertices that are connected. You want to make a binary classification (e.g., is there a disturbance or not) that depends on the inputs x_v, sitting at each node.

   (a) (2 pts) You decide to design a *graph* self-attention, by remaining consistent with the graph at any given layer, i.e., uses only inputs connected to each vertex. Write the self-attention equations at a node v in clear mathematical notation.

   (b) (3 pts) Explain (1) how you would build transformer layers using this modified self-attention, (2) whether or not the outputs of the last layer will only depend on neighboring inputs in the graph, and (3) how you would use the last layer outputs to make the binary classification with a clear description of the architecture and loss function.