Homework 03

ECE 449/590, Fall 2022

Due Date: 11/09 by the end of the day (Chicago time)

1. *(30 points)* Let $\{x^{(1)}, \ldots, x^{(m)}\}$ be sampled i.i.d. from a Gaussion distribution of mean μ anf variance σ^2. Let $\hat{\mu}_m = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$ be an estimator of Gaussian mean μ. Show that $\mathrm{Var}(\hat{\mu}_m) = \frac{\sigma^2}{m}$.

2. *(70 points)* Considering the following neural network with inputs (x_1, x_2), outputs (z_1, z_2, z_3), and parameters $\boldsymbol{\theta} = (a, b, c, d, e, f, i, j, k, l, m, n, o, p, q)$,

$$\begin{aligned}
\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} &= \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}, \\
\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} &= \begin{pmatrix} \mathrm{ReLU}(g_1) \\ \mathrm{ReLU}(g_2) \end{pmatrix}, \\
\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} &= \begin{pmatrix} i & j \\ k & l \\ m & n \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} o \\ p \\ q \end{pmatrix}.
\end{aligned}$$

   A. *(15 points)* For a minibatch containing a single training sample $(x_1, x_2, y = 2)$, apply softmax and write down the cross-entropy loss function $J(\boldsymbol{\theta})$ as a function of (z_1, z_2, z_3). Compute $\frac{\partial J}{\partial z_1}, \frac{\partial J}{\partial z_2}, \frac{\partial J}{\partial z_3}$ as functions of (z_1, z_2, z_3).

   B. *(20 points)* Base on A., apply backprop to compute $\frac{\partial J}{\partial i}, \frac{\partial J}{\partial j}, \frac{\partial J}{\partial k}, \frac{\partial J}{\partial l}, \frac{\partial J}{\partial m}, \frac{\partial J}{\partial n}, \frac{\partial J}{\partial o}, \frac{\partial J}{\partial p}, \frac{\partial J}{\partial q}, \frac{\partial J}{\partial h_1}, \frac{\partial J}{\partial h_2}$.

   C. *(20 points)* Base on B., apply backprop to compute $\frac{\partial J}{\partial a}, \frac{\partial J}{\partial b}, \frac{\partial J}{\partial c}, \frac{\partial J}{\partial d}, \frac{\partial J}{\partial e}, \frac{\partial J}{\partial f}$. Explain why you don't need to compute $\frac{\partial J}{\partial x_1}$ and $\frac{\partial J}{\partial x_2}$.
   (Hint: use the step function $u(x)$ as the derivative of $\mathrm{ReLU}(x)$.)

   D. *(15 points)* For the learning rate ϵ, show the equation to apply the simple SGD algorithm to update $\boldsymbol{\theta}$ for this minibatch.