HW Week 5: Normalization, ResNet and Multigrid
(*due date: 06/19/2020*)

MATH 497 Summer 2020

Problem 1 Assume that we have the following data

$$X = \{x_1, x_2, \cdots, x_N\},$$

where

$$x_i = \begin{pmatrix} [x_i]_1 \\ [x_i]_2 \\ \vdots \\ [x_i]_d \end{pmatrix} \in \mathbb{R}^d.$$

Then consider the following normalization operation via each component

$$[\tilde{x}_i]_j = \frac{[x_i]_j - [\mu_X]_j}{\sqrt{[\sigma_X]_j}},$$

where

$$[\mu_X]_j = \frac{1}{N}\sum_{i=1}^{N}[x_i]_j, \quad [\sigma_X]_j = \frac{1}{N}\sum_{i=1}^{N}([x_i]_j - [\mu_X]_j)^2.$$

1. (**10%**) Prove that

$$[\mu_{\tilde{X}}]_j = 0 \quad \text{namely} \quad \frac{1}{N}\sum_{i=1}^{N}[\tilde{x}_i]_j = 0,$$

   for any $j = 1, 2, \cdots, d$.

2. (**10%**) Prove that

$$[\sigma_{\tilde{X}}]_j = 1 \quad \text{namely} \quad \frac{1}{N}\sum_{i=1}^{N}\left([\tilde{x}_i]_j - \frac{1}{N}\sum_{i=1}^{N}[\tilde{x}_i]_j\right)^2 = 1.$$

   for any $j = 1, 2, \cdots, d$.

Problem 2 Define

$$f(x, W) = h(W * x),$$

where

$$W \in \mathbb{R}^{3\times 3} \quad \text{and} \quad x \in \mathbb{R}^{4\times 4},$$

the convolution is done for one channel with stride one and zero padding as discussed in class and

$$h : \mathbb{R}^{4\times 4} \mapsto \mathbb{R},$$

defined by

$$h(Y) = \sum_{1\le i,j\le 4} e^{Y_{ij}},$$

for $Y \in \mathbb{R}^{4\times 4}$.

Consider the following kernel

$$K = \begin{pmatrix} 0 & 1 & -1 \\ 0 & -1 & -1 \\ 0 & 0 & 2 \end{pmatrix}$$

and the following 4×4 image (tensor)

$$g = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{pmatrix}. \tag{1}$$

1. (**10%**) Calculate the derivative of h at $K * g$

$$\left.\frac{\partial h(Y)}{\partial Y}\right|_{Y=K*g} \in \mathbb{R}^{4\times 4}$$

2. (**10%**) Calculate $\left.\frac{\partial f}{\partial x}\right|_{W=K,x=g}$ and $\left.\frac{\partial f}{\partial W}\right|_{W=K,x=g}$.
   Hint: Use chain rule.

Problem 3 Consider the following Parametric Rectified Linear Unit (PReLU) function

$$\text{PReLU}(t) = \begin{cases} t, & \text{if } t \ge 0 \\ at, & \text{if } t < 0, \end{cases}$$

for some $a \in \mathbb{R}$. Let assume that X is an random variable on $\mathbb{R}$ with symmetric probability density function $p(x)$, i.e.

$$p(x) = p(-x), \quad \forall x \in \mathbb{R}.$$

1. (**10%**) Prove that

$$\mathbb{E}[X] = 0.$$

   Hint: Use the definition of expectation and the symmetric property of $p(x)$.

2. (**10%**) Further prove that

$$\mathbb{E}[(\mathrm{PReLU}(X))^2] = \frac{1+a^2}{2}\mathbb{V}[X].$$

Hint: Follow the proof of ReLU case in the notes.

Problem 4 Consider the convolution for one channel with stride one and zero padding. Given a kernel $A = [-1, 2, -1]$, we recall that λ is an eigenvalue of A and $\xi \in R^n \setminus \{0\}$ is a corresponding eigenvector if

$$A * \xi = \lambda \xi.$$

1. (**15%**) Verify that all the n eigenvalues, λ_k, and the corresponding eigenvectors, $\xi^k = (\xi^k_j)$, of $A*$ can be obtained, for $1 \leq k \leq n$, as follows:

$$\lambda_k = 4\sin^2\frac{k\pi}{2(n+1)}, \quad \xi^k_j = \sin\frac{kj\pi}{n+1}\ (1 \leq j \leq n).$$

2. (**5%**) Prove that the eigenvectors $\xi^k = (\xi^k_j)$, $1 \leq k \leq n$ are orthogonal, namely

$$(\xi^k, \xi^l) = \sum_{j=1}^{n} \xi^k_j \xi^l_j = 0 \text{ if } k \neq l.$$

Problem 5 Consider the convolution for one channel with stride one and zero padding $A* : R^{n\times n} \mapsto R^{n\times n}$

$$A * u = f, \tag{2}$$

where

$$A = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}.$$

Consider following two iterative methods for equation (2).

Given u^0,

$$\begin{aligned} &\text{for} \quad \ell = 1, 2, \cdots, 2m \\ &\qquad u^\ell = u^{\ell-1} + S_0 * (f - A * u^{\ell-1}) \end{aligned} \tag{3}$$

with $S_0 = \frac{1}{8}$.

And

Given $\tilde{u}^0 = u^0$,

$$\begin{aligned} &\text{for} \quad \ell = 1, 2, \cdots, m \\ &\qquad \tilde{u}^\ell = \tilde{u}^{\ell-1} + S_1 * (f - A * \tilde{u}^{\ell-1}) \end{aligned} \tag{4}$$

with

$$S_1 = \frac{1}{64}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 12 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{5}$$

1. (**15%**) Prove that when $m = 1, u^2 = \tilde{u}^1$.
2. (**5%**) Prove that for any $m, u^{2m} = \tilde{u}^m$.