HWK4 part 3

1. Consider a dataset with 1000 categories and images which are 256 by 256 RGB images. Suppose that we process all this with a one-layer network without bias.

   (a) Let W be the weight matrix. And let $\mathbf{x}$ be an image that has been flatten out. What are the sizes of W and $\mathbf{x}$?

   (b) Write down the formula that describe how a single flattened-out image $\mathbf{x}$ is being processed by the network and how the weight matrix is being updated.

   $$\mathbf{y} =$$

   $$\mathbf{p} =$$

   $$\mathbf{e} =$$

   $$W = W+$$

   (c) Explain intuitively the meaning of the last formula (explain it in term of template matching)

   (d) Write down the formula that describe how a minibatch X of 200 flattened out images is being processed by the network and how the weight matrix is being updated.

   $$Y =$$

   $$P =$$

   $$E =$$

   $$W = W+$$

(e) Make sure that you understand how the outer product appearing in the equations without batch becomes a matrix multiplication in the equation with batch.

(f) What are the sizes of the matrices X, Y, P and E used in the above equations? What should be the sizes of these matrices when you use Pytorch?

2. Compute the gradient of the following functions:

(a) $\boxed{f(x,y,z) = xy + yz + xz}$. Find $\nabla f(x,y,z)$ as well as $\nabla f(1,2,5)$.

(b) $\boxed{f(x,y,z) = \sqrt{x^2+y^2+z^2}}$. Find $\nabla f(x,y,z)$ as well as $\nabla f(1,2,5)$.

(c) $\boxed{f(x_1,x_2,x_3,\ldots,x_n) = w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n}$ where the w_i's are constants. Find $\nabla f(x_1,x_2,x_3,\ldots,x_n)$.

(d) $\boxed{L(w_1, w_2, w_3) = A(w_1)^2 + B(w_2)^2 + C(w_3)^2}$ where A, B and C are constants. Find $\nabla L(w_1, w_2, w_3)$

(e) $\boxed{L(w_1, w_2, w_3) = \log\left(e^{w_1 x_1 + w_2 x_2 + w_3 x_3}\right)}$ where the x_i's are constants. Find $\nabla L(w_1, w_2, w_3)$

(f) $\boxed{L(w_{11}, w_{12}, w_{21}, w_{22}) = x_1 p_1 w_{11} + x_1 p_2 w_{12} + x_2 p_1 w_{21} + x_2 p_2 w_{21}}$ where the x_i's and p_i's are constants. Find $\nabla L(w_{11}, w_{12}, w_{21}, w_{22})$.