HW #5 Due: 6/13/2019

1. The following is a simplified version of the variational autoencoder (VAE) to explain the reparameterization trick. In the figure, a node with "x" means multiplication, with "+" means addition, and with "exp" means $out = \exp(in)$. All other nodes use the sigmoid activation function. In addition, ℓ_1 is a random number generated from a Gaussian random variable with zero mean and uni-variance. Let the loss function $\mathcal{L} = -(\mathcal{L}_1 + \mathcal{L}_2)$, where

$$\mathcal{L}_1 = \sum_{i=1}^{2} [x_i \ln y_i + (1 - x_i) \ln(1 - y_i)]$$

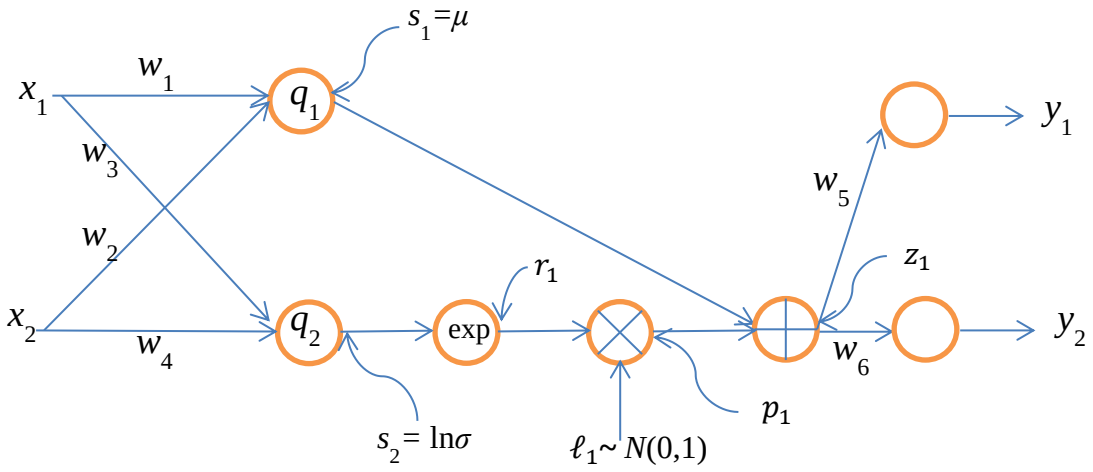
and

$$\mathcal{L}_2 = \frac{1}{2}\left(\mu^2 + \sigma^2 - \ln \sigma^2 - 1\right).$$

(a) Find $\frac{\partial \mathcal{L}_1}{\partial z_1}$.

(b) Find $\frac{\partial \mathcal{L}_1}{\partial w_4}$.

(c) Find $\frac{\partial \mathcal{L}_2}{\partial w_4}$.



2. Use the equations of optimal margin (linear) SVM (in pp. 12) to find $\mathbf{w}$ given $\mathbf{x}_1 = [1 \quad 1]^T \in C_{+1}$, $\mathbf{x}_2 = [-1 \quad 1]^T \in C_{-1}$, $\mathbf{x}_3 = [2 \quad 2]^T \in C_{+1}$, $\mathbf{x}_4 = [-2 \quad 2]^T \in C_{-1}$. Hint: Recall the meaning of support vectors.
3. Assuming that the following is a part of subpixel convolution networks with stride $\frac{1}{2}$. Compute the resultant values with the ReLU activation function. For simplicity, you may assume that the resultant plane has a size of 9×9.

Input plane

6	0	-4	0	1
4	4	0	2	1
3	-7	1	4	2
-2	2	1	-4	2
5	1	2	4	-1

Kernel

3	-1	2
-2	1	-3
-2	0	3

4. In the version 1 model of the LSTM, the activation function for f_t is sigmoid (represented by σ in pp. 21 of the notes). If we use ReLU instead, can the LSTM nodes still properly work? Why or why not?
5. In the lecture of Adaboost, we mentioned that we use three weak classifiers with a voting to correctly classify all samples (pp. 5). As the dataset is linearly non-separable, this problem cannot be correctly classified with a linear classifier, nor with the linear combination of some linear classifiers. Where did we introduce the nonlinear function when making decisions?