HW #6 Due: 6/19/2018

1. We mentioned that the parameters γ and β in batch normalization are trained through backprop. Explain why we don't want to directly compute these two parameters from the training samples and also give the details how the training is carried out.
2. We use the "Reparameterization Trick" in training Variational AutoEncoder (VAE). Explain why this trick is necessary and how it is accomplished.
3. Use the equations of optimal margin (linear) SVM (in pp. 12) to find $\mathbf{w}$ given $\mathbf{x}_1 = [1 \quad -1]^T \in C_{+1}$ and $\mathbf{x}_2 = [-1 \quad -1]^T \in C_{-1}$.
4. Implement a discrete HMM training program. Use the three-urn example (in pp. 12 of the PPT file) to test your program and produce the training results after 100 iterations. Use the red and blue balls in each urn to compute the initial emission probability. The initial transition probability $\mathrm{A} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ and $\pi = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$.

5. 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.

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