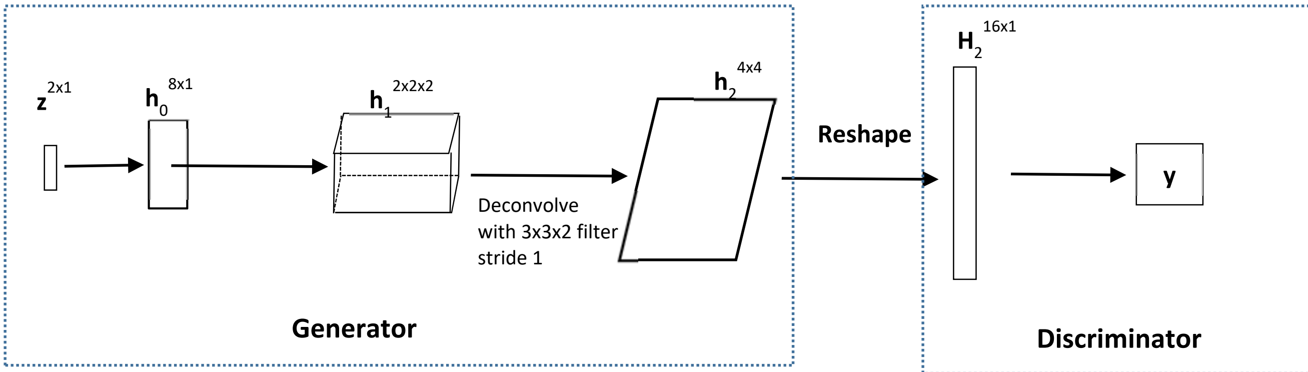
Assignment 6, due 4pm , April 12th

1. **Problem 1** Forward and backward propagation for a simple GAN below [20 points]



Generator:

Random input: z = $=\begin{pmatrix}0.4\\0.8\end{pmatrix}$

Project: $h_0 = \mathrm{ReLu}(w_0^t z_1 + b_0), w_0 = \begin{bmatrix}0.1 & 0.3\\0.7 & 0.9\\0.5 & 0.7\\0.1 & 0.9\\0.3 & 0.1\\0.4 & 0.6\\0.7 & 0.8\\0.2 & 0.1\end{bmatrix}, b_0 = \begin{bmatrix}0.1\\0.4\\0.2\\0.3\\0.5\\0.9\\0.7\\0.9\end{bmatrix}$

h_1=reshape(h_0)

$$h_2 = \mathrm{ReLu}(Deconv(h_1, w_1, b_1) \in R^{4x4}, w_1^{3x3x2} = [w_1^1\ w_1^2],$$

Deconvolution: $w_1^1 \in \begin{bmatrix}0.1 & 0.4 & 0.5\\0.1 & 0.7 & 0.1\\0.2 & 0.4 & 0.3\end{bmatrix}, w_1^2 \in \begin{bmatrix}0.2 & 0.1 & 0.6\\0.3 & 0.9 & 0.2\\0.5 & 0.2 & 0.3\end{bmatrix} \quad b_1 = \begin{bmatrix}0.1 & 0.5 & 0.6 & 0.3\\0.2 & 0.4 & 0.8 & 0.1\\0.1 & 0.3 & 0.7 & 0.2\\0.3 & 0.1 & 0.2 & 0.1\end{bmatrix}$

Binary Discriminator:

$$\mathbf{H}_2 = \mathrm{Flatten}(\mathrm{h}_2) \in R^{16x1}$$

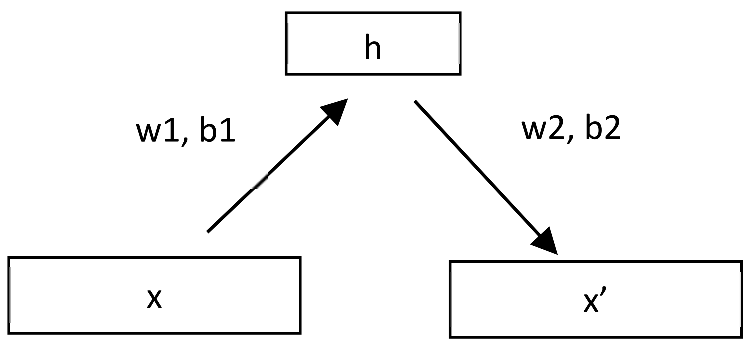
$p(\mathbf{y}=0)=\sigma(w_2^t \mathbf{H}_2 + b_2), w_2 = [0.1\ 0.2\ \ 0.3\ \ 0.2\ 0.4\ \ 0.6\ \ 0.7\ \ 0.1\ \ 0.1\ \ 0.2\ \ 0.5\ \ 0.3\ \ 0.1\ 0.2\ \ 0.6\ \ 0.9]^t\ b_2 = 0.3$

Tasks:

1) Using the forward propagation equations and given the initial weights, produce the values for h_0, h_1, and h_2 from a given z.
2) Given the logistic regression discriminator classifier parameters $\theta^D = \{w_2, b_2\}$ above, derive the gradients for w_1 and w_0 by maximizing the $\log p(y=0)$. Show your process.

Problem 2 Auto-encoder [20 points, only for those taking the class at 6000 level]

For the encoder and decoder below,



$$h = \mathrm{Re}\,Lu(w_1^t x + b_1)$$
$$x^{'} = \sigma(w_2^t x + b_2)$$
$$w^1, w^2 = \arg\min_{w^1, w^2} \sum_{i=1}^{N} (x_i - x_1^{'})^2$$

Derive the gradient equations for w_1 and w_2 by minimizing the reconstruction errors above

3. **Problem 3** Variational Auto-encoder (VAE) [20 points, **extra credit**]

For an VAE, prove maximizing the decoder distribution p(x|ϕ) is equivalent to maximizing the following function, where q(z|x,θ) is the encoder function and KL the KL divergence. Hint: using Jensen's inequality

$$E_{q(z|x,\theta)}(p(x|z,\phi)) - KL(q(z|x,\theta)\,||\,p(z))$$