Problem 1:

In this problem we have the intractable integral $p_{theta}(x) = \int p_{theta}(z)*p_{theta}(x|z)\, dz$. We can infer $p_{theta}(z)$ from $p_{theta}(z|x)$, but we do not know this either. So we are going to use Q(z|x) from $\log(p_{theta}(x)) = E_{qsigma}(\log p_{theta}(x))$ to infer P(z|x) to derive a loss function L(x, theta, sigma). To start with we will use KL divergence to get

$D_{KL}[Q(z|x)||P(z|x)] = E_{qsigma}[\log(Q(z|x) / P(z|x)] = E_{qsigma}[\log(Q(z|x) – \log(P(z|x)]$.

From here Bayes rules can be used to further modify the equation.

$D_{KL}[Q(z|x)||P(z|x)] = E_{qsigma}[\log(Q(z|x) – \log(P(x|z)-\log(z)] + \log(P(x))$

$D_{KL}[Q(z|x)||P(z|x)] - \log(P(x)) = E_{qsigma}[\log(Q(z|x) – \log(P(x|z)-\log(P(z))]$

This equation can then be rearranged and multiplied by -1 to get

$\log(P(x)) - D_{KL}[Q(z|x)||P(z|x)] = E_{qsigma}[\log(P(x|z)- (\log(Q(z|x) - \log(P(z))]$

$\log(P(x)) - D_{KL}[Q(z|x)||P(z|x)] = E_{qsigma}[\log(P(x|z)]- E_{qsigma} [(\log(Q(z|x) - \log(P(z))]$

$\log(P(x)) - D_{KL}[Q(z|x)||P(z|x)] = E_{qsigma}[\log(P(x|z)]- D_{KL}[Q(z|x)||P(z)]$

From here the loss function can be written as

$L(x, theta, sigma) = E_{qsigma}[\log(P(x|z)]- D_{KL}[Q(z|x)||P(z)] + D_{KL}[Q(z|x)||P(z|x)]$

It is important to note that P(z|x) is an unknown variable and thus needs to be dropped from the equation. So, we get the following for the loss

$L(x, theta, sigma) = E_{qsigma}[\log(P(x|z)]- D_{KL}[Q(z|x)||P(z)]$

For the full symbol format $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}^{(i)}) = -D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})||p_{\boldsymbol{\theta}}(\mathbf{z})) + \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z})\right]$

Question 2:

The LSGAN equation can be rewritten as

$V(D) = \int 0.5(p_{data}(x)(D(x)-b)^2 + P_{model}(x)(D(x)-a)^2)dx$

Taking the derivative equal to zero gives

$V(D)^* = 0 = P_{data}(D(x)-b) + P_{model}(D(x)-a)$

Rearranging

$P_{data}(D(x)) + P_{model}D(x) = P_{data}*b + P_{model}*a$

$D^*_G(x) = (P_{data} *b + P_{model} *a)/ (P_{data} + P_{model})$

Two major differences stand out between the optimal value of D for LSGAN and GAN. The first is that the top part of the difference now contain a Pmodel factor for LSGAN that is not present in the GAN version. The second difference is that the value in the top of the difference in LSGAN are multiplied by a constant value, b and a.