Problem 1:

We want to estimate $\log(p_{theta}(x)) = E_{qsigma}[\log p_{theta}(x))$ where we are using Q(z|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 a unknown variable and thus needs to be dropped from the equation. So we get the following for the loss

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

Question 2:

The LSGAN equation can be rewritten as

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

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

Rearranging

$P_{data}(D(x)) + P_zD(x) = P_{data}b + P_za$

$D(x) = (P_{data}b + P_za)/(P_{data} + P_z)$

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 Pg 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.