1. NA
2. NA
3. a) NA

3. b)

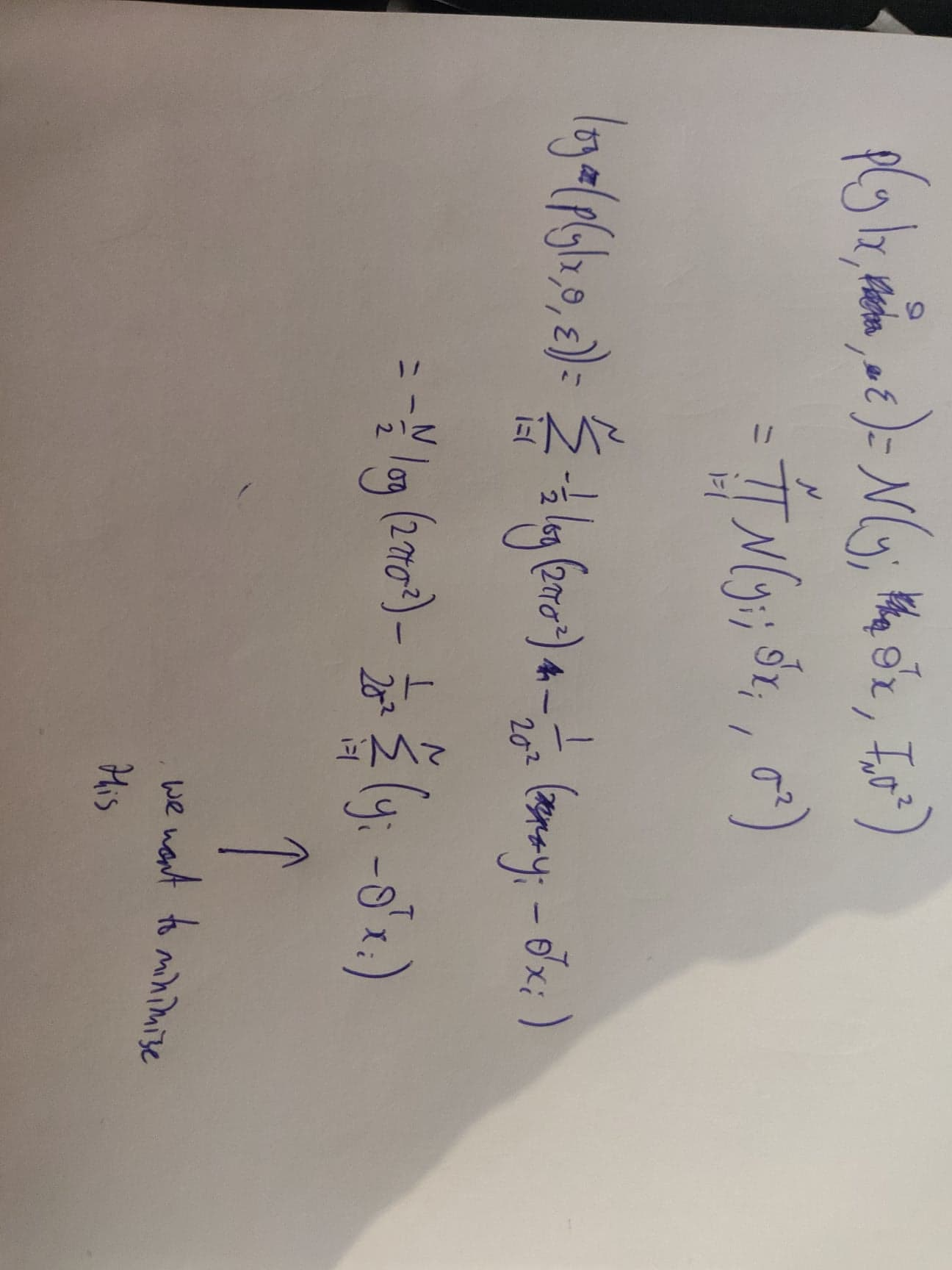
i)

-write the likelihood

-take log

-see that optimizing w.r..t. theta is minimizing square error between y\_i and \theta^T x\_I

--------------------------------------------------------------------------------------------------------------------



Idk what to do after this though

--------------------------------------------------------------------------------------------------------------------------

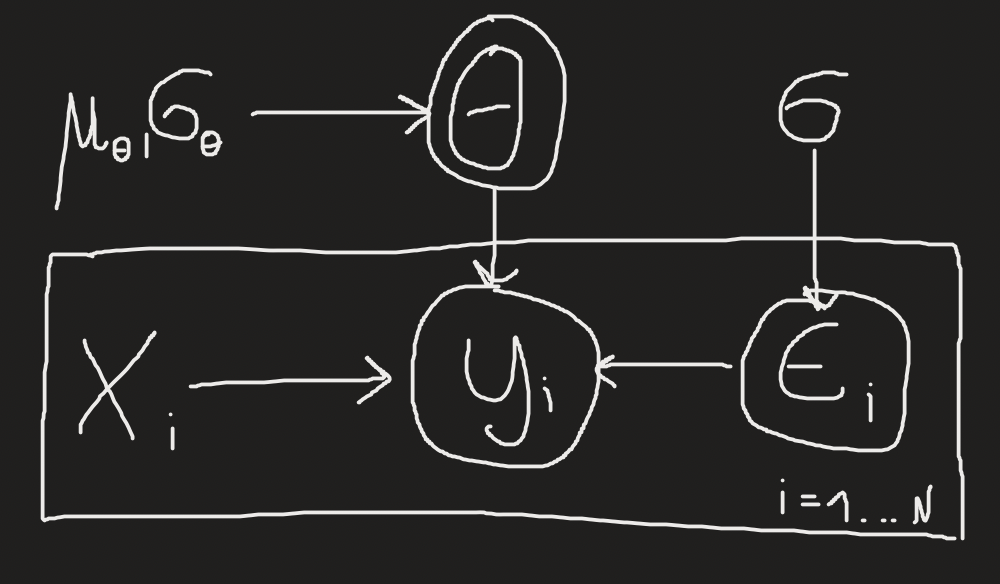
ii) overfitting happens when the model is too flexible and instead of learning the pattern, the system just memorizes the observed data. Overfitting is bad because it harms generalization.

iii) MAP takes the Prior into account. We can reduce overfitting by having lower prior probability of extreme parameters.

iv) I would use a Gaussian prior on theta. Then, prior and likelihood would be jointly Gaussian since y is obtained through linear transformation of theta (with some additional noise). Then the posterior probability of theta would also be gaussian since, the marginalization of Gaussian is also Gaussian.

(I’m not sure about this one – integrate theta from where (prediction of f\*?))

v)



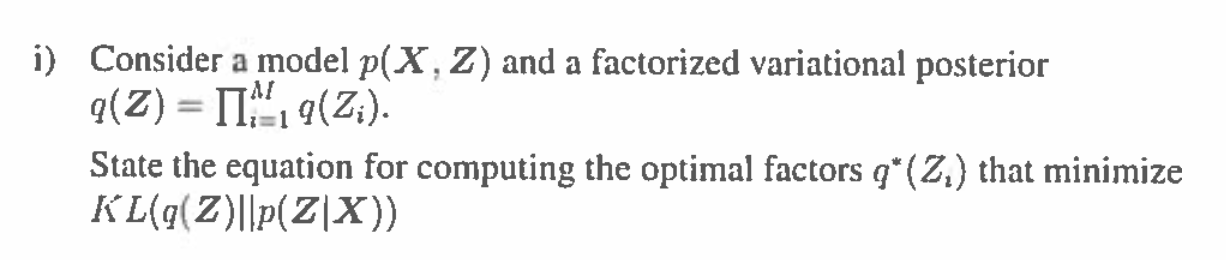
4.a)

i) detailed balance is a property of MCMC samplers where probability of being in state s1 and transitioning to s2, is the same as probability of being in the state s2 and transitioning to s1. :  
p(x)T(x’|x)=p(x’)T(x|x’)

ii) if detailed balance holds then one of the desired properties of MCMC is achieved i.e. if x is drawn from p, and then we sample x’ according to T(x’|x) marginal distribution of x’ is p as well. T leaves p invariant.  
It also ensures that the associated markov chain is reversible.

iii) NA (Gibbs Sampling)

b)



(1. Is this question about ELBO?)

(2. where are any variational parameters? )

Just gonna have a guess:

Log q(z) = log INTEGRAL q(z|x)q(x) dx

= log INTEGRAL q(z|x)q(x) p(x)/p(x) dx

= log INTEGRAL q(z|x)(q(x)/p(x)) p(x) dx

= log E\_p [q(z|x)q(x)/p(x)]

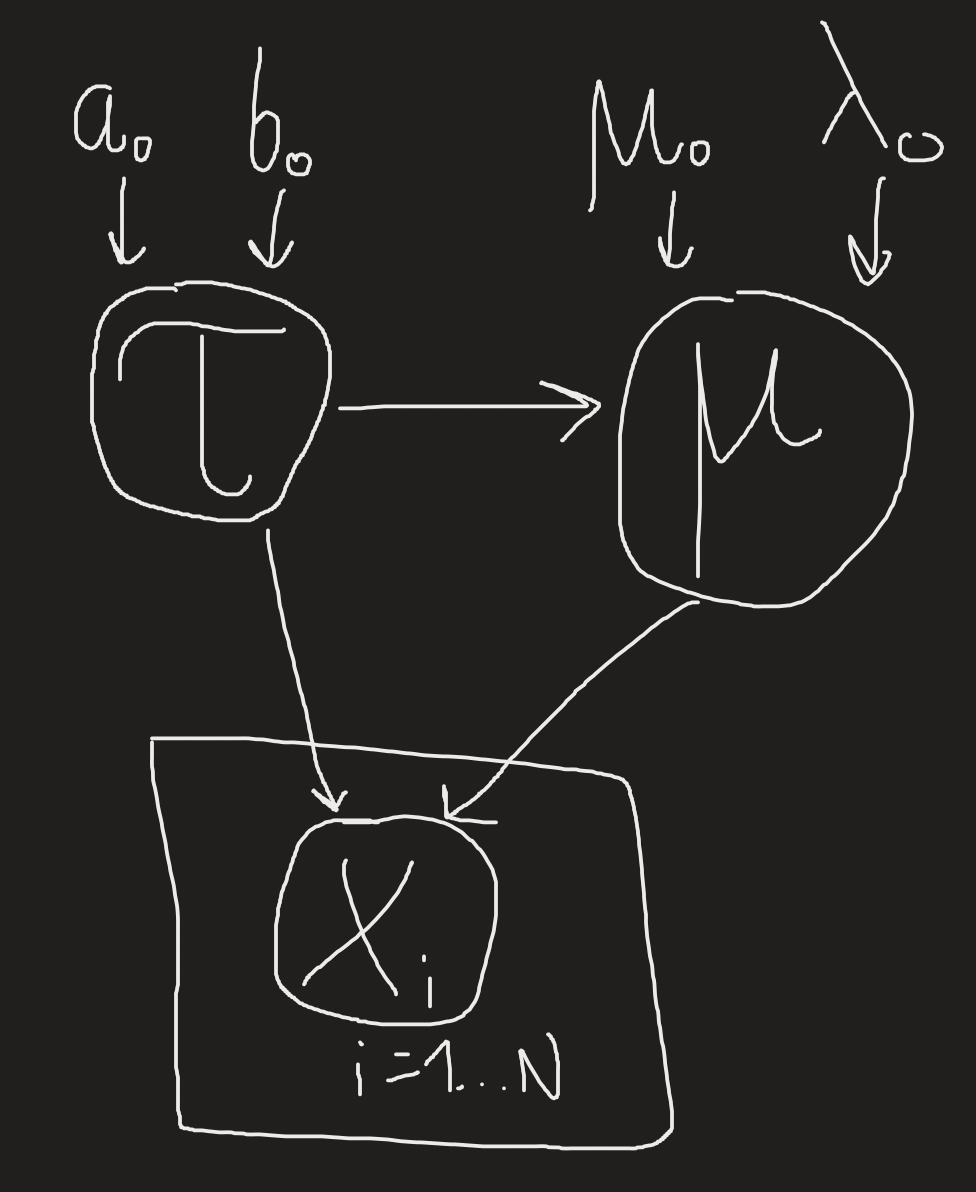
>= E\_p[ log q(z|x)q(x)/p(x) ]

= E\_p[ log q(z|x) ] - E\_p [log p(x)/q(x)]

Nvm it’s wrong idk

ii)

A)



B)

~~-write down log p(mu|X) as log of an integral,~~

~~-use Jensen inequality w.r.t p(tau) (or possibly q(tau) I'm not sure)~~

~~-do the math (ignore most of the things which do not depend on mu – just write const)~~

C)

log q is proportional to –½ times quadratic in mu, this is the same shape as gaussian. Since q(mu) is a pdf, it must integrate to 1, and thus it is Gaussian.

(not sure about this one)

D)

* write down as n -> infinty mu\_n = mean(x), lambda\_n->infinty

For Variational posterior of mu converges to the mean which is the maximum likelihood estimate of x. The precision goes to infinity <=> the variance goes to 0, thus as the number of points increases the MLE is more appropriate, while at N->infinity it is exact.

c)

-use Var(x)=E[x^2]-E[x]^2

-cancel 2 terms

-expand as integrals

-expand the square as multiplication of 2 terms

-match the integrals

