**Question 1 (1pt)**

**Given that the data is high-dimensional (in our case, images), what could be the motivation for training a latent variable model, as opposed to directly trying to model p(x)?**

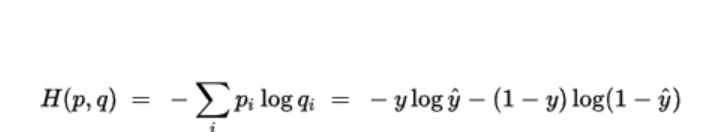
Link: <http://www.econ.upf.edu/~michael/latentvariables/lecture1.pdf>

* The advantage is that we won’t just learn the distribution of the input image X but the parameters that rule this distribution. That will allow us to create new images which follow a new distribution but which…..
* We will train the model in the latent space, which has a smaller dimensionality than the feature space, so less parameters needs to be learned.
* Learning the latent variable model provides more flexibility and generalization when creating new images, which slightly vary from the original.
* The latent variable model is much simpler than the feature variable model is much simpler to work with and we will get the same efficiency without reducing the flexibility of the model.
* A latent variable model expresses the uncertainty of the model, this is representing the effect of unobservable covariates, factors and inner structure of our data.
* Latent variable models summarizes different measurements of the same (directly) unobservable characteristics.

**Question 2 (1pt)**

**Write out log p(x|z) for this discrete model, and simplify the expression as much as possible. Can you relate this expression to a commonly used loss function for neural networks?**

BINARY!!!! Cross- Entropy loss:



**Question 3 (1pt)**

**Write out log p(x|z) for this continuous model, and simplify the expression as much as possible. Can you relate this expression to a commonly used loss function for neural networks? (Hint: note that terms that are constant w.r.t. the learned parameters will not affect the learning, as their derivative will be zero.)**

MSE loss: