Midterm

1. Summarize how the change of variable theorem is used to estimate the log-density in invertible neural networks. In other words, describe the mechanism behind the following equation.

With an invertible F: $$F(x) = z \implies \log p(x) = \log p(z) + \log \left| \det \frac{\partial F}{\partial x} \right|$$

There are several types of generative models, and the typical generative models that we think of are GAN and VAE. However, there is also another model that is flow-based model. The main difference of the flow-based model compared to the other two models is that the flow-based model encodes and decodes the latent vector through inverses. Using inverse has some advantages, since there is no reduction in dimension and expansion during the encoding and decoding phase.

To understand the change of variable theorem, let us first consider x as the input vector and z as the latent vector. $x = f(z)$, and if f is invertible, $z = f^{-1}(x)$. Let us also consider $p(x)$ as and $p(z)$ to be the probability distribution function of x and z. Then, by the definition of the change of variable theorem, $\int p(x)dx = \int p(z)\, dz = 1$, and if we take the derivative on both sides, we'll get $p(x) = p(z)\left|\frac{dz}{dx}\right| = p\left(f^{-1}(x)\right)\left|\frac{df^{-1}}{dz}\right|$, and using multi variable, we can represent it using Jacobian Determinant, thus $p(x) = p(z)|\det\frac{dz}{dx}|$. Then, if we take log on both sides, we finally get the following equation: $\log p(x) = \log p(z) + \log|\det\frac{dF}{dx}|$.

Doing this inverse can be very beneficial in computation, under some conditions. First of all, when computing determinant, if we have the form of $\begin{bmatrix} I_d & 0 \\ \frac{\partial y_{I_2}}{\partial x_{I_1}} & \frac{\partial y_{I_2}}{\partial x_{I_2}} \end{bmatrix}$ the determinant of this form becomes extremely simple, or 'nice'.

2. Describe your research topic and explain how invertible neural networks can be used for your research topic. If you do not have a research topic yet, you can choose any random topic.

My research topic is figuring out cold-start categorization on documents that are collected in real world industries. Most of the industries these days have numerous documents flooding into their data center. Some of the documents may be system messages which can be directly categorized if pre-set, but some of the other documents may be human written notes, which can theoretically be categorized if one chooses the category before writing the memo, but this does not happen that often. What is worse is that most industries did not start off with strategies to categorize the data, so they do not have a fine-tuned categorization system. I aim to combine the documents with minimal human supervision, to come up with weak supervisions to 'generally' categorize the documents, then use NLP models like BERT to locate clusters within the documents, then finally combine them using the new and appropriate categories.

I wish to I could use invertible neural networks for my research topics, but since I'm using the pre-defined methods like BERT and SVDs, I do not have much flexibility in giving variations to the methods that I'll be using. Also, for the input to be invertible, the input dimension must be square-sized (n * n) matrix. However, for NLP models, especially for BERT, the inputs rarely become square-sized matrix. I found a paper, 'Two Birds with One Stone: Investigating Invertible Neural Networks for Inverse Problems in Morphology' by Gozde Gul Sahin and Iryna Gurevych, which attempts to use INN in evaluating morphology, but unfortunately it is not much related to my topic.