# Exploring Variational Autoencoders

Dami Akinniyi | July 2024

**Abstract**

*Like many data scientists, my initial introduction to machine learning revolved around learning packages and understanding which functions to call to train a model. As a master's student applying machine learning to projects with tight deadlines, I somewhat accepted the black-box approach to understanding these models. Over the years, my priorities have shifted, and I now find myself eager to dissect the black boxes of AI to grasp the underlying mechanisms of the many models integrating into our lives.*
*While older than many of the generative models available today, variational autoencoders [VAEs] have a particular charm. Whether it's their rigorous and detailed formulation or the wealth of information contained in related papers, VAEs and their subsequent families are as fascinating as they are intimidating.*
*My first foray into VAEs involved running and modifying a hand-me-down code. After several rounds of training, it became clear that I was out of my depth and uncertain about which changes led to improvements or setbacks. After setting it aside for a while, I've finally returned to look under the hood and understand the intricacies of VAEs.*

*An Introduction to Vairational Autoencoders by Diederik P. Kingma and Max Welling*
*Autoencoding Variational Bayes by Diederik P. Kingma and Max Welling*

---

# 1 Introduction to Variational Autoencoders

Variational autoencoders are .... $x^2 + y^2 = z^2$

# 2 Understanding Variational Inference

What is this Write about concepts here. Starting a new paragraph here. Loern ipsum. What is the understanding of this topic. I need to know what this is about and consolidate my knowledge and understanding of it. Without consistent support, it becomes rather pointless that the formatting and what is not. I dont know what to do here because this is not so great for me.
Why are these paragraphs indented? I think this is a rather silly behaviour of latex.

## 2.1 Posterior Probability and Intractability

What does this have to do with anything

# 3 Code Exploration

The loss function

- minimizing evidence of the data and why mse works for fashion mnist
- in the same manner, why categorical cross entropy works for mnist digit
- Understanding expected behaviour of kl-divergence

What I have written in code and my understanding of it.