**2016- A note on the evaluation of generative models**

* Generative models have man different uses like compression, denoising, semi- supervised learning,etc.
* This article involves properties relating to evaluation and interpretation of generative models (on image models).
* Three mostly used criteria - average log-likelihood, Parzen window estimates, and visual fidelity of samples
* This shows that above three methods are independent of each other when the data is high-dimensional (meaning good performance w.r.t one criterion doesn’t mean other two will show good performance too).
* This also show why Parzen window estimates are generally avoided.

Log-likelihood: -

* For density estimation, this is used for training generative models
* Measure for quantifying generative image modelling performance
* For semi supervised learning

We then discuss the relationship between log-likelihood, classification performance,

visual fidelity of samples and Parzen window estimates. (all mathematics calculation)

**Conclusion**: - Different metrics can lead to different trade-offs, and different evaluations favor different models.