Joint Data Modeling using Variational Autoencoders

Achint Kumar

Duke University

January 23, 2023

Desiderata

Desiderata

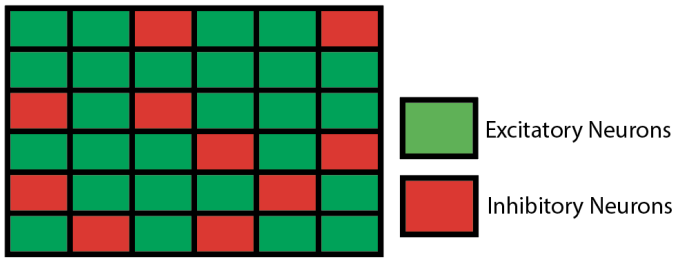
Introduction: Simplicity begets Complexity

Why complicated physical systems are often governed by small number of parameters?

1. Renormalization Group View: Coarse-graining in length or time scale eliminates many parameters.
2. Complex Systems View: Simple (nonlinear) rules generically give rise to complicated behaviour
3. Manifold Hypothesis View: Real world data are highly structured and correlated so they lie on low dimensional manifold

Wilson-Cowan Equation

$$\tau \frac{dr(\vec{x},t)}{dt} = -r(\vec{x},t) + [w * \phi(r)](\vec{x},t)$$



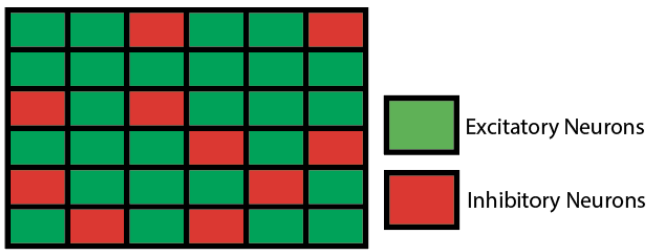
Introduction: Simplicity begets Complexity

Why complicated physical systems are often governed by small number of parameters?

1. Renormalization Group View: Coarse-graining in length or time scale eliminates many parameters.
2. Complex Systems View: Simple (nonlinear) rules generically give rise to complicated behaviour
3. Manifold Hypothesis View: Real world data are highly structured and correlated so they lie on low dimensional manifold

Wilson-Cowan Equation

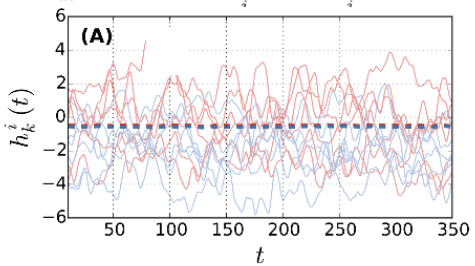
$$\tau \frac{dr(\vec{x},t)}{dt} = -r(\vec{x},t) + [w * \phi(r)](\vec{x},t)$$



Firing Rate Equation

$$\tau_E \frac{E_i(t)}{dt} = -E(t) + \phi(\sum_j J_{ij}^{EE} E_j - \sum_j J_{ij}^{EI} I_j)$$

$$\tau_I \frac{I_i(t)}{dt} = -I(t) + \phi(-\sum_i J_{ij}^{II} I_j + \sum_i J_{ij}^{IE} E_j)$$



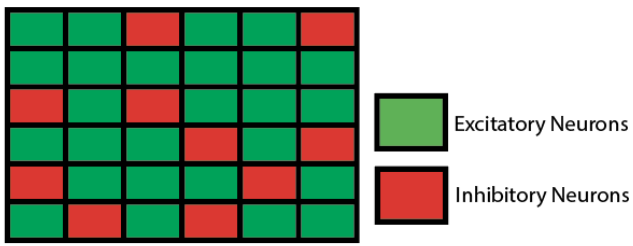
Introduction: Simplicity begets Complexity

Why complicated physical systems are often governed by small number of parameters?

1. Renormalization Group View: Coarse-graining in length or time scale eliminates many parameters.
2. Complex Systems View: Simple (nonlinear) rules generically give rise to complicated behaviour
3. Manifold Hypothesis View: Real world data are highly structured and correlated so they lie on low dimensional manifold

Wilson-Cowan Equation

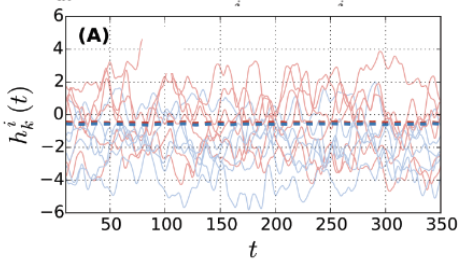
$$\tau \frac{dr(\vec{x},t)}{dt} = -r(\vec{x},t) + [w * \phi(r)](\vec{x},t)$$



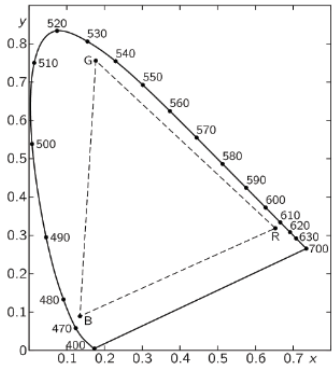
Firing Rate Equation

$$\tau_E \frac{E_i(t)}{dt} = -E(t) + \phi(\sum_j J_{ij}^{EE} E_j - \sum_j J_{ij}^{EI} I_j)$$

$$\tau_I \frac{I_i(t)}{dt} = -I(t) + \phi(-\sum_i J_{ij}^{II} I_j + \sum_i J_{ij}^{IE} E_j)$$



Color Space is 2 dimensional



Main Offering of This Talk

How to infer underlying latent variables of high-dimensional data?

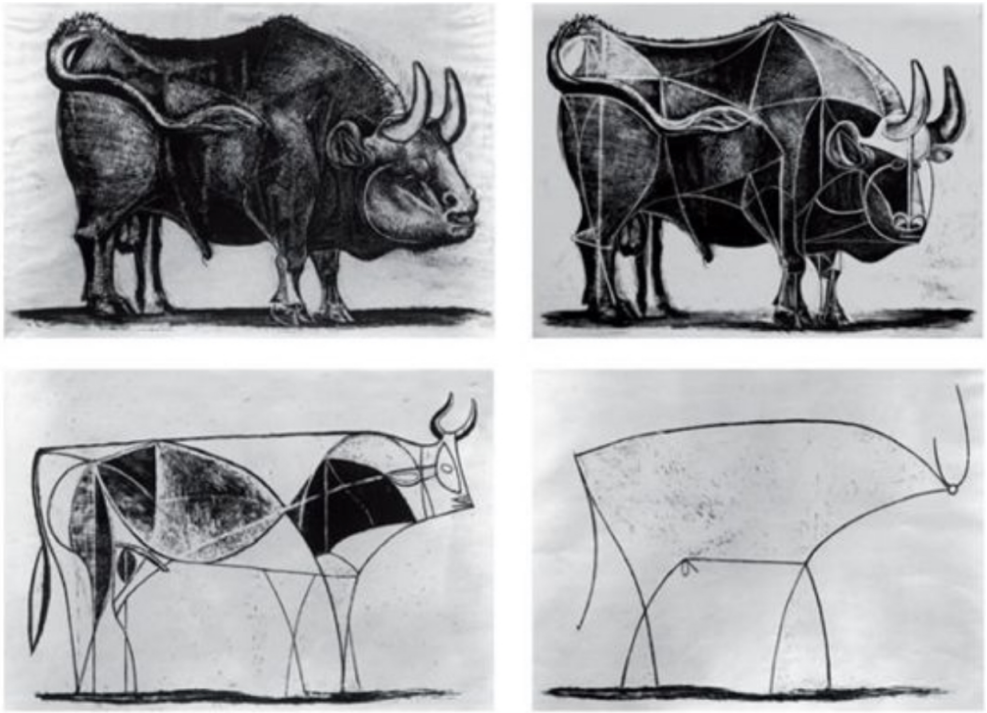
1. Project 1: Latent variable modelling of unimodal data (applications project)
2. Project 2: Latent variable modelling of multimodal data (methods project)

Main Offering of This Talk

How to infer underlying latent variables of high-dimensional data?

1. Project 1: Latent variable modelling of unimodal data (applications project)
2. Project 2: Latent variable modelling of multimodal data (methods project)

Essence of Latent Variable Modelling



The Bull by Picasso (1945)

- Six strokes capture the essence of the bull (latent variables)
- Six strokes is scaffolding for any bull (generative modelling)

Why should we care?

- Latent variable models: Working in latent space is simpler than working in data space.

  "Make things as simple as possible, but not simpler"

  Albert Einstein

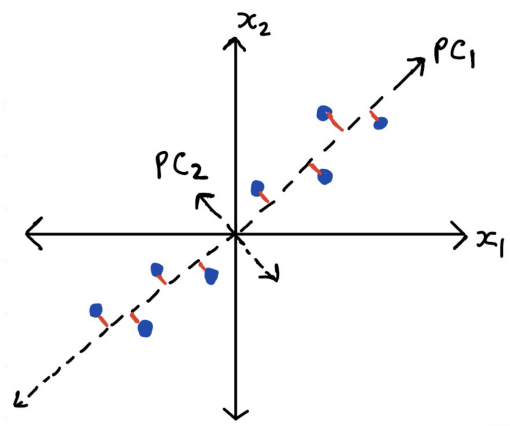
- Generative models: We can generate new data. Abstractly, it lets us represent and sample from high dimensional data distribution, p(x) efficiently.

  "What I cannot create, I do not understand"

  Richard Feynman

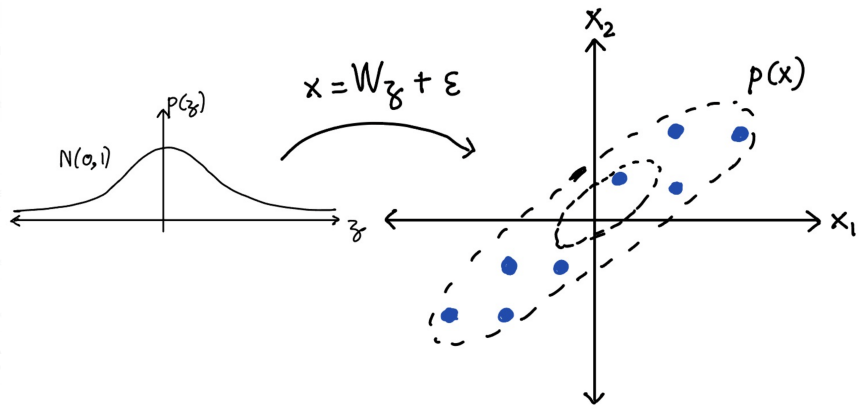
Latent Variable Modelling: Historical Perspective

Principal Component Analysis (Pearson, 1901. Hotelling, 1933)



- Maximize variance in projected space.
- Projected space has less noise, no redundancy
- Linear, not generative, needs covariance matrix diagonalization

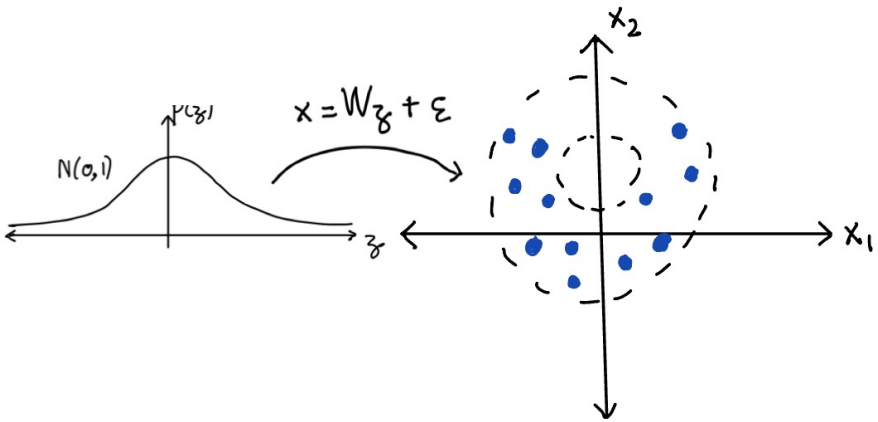
Probabilistic PCA (Tipping & Bishop, 1999)



- Maximize $\log p(x)$ wrt W
- Both latent variable and generative model
- Linear, needs covariance matrix diagonalization

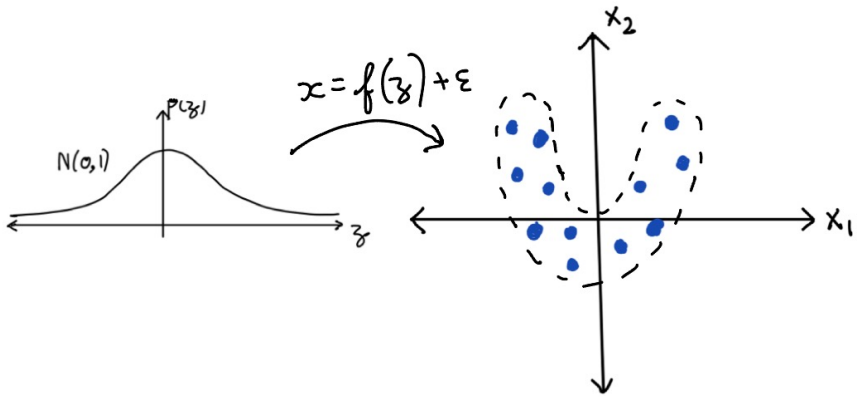
pPCA to Variational Autoencoders

Probabilistic PCA (Tipping & Bishop, 1999)



- Maximize $\log p(x)$
- Both latent variable and generative model
- Linear, needs covariance matrix diagonalization

Variational Autoencoders (Kingma & Welling, 2013)



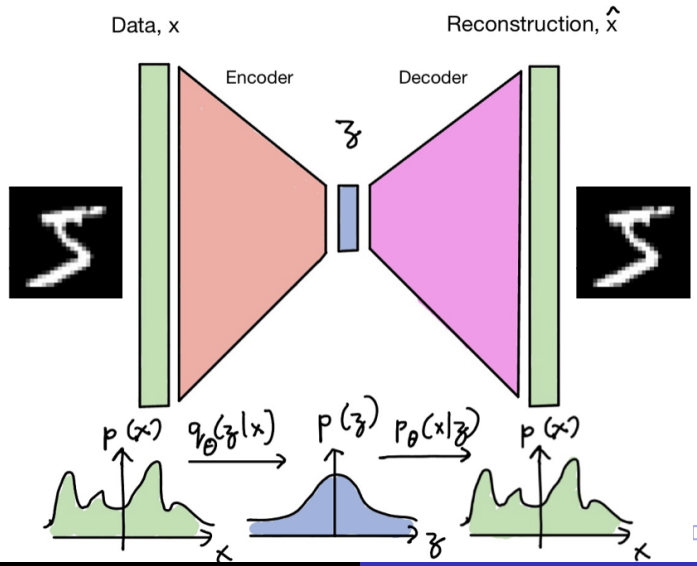
- Minimize variational free energy
- Non-linear, scalable and works for high dimensional data
- Fuzzy generation, non-interpretable latent space

Variational Autoencoders: Network Architecture

State-of-the-art framework for latent variable modelling. It consists of 2 neural networks:

1. Encoder: Data, **x** are input and its latent representation, **z** is output.
2. Decoder: Latent representation, **z** is input and reconstruction of data, $\hat{\mathbf{x}}$ are output.

To create probabilistic framework, we add noise to latent space.



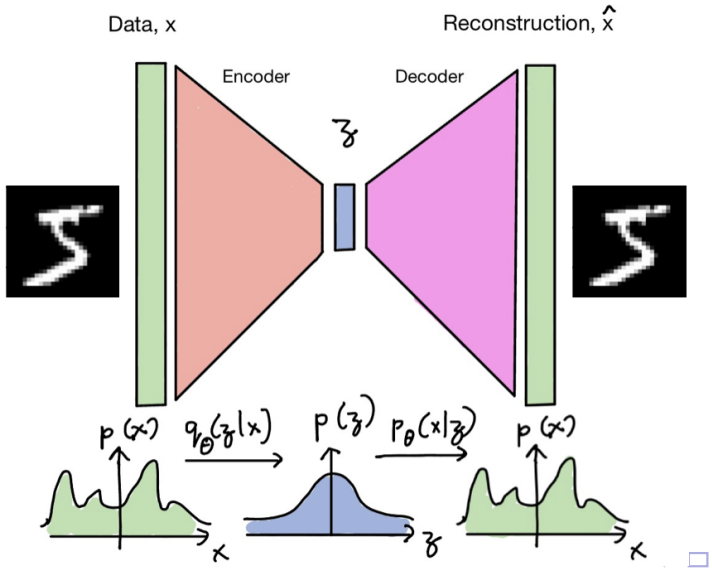
Variational Autoencoders: Training Objective

Want: $\max_\theta \log p_\theta(x) = \max_\theta \log \int p_\theta(x|z)p(z)\,dz$.

For Vanilla VAE, we assume

$$p(z) \sim \mathcal{N}(0, \mathbb{I})$$
$$p(x|z) \sim \mathcal{N}(f(z), \mathbb{I})$$

Problem: Calculating $p_\theta(x)$ leads to an intractable integral!
Solution: Exploit connection to statistical physics.



Training Objective: Statistical Physics to rescue

In Statistical Physics, partition function (ML: $p_\theta(x)$) is related to free energy.

$$-\log[p_\theta(x)] \doteq \underbrace{\mathcal{F}(x)}_{\text{Free Energy}} = \underbrace{\langle E(x) \rangle}_{\text{Energy}} - T \underbrace{\langle S(x) \rangle}_{\text{Entropy}}$$

where

$$\langle E(x) \rangle = -\mathbb{E}_{p(z|x)}[\log(p(x,z))]$$
$$\langle S(x) \rangle = -\mathbb{E}_{p(z|x)}[\log(p(z|x))]$$

This relation gives another way of calculating, $p_\theta(x)$.

Problem: We don't know the posterior $p(z|x)$.

Solution: Variational Inference

Training Objective: Statistical Physics to rescue

In Statistical Physics, partition function (ML: $p_\theta(x)$) is related to free energy.

$$-\log[p_\theta(x)] \doteq \underbrace{\mathcal{F}(x)}_{\text{Free Energy}} = \underbrace{\langle E(x) \rangle}_{\text{Energy}} - T \underbrace{\langle S(x) \rangle}_{\text{Entropy}}$$

where

$$\langle E(x) \rangle = -\mathbb{E}_{p(z|x)}[\log(p(x,z))]$$
$$\langle S(x) \rangle = -\mathbb{E}_{p(z|x)}[\log(p(z|x))]$$

This relation gives another way of calculating, $p_\theta(x)$.
Problem: We don't know the posterior $p(z|x)$.
Solution: Variational Inference

Training Objective: Variational Inference

Approximate true posterior, $p_\phi(z|x)$ by Gaussian distribution, $q_\phi(z|x) \sim \mathcal{N}(\mu(x), \sigma^2(x))$.

Variational free energy upper bounds true free energy,

$$\underbrace{\mathcal{F}_{\text{true}}(x)}_{\text{Free Energy}} \leq \mathcal{F}_{\text{var}}(x) = \underbrace{E_{\text{var}}(x)}_{\text{Energy}} - T\underbrace{S_{\text{var}}(x)}_{\text{Entropy}}$$

After making substitutions, variational free energy becomes

$$\mathcal{F}_{\text{var}}(x) \propto \underbrace{||x - \hat{x}||_2^2}_{\text{reconstruction}} + \underbrace{D_{KL}(q_\phi(z|x)||p(z))}_{\text{regularizer}}.$$

VAEs minimize $\mathcal{F}_{\text{var}}(x)$.

Training Objective: Variational Inference

Approximate true posterior, $p_\phi(z|x)$ by Gaussian distribution, $q_\phi(z|x) \sim \mathcal{N}(\mu(x), \sigma^2(x))$.

Variational free energy upper bounds true free energy,

$$\underbrace{\mathcal{F}_{\text{true}}(x)}_{\text{Free Energy}} \leq \mathcal{F}_{\text{var}}(x) = \underbrace{E_{\text{var}}(x)}_{\text{Energy}} - T\underbrace{S_{\text{var}}(x)}_{\text{Entropy}}$$

After making substitutions, variational free energy becomes

$$\mathcal{F}_{\text{var}}(x) \propto \underbrace{||x - \hat{x}||_2^2}_{\text{reconstruction}} + \underbrace{D_{KL}(q_\phi(z|x)||p(z))}_{\text{regularizer}}.$$

VAEs minimize $\mathcal{F}_{\text{var}}(x)$.

Story so far: Executive Summary

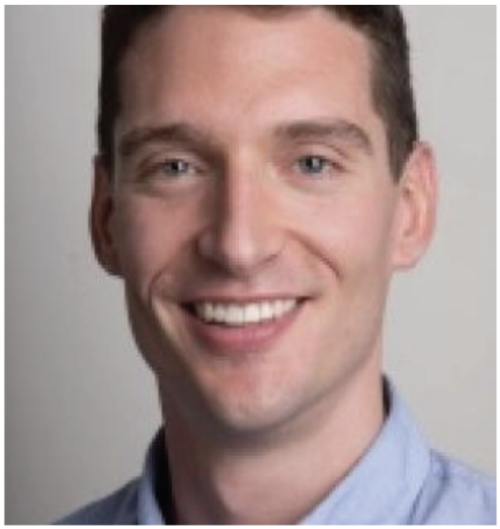
How to infer underlying latent variables of high-dimensional data?

Variational Autoencoders (VAEs) provide a framework for generative and latent variable modelling.
Training VAEs involves 3 steps:

1. Assume a form for prior distribution, $p(z)$ and variational posterior distribution, $q_\phi(z|x)$
2. Calculate variational free energy, $\mathcal{F}_{\text{var}}(x) = \text{reconstruction} + \text{regularizer}$
3. Minimize $\mathcal{F}_{\text{var}}(x)$ by backpropagation

Desiderata

Collaborators



Tom Harmon

John Pearson

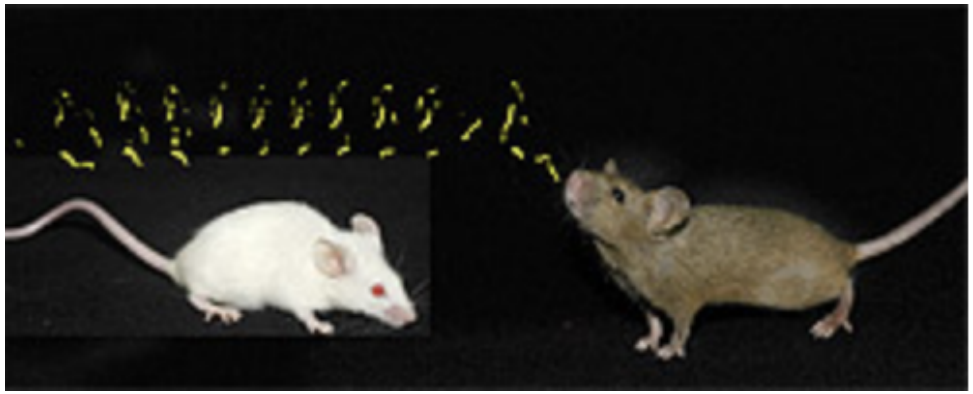
Richard Mooney

- Many thanks to Jack Goffinet for the code base. Also thanks to Miles Martinez for helpful discussions.

Background

Mice produce ultrasonic vocalizations (USVs) in frequency range between 30kHz- 110kHz under two circumstances:

- Isolation calls: Produced by pups when separated from litter
- Courtship calls: Produced by adult males in presence of female



Debate: Is there plasticity in mouse vocalization?

For the motion

- Mouse USVs share characteristics with songbirds like sequential vocalization with complicated and idiosyncratic structure [Holy & Guo, 2005]
- Based on a visual syllable classification scheme, its claimed hearing and deaf mice vocalize differently [Arriaga, et. al. 2012]

Against the motion

- Mouse USVs arise through innate processes and not through tutoring as in songbirds
- Based on a visual syllable classification scheme, its claimed hearing and deaf mice vocalize the same [Hammerschmidt, 2012. Mahrt et. al. 2013]

What's at stake?

If it turns out there is plasticity in mouse vocalization:

- Mice can serve as model organism for understanding vocal learning and disorders.
- Mice have a much more developed genetic manipulation toolset than vocal learners like songbirds.

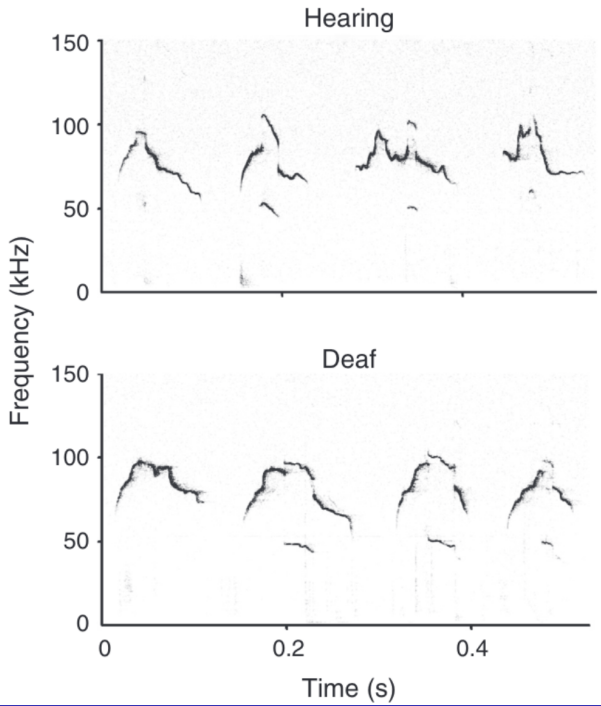
Problem Statement

Question

Are there any differences in vocalization produced by deaf mice (without auditory feedback) and hearing mice (with auditory feedback)?

My work improves on earlier experiments:

- I am using a more agnostic metric for comparing syllables.
- The deaf mice have a cleaner phenotype (only inner hair cells are affected) and are genetically very similar to hearing mice.



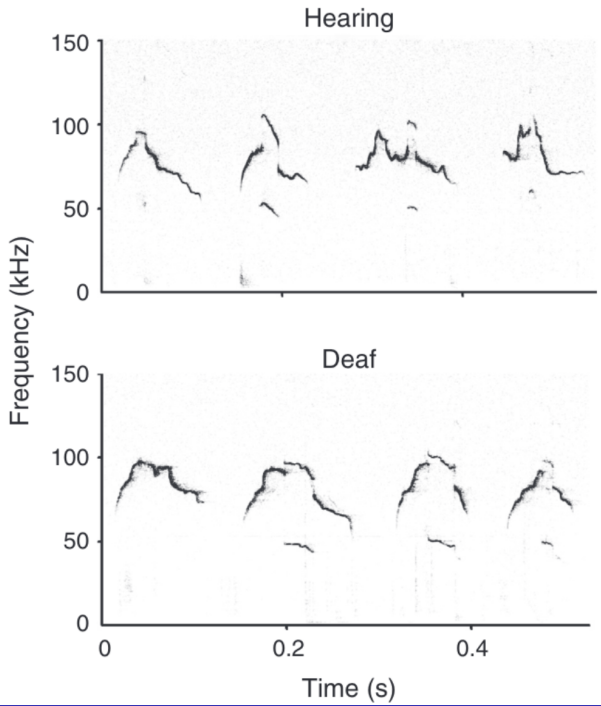
Problem Statement

Question

Are there any differences in vocalization produced by deaf mice (without auditory feedback) and hearing mice (with auditory feedback)?

My work improves on earlier experiments:

- I am using a more agnostic metric for comparing syllables.
- The deaf mice have a cleaner phenotype (only inner hair cells are affected) and are genetically very similar to hearing mice.

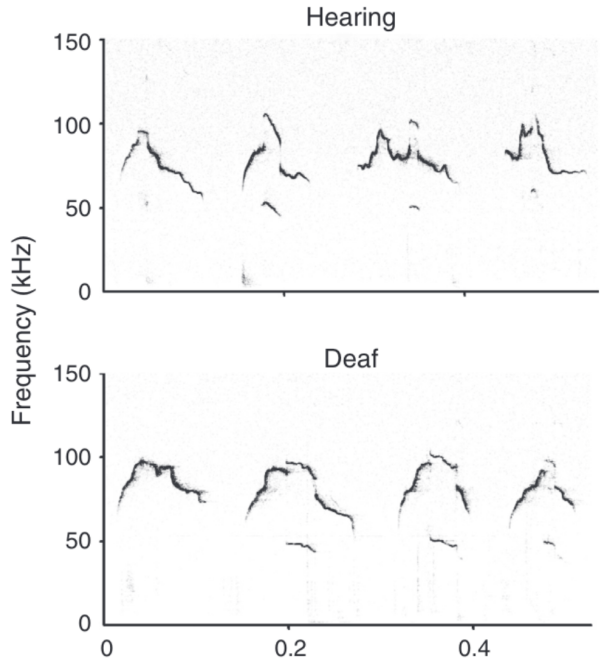


Why is this problem non-trivial?

- Both hearing and deaf mice syllables are high dimensional images.
- I am trying to answer whether two high dimensional images are generated from the same distribution or not.

My strategy consists of 2 steps:

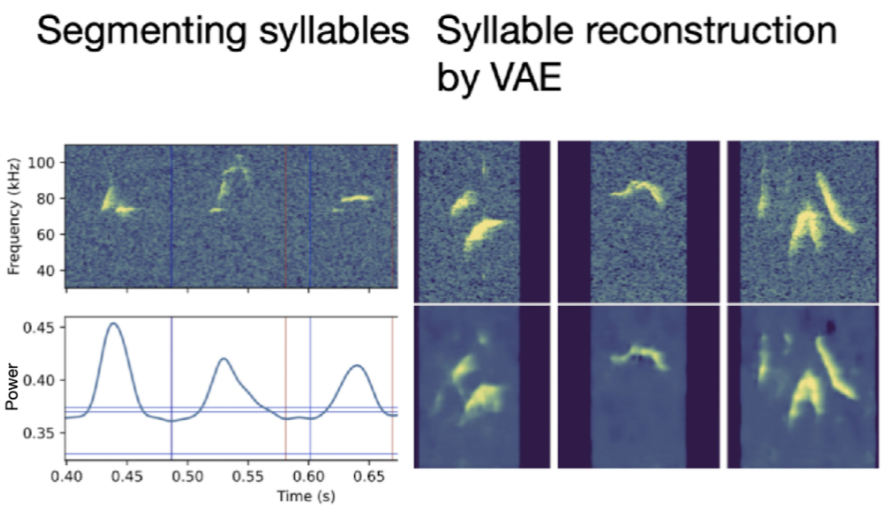
- Use VAE to reduce dimensionality of the distribution
- Use a statistical tool called MMD to compute similarity between latent distributions.

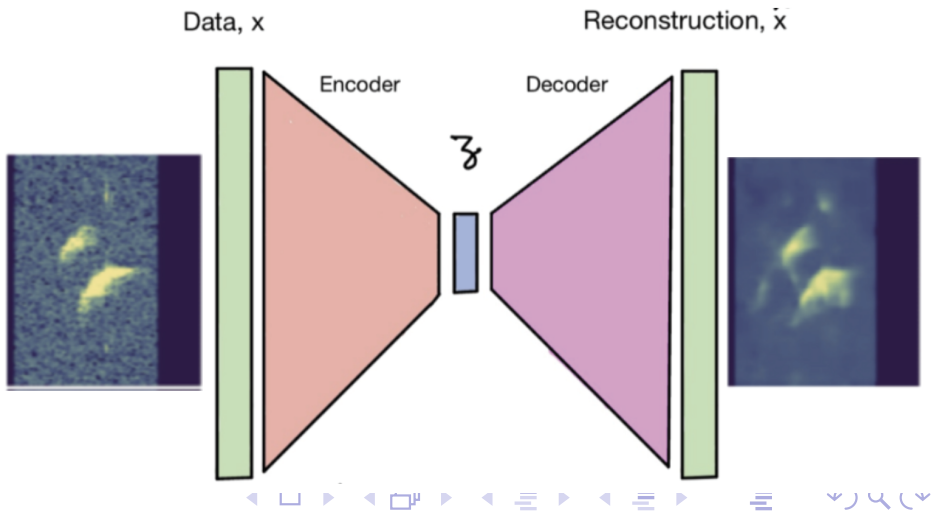


Training VAE

Model training involved following steps:

1. Create spectrograms of vocalizations of hearing and deaf mice
2. Segment the spectrogram into individual syllables by using amplitude thresholds
3. Feed the syllables into VAE

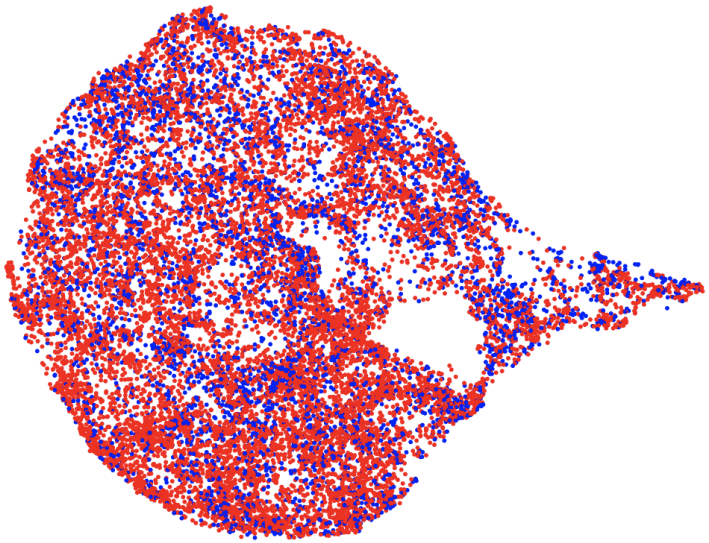




Visualizing Latent Space

Uniform Manifold Approximation and Projection (UMAP) is a non-linear dimensionality reduction technique which tries to preserve topology of data manifold.

- 32 dimensional latent representation was projected to 2 dimensions using UMAP for visualization.
- UMAP projection of hearing (blue) and deaf (red) mice are highly overlapping.



Comparing high dimensional data distributions

Are latent representation of syllables from hearing and deaf mice coming from different probability distributions?

$$X = \{x_1, \ldots x_n\} \sim p(x)$$
$$Y = \{y_1, \ldots y_n\} \sim q(y)$$

Maximum Mean Discrepancy (MMD) analysis is 'distance' metric to compare two probability distributions.

$$\mathrm{MMD}(p, q) = \max_{\mathcal{F}} ||\hat{\mu}_p(f(x)) - \hat{\mu}_q(f(y))||$$

where $f \in \mathcal{F}$ is are 'special' functions and $\hat{\mu}$ is the mean operation.

Comparing high dimensional data distributions

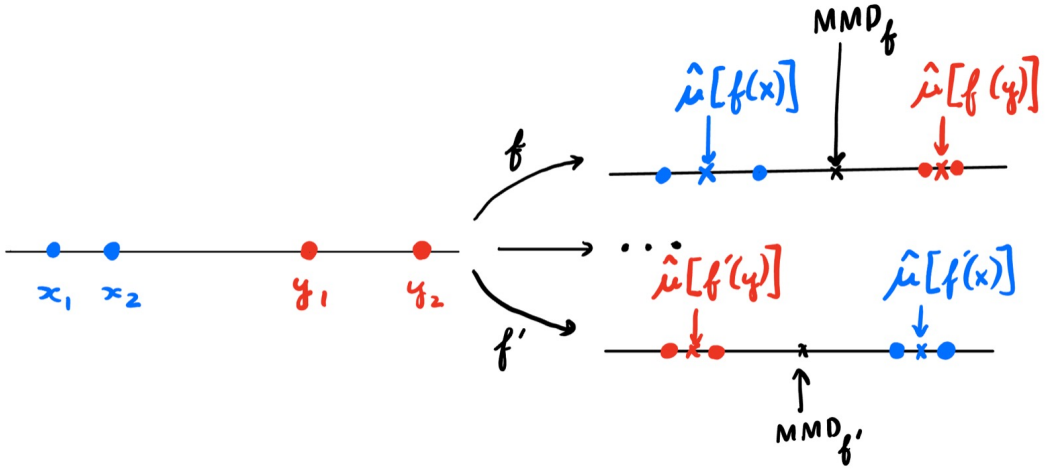
Are latent representation of syllables from hearing and deaf mice coming from different probability distributions?

$$X = \{x_1, \ldots x_n\} \sim p(x)$$
$$Y = \{y_1, \ldots y_n\} \sim q(y)$$

Maximum Mean Discrepancy (MMD) analysis is 'distance' metric to compare two probability distributions.

$$\mathrm{MMD}(p, q) = \max_{\mathcal{F}} ||\hat{\mu}_p(f(x)) - \hat{\mu}_q(f(y))||$$

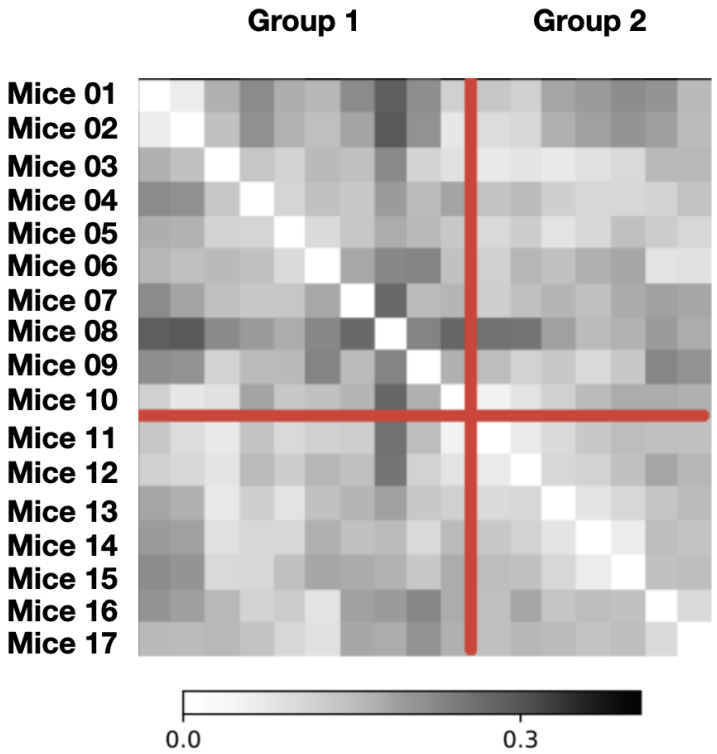
where $f \in \mathcal{F}$ is are 'special' functions and $\hat{\mu}$ is the mean operation.



Maximum Mean Discrepancy (MMD) analysis

- When $p = q$, then MMD value is close to zero
- When $p \neq q$, then MMD value is far from zero

MMD values are consistent with no difference in vocalization between the hearing and deaf mic



Conclusion

- I found no statistical difference in vocalizations produced by hearing and deaf mice.
- My results provides support for 'against the motion' side of the debate on whether there is vocal learning in mice.
- This project illustrates the power of latent variable modelling to answer scientific questions.

Desiderata

Collaborators

Ziyi Gong

John Pearson

- Many thanks to Miles Martinez and Daniela Albuquerque for helpful discussions as well.

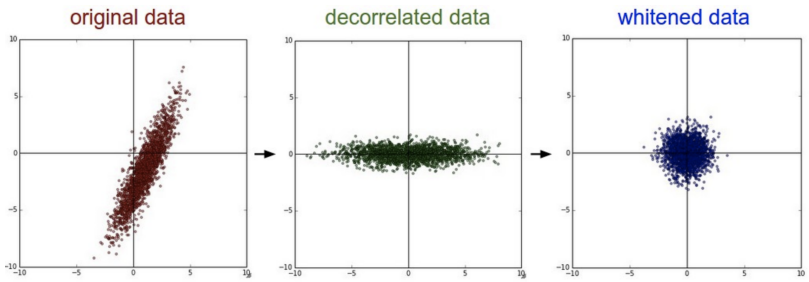
Introduction

Modern experiments are often multi-modal and it is of interest to jointly model them. Some examples:

1. Multi-messenger astronomy: Data of celestial objects simultaneously collected from neutrinos, electromagnetic radiation and gravitational waves
2. Neuroscience: Behaving animal and corresponding brain activity
3. Engineering: Cameras, lidar jointly model environment in self-driving cars

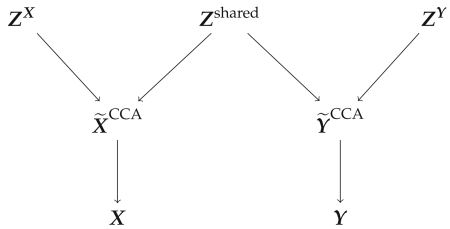
Multi-modal data analysis: Classical Methods

Canonical Correlation Analysis (CCA) (Hotelling, 1936)



- Let X (behavior) and Y (brain activity) be 2 datasets.
- Whiten X to $\tilde{X}^{CCA}$ and Y to $\tilde{Y}^{CCA}$ such that: $\text{corr}(\tilde{X}^{CCA}, \tilde{Y}^{CCA}) = \text{diagonal}$
- Whitening creates uncorrelated independent factors. CCA finds correlation between particular features in behavior and brain activity
- Linear, non-generative method

Probabilistic Canonical Correlation Analysis (Bach & Jordan, 2005)



- Assume, two private latent spaces Z^X, Z^Y and shared latent space Z^{shared} such that,

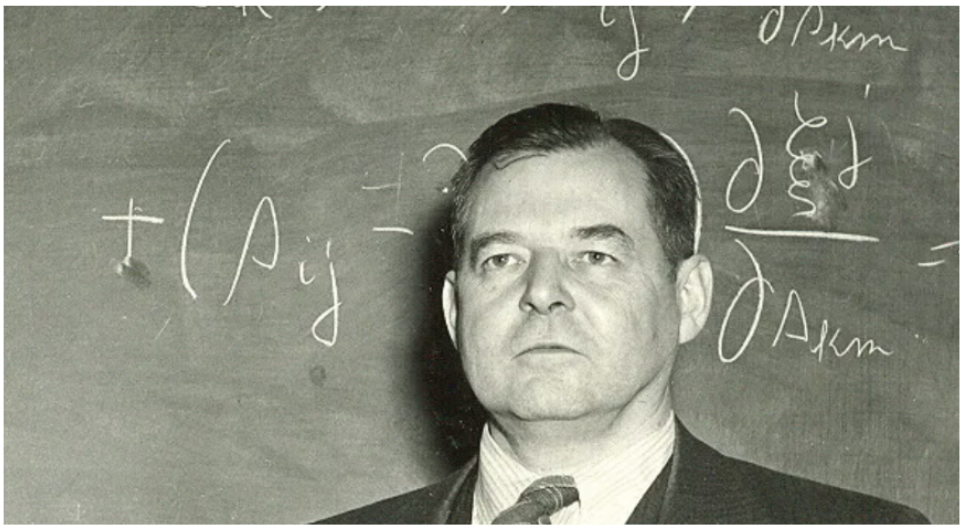
$$\tilde{X}^{CCA} = (\dots)Z^X + (\dots)Z^{shared}$$
$$\tilde{Y}^{CCA} = (\dots)Z^Y + (\dots)Z^{shared}$$

- Generative model
- Linear method

Standing on shoulders of giant: Harold Hotelling

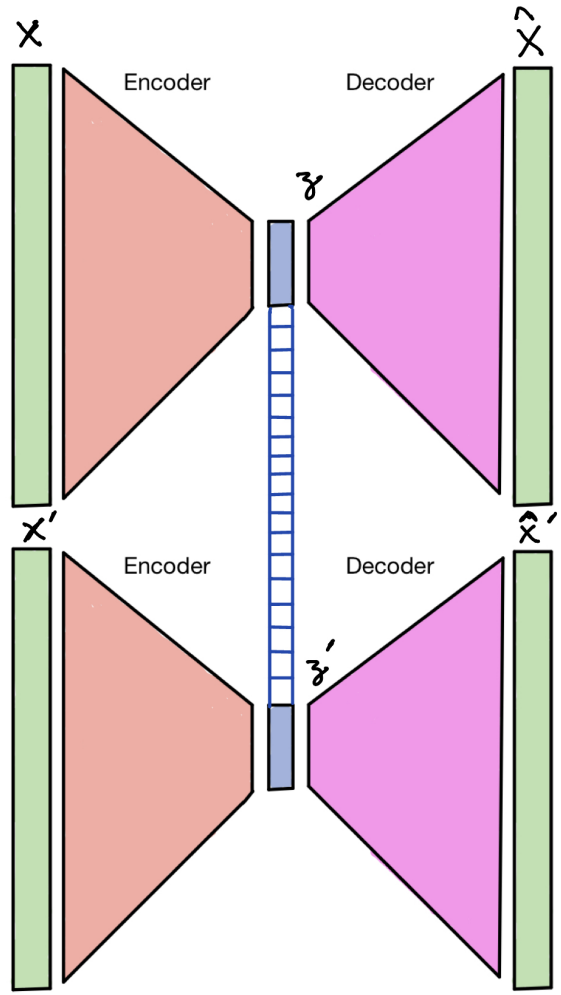
- Discovered PCA in 1933
- Discovered CCA in 1936

My work stands on the ideas pioneered by him.



Multi-modal Variational Autoencoders

1. Multi-modal VAE is a variant of vanilla VAE in which multiple datasets can be jointly input to the network.
2. It can model nonlinear correlations between modalities. It is much more expressive than CCA.



Motivation

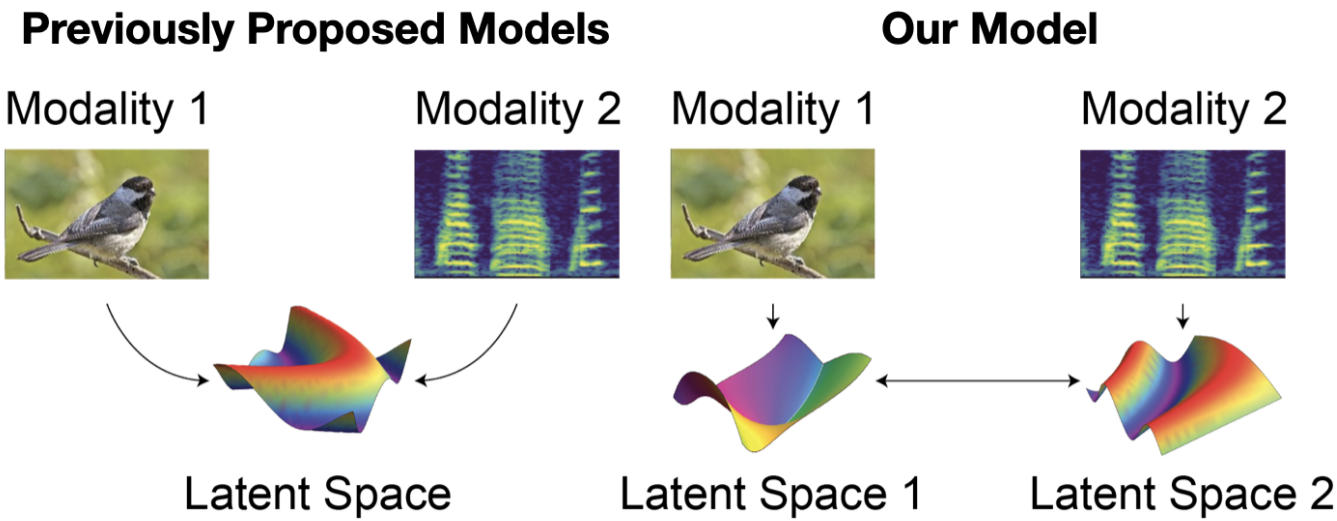
1. The lower dimensional latent representation in VAEs are often non-interpretable
   - My latent space learns the true latent variables up to a linear transformation (identifiable).
2. When high dimensional and low dimensional data are collected simultaneously, learning the features of low dimensional data becomes challenging.
   - My model has separate latent space for both modalities. This ensures that the modalities are being weighted equally in the learning process.

Motivation

1. The lower dimensional latent representation in VAEs are often non-interpretable
   - My latent space learns the true latent variables up to a linear transformation (identifiable).
2. When high dimensional and low dimensional data are collected simultaneously, learning the features of low dimensional data becomes challenging.
   - My model has separate latent space for both modalities. This ensures that the modalities are being weighted equally in the learning process.

Introducing POISE-VAE

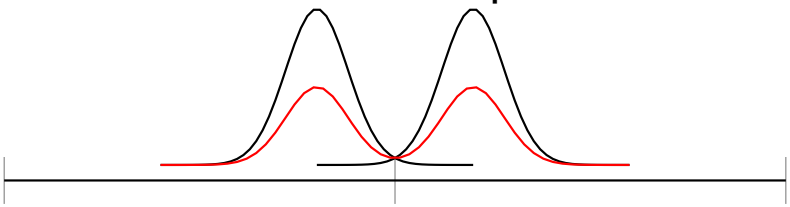
Product of Identifiable, Sufficient Experts (POISE) VAE.



Formulating Posterior Distribution

There are two ways to define the multimodal posterior, $q_\phi(z_1, z_2|x_1, x_2)$ in terms of unimodal posterior, $q_\phi(z_1, z_2|x_1)$ and $q_\phi(z_1, z_2|x_2)$:

Mixture of Experts



Black: Unimodal posterior
Red: Multimodal posterior

- Posterior is written as sum:

$$q_\phi(z_1, z_2|x_1, x_2) = \sum_i \alpha_i q_{\phi_i}(z_1, z_2|x_i)$$

- Like a healthy relationship (each expert can make decision). Does what either likes.

Product of Experts

Black: Unimodal posterior
Red: Multimodal posterior

- Posterior is written as product:

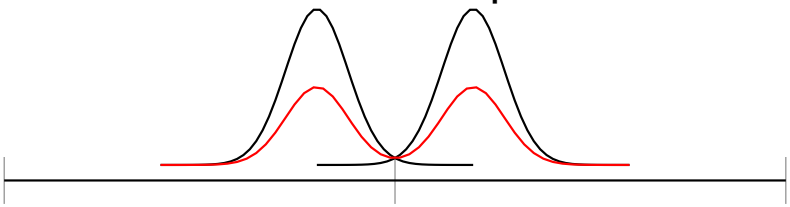
$$q_\phi(z_1, z_2|x_1, x_2) = p(z_1, z_2) \prod_{i=1}^{2} q_{\phi_i}(z_1, z_2|x_i)$$

- Like UN Security Council (each expert has veto power). Does what neither likes.

Formulating Posterior Distribution

There are two ways to define the multimodal posterior, $q_\phi(z_1, z_2|x_1, x_2)$ in terms of unimodal posterior, $q_\phi(z_1, z_2|x_1)$ and $q_\phi(z_1, z_2|x_2)$:

Mixture of Experts



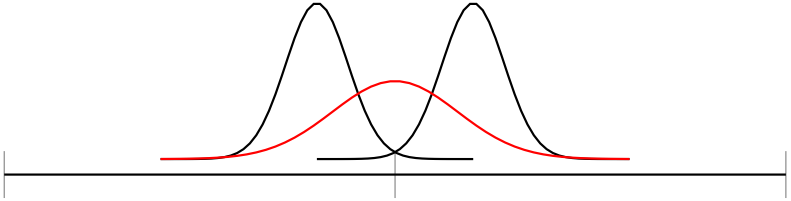
Black: Unimodal posterior
Red: Multimodal posterior

- Posterior is written as sum:

$$q_\phi(z_1, z_2|x_1, x_2) = \sum_i \alpha_i q_{\phi_i}(z_1, z_2|x_i)$$

- Like a healthy relationship (each expert can make decision). Does what either likes.

Product of Experts



Black: Unimodal posterior
Red: Multimodal posterior

- Posterior is written as product:

$$q_\phi(z_1, z_2|x_1, x_2) = p(z_1, z_2)\prod_{i=1}^{2} q_{\phi_i}(z_1, z_2|x_i)$$

- Like UN Security Council (each expert has veto power). Does what neither likes.

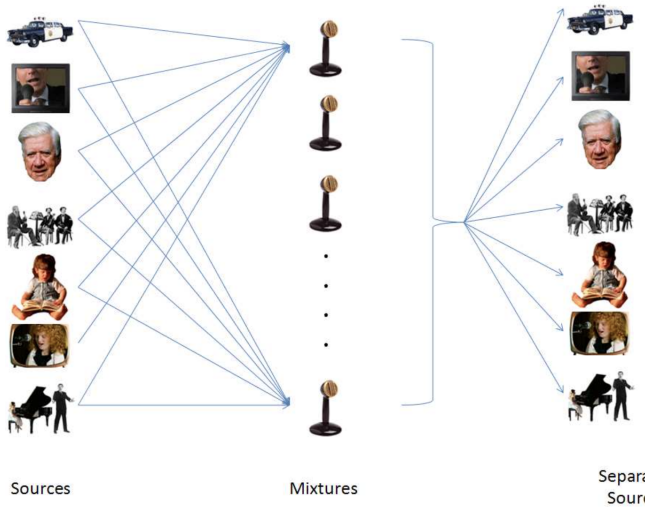
Digression: Notion of identifiability

Source-Separation problem

There are N sources of sound, $z(t) \in \mathbb{R}^N$ and N microphones detecting the cacophony of sounds, $x(t) \in \mathbb{R}^N$. Assume,

$$x(t) = Az(t)$$

Given x(t) can we *identify* individual sound sources, z(t)?



Independent Component Analysis

Source-Separation problem can be solved by Independent Component Analysis up to permutation and scaling.

Identifable system

A system is *identifiable* if we can recover the true $\mathbf{z}$ up to a linear transformation. So, if our estimate for $\mathbf{z}$ is $\hat{\mathbf{z}}$ then,

$$\mathbf{z} = \mathbf{W}\hat{\mathbf{z}}$$

Here, $\mathbf{W} = \mathbf{PS}$ is a $n \times n$ matrix. $\mathbf{P}$ is permutation matrix and $\mathbf{S}$ is a diagonal scaling matrix.

Can we have identifiable systems in which number of microphones are less than number of sources and they are related by non-linear transformation?

Independent Component Analysis

Source-Separation problem can be solved by Independent Component Analysis up to permutation and scaling.

Identifable system

A system is *identifiable* if we can recover the true $\mathbf{z}$ up to a linear transformation. So, if our estimate for $\mathbf{z}$ is $\hat{\mathbf{z}}$ then,

$$\mathbf{z} = \mathbf{W}\hat{\mathbf{z}}$$

Here, $\mathbf{W} = \mathbf{PS}$ is a $n \times n$ matrix. $\mathbf{P}$ is permutation matrix and $\mathbf{S}$ is a diagonal scaling matrix.

Can we have identifiable systems in which number of microphones are less than number of sources and they are related by non-linear transformation?

Non-linear Independent Component Analysis

Problem

Given, $z(t) \in \mathbb{R}^M$, $x(t) \in \mathbb{R}^N$ and f is some invertible non-linear mapping between them,

$$x = f(z) + \epsilon$$

Can we still *identify* the sources, z?

In general, answer is no. But recently, certain situations where it is possible has been identified (no pun intended).

Solving Non-linear ICA

Key idea: Impose constraints $\rightarrow$ Unique solution.
For source-separation problem, suppose in addition to audio recordings x(t), we also have video of people producing sounds, u(t).

Theorem: (Khemakhem, et. al. 2020)

If we assume z becomes conditionally independent given u, that is

$$p(z|u) = \prod_i p(z_i|u).$$

Then the system is identifiable.

Notice, VAE framework is very similar to non-linear ICA. I will be exploiting this fact in building my VAE model.

Three notions of identifiability

Standard definition

A system is identifiable if we can recover z exactly, that is

$$z = \hat{z}$$

Classical ICA definition

A system is identifiable if we can recover z up to a permutation (P) and scaling (S), that is

$$z = W\hat{z} = PS\hat{z}$$

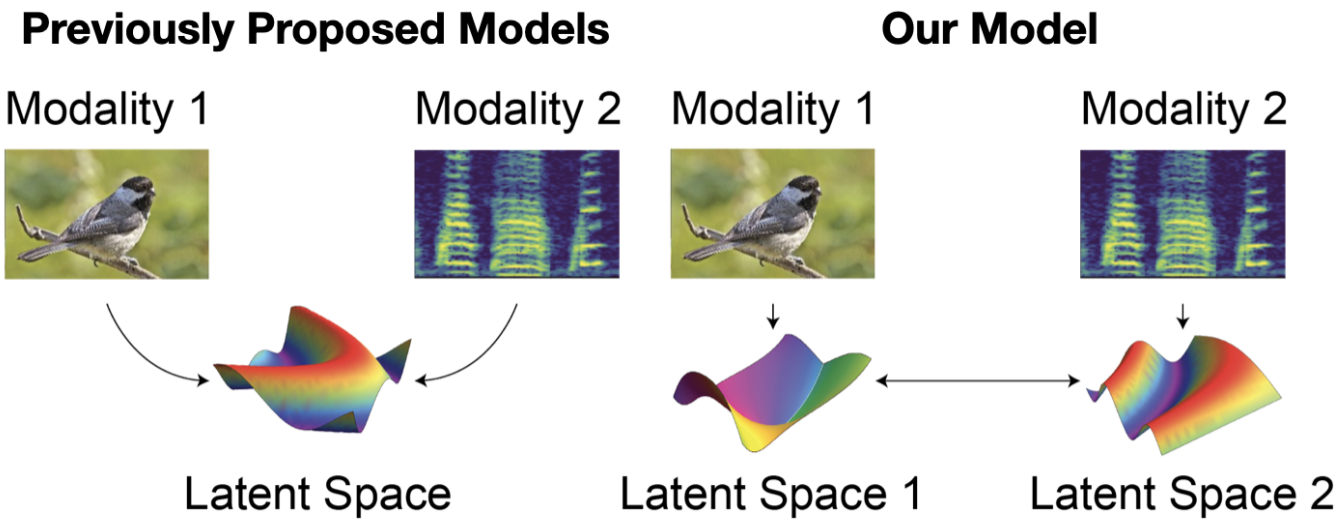
Nonlinear ICA definition

A system is identifiable if we can recover *sufficient statistics* of z, $T(z)$ up to scaling (S) and translation (c), that is

$$T(z) = ST(\hat{z}) + c$$

Introducing POISE-VAE

Product of Identifiable, Sufficient Experts (POISE) VAE.



Constructing POISE-VAE

For any multi-modal VAE variational free energy is given by,

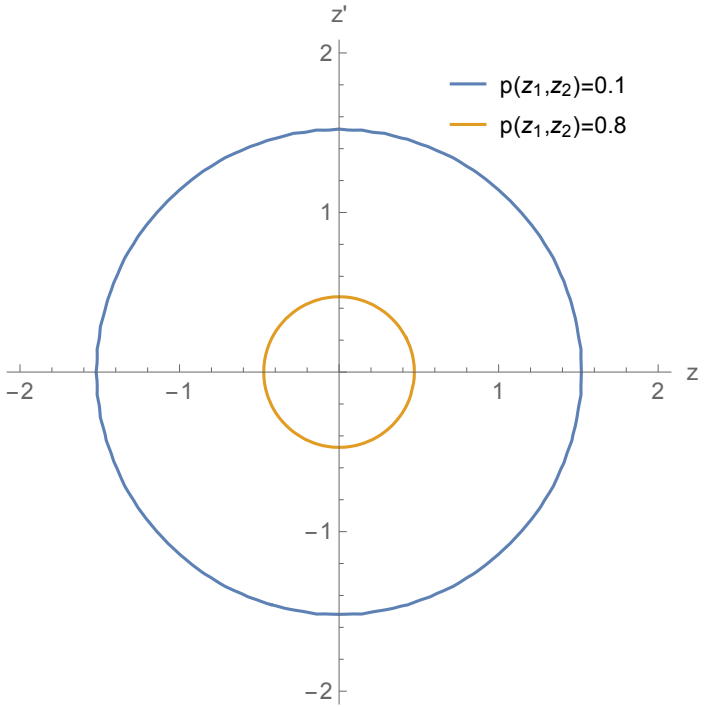
$$\mathcal{F}_{var}(x, x') = \underbrace{||x - \hat{x}||_2^2 + ||x' - \hat{x'}||_2^2}_{reconstruction} + \underbrace{D_{KL}(q_\phi(z, z'|x, x')||p(z, z'))}_{regularizer}$$

We need to construct the right form for prior, $p(z, z')$ and posterior, $q_\phi(z, z'|x, x')$ for identifiability.

Formulating Prior Distribution

1. Ansatz 1:
   $p(z, z') \sim \exp[-z^2 - z'^2]$

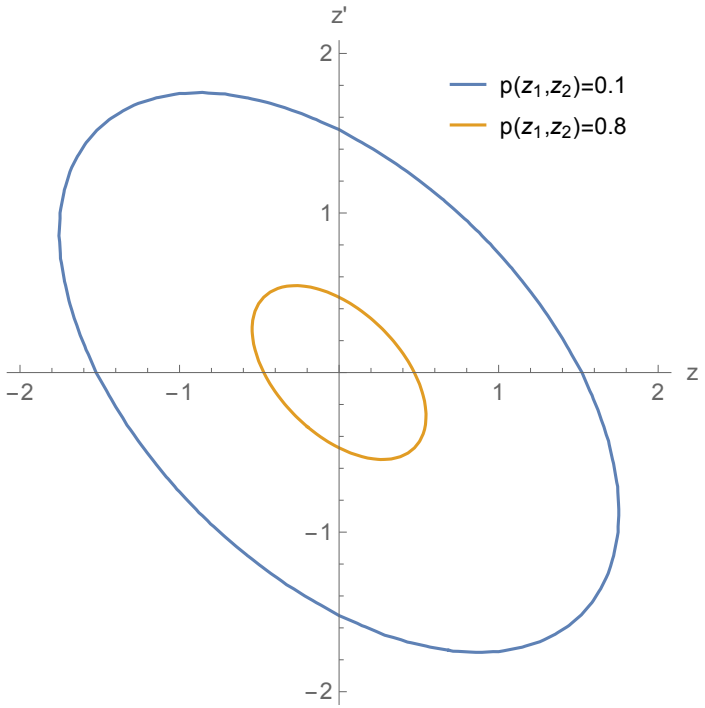
2. Latent spaces are independent. So, we won't learn relationship between the modalities



Contour lines for bivariate Gaussian distribution

Formulating prior distribution

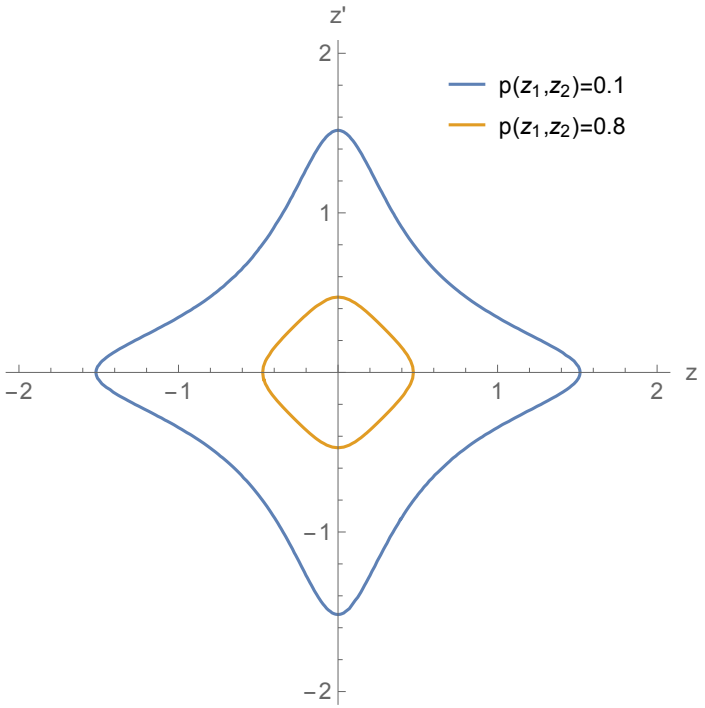
1. Ansatz 2:
   $p(z, z') \sim \exp[-z^2 - z'^2 + gzz']$
2. We are learning correlations between modalities



Correlated bivariate Gaussian distribution ($g = -1$)

Formulating Prior Distribution

1. Ansatz 3:
   $p(z, z') \sim \exp[-z^2 - z'^2 + g_1 z z' + g_2 z^2 z'^2]$
2. Concavity of level curves ensure that sparse subset of latent variables make decisions
3. Adding g's makes the model identifiable.



Contour plot($g_1 = 0, g_2 = -10$)

Identifiability of POISE-VAE

Theorem: (Khemakhem, et. al. 2020)

If we assume z becomes conditionally independent given u, that is

$$p(z|u) = \prod_i p(z_i|u).$$

Then the system is identifiable.

What if we let u=z'? That is, we are using z' as the auxiliary variable for z and vice-versa.

Theorem:

If we assume z becomes conditionally independent given z' and vice-versa, that is

$$p(z|z') = \prod_i p(z_i|z')$$

$$p(z'|z) = \prod_i p(z'_i|z)$$

Then the system is identifiable and the two latent variables are learnt up to a *different* linear transformation.

Identifiability of POISE-VAE

Theorem: (Khemakhem, et. al. 2020)

If we assume z becomes conditionally independent given u, that is

$$p(z|u) = \prod_i p(z_i|u).$$

Then the system is identifiable.

What if we let u=z'? That is, we are using z' as the auxiliary variable for z and vice-versa.

Theorem:

If we assume z becomes conditionally independent given z' and vice-versa, that is

$$p(z|z') = \prod_i p(z_i|z')$$

$$p(z'|z) = \prod_i p(z'_i|z)$$

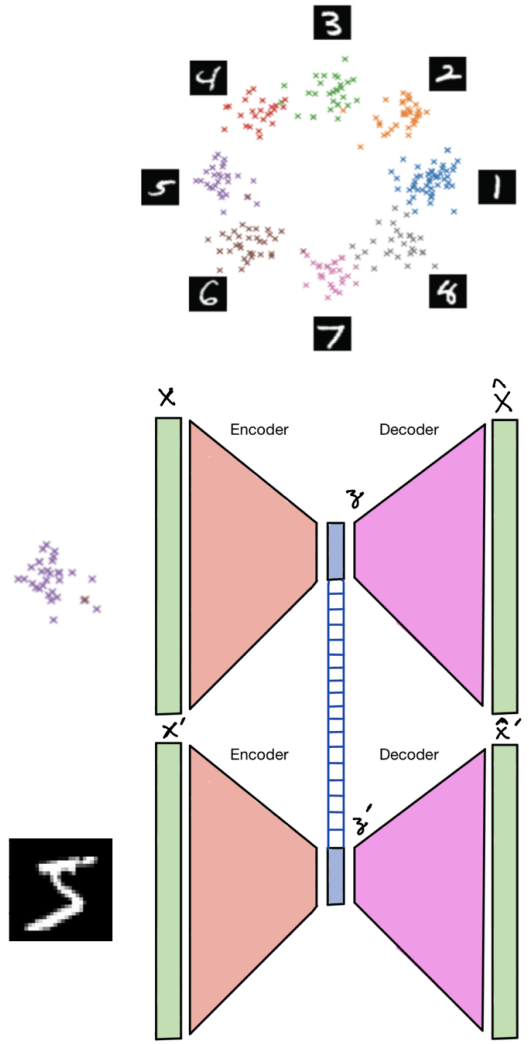
Then the system is identifiable and the two latent variables are learnt up to a *different* linear transformation.

Mode Starvation Experiment

- In experiments, often high and low dimensional data are collected simultaneously.
- We don't want learning of the low dimensional mode to be starved due to the high dimensional mode.
- This experiment tests whether POISE-VAE mitigates this problem, since it has a private latent space for both modalities.

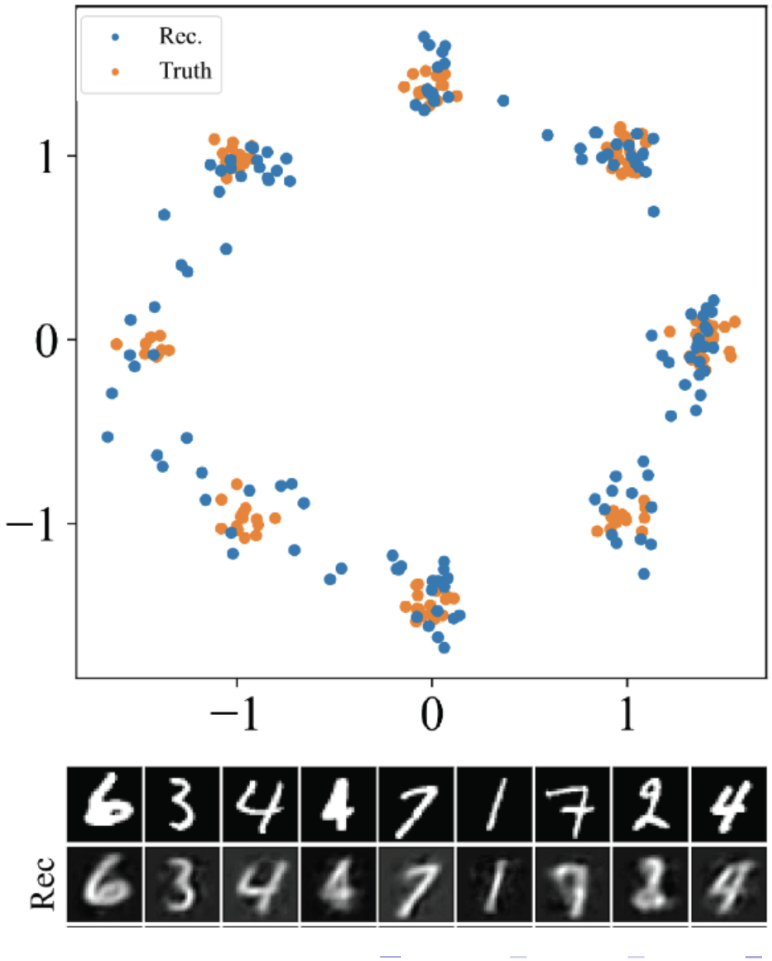
MNIST-Gaussian Experiment

1. Input to POISE-VAE:
   - Modality 1: Mixture of eight Gaussians arranged in a circle
   - Modality 2: MNIST
2. Methodology: Each mode of mixture of Gaussian was paired with a specific MNIST digit from 1 to 8
3. Motivation: Can POISE-VAE learn both modalities simulatenously?



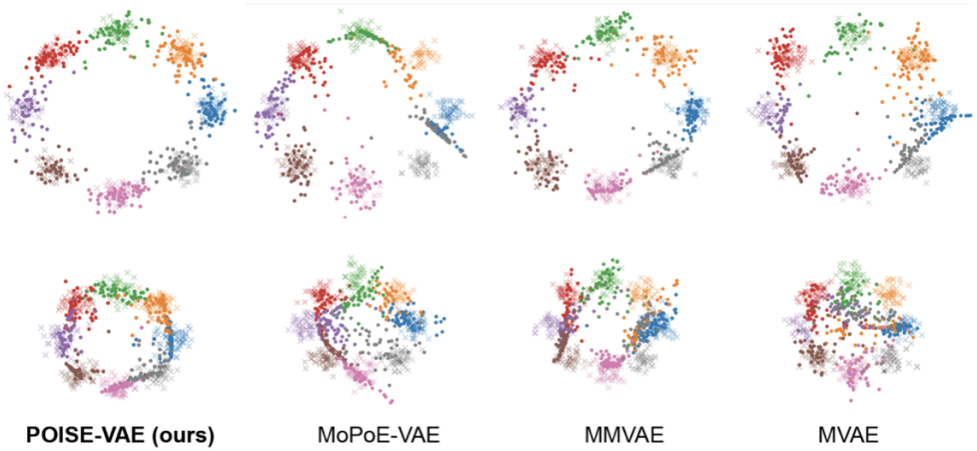
MNIST-Gaussian Results

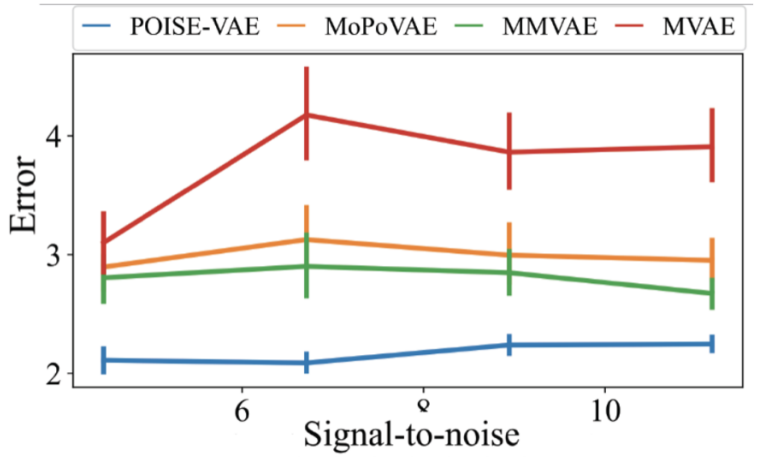
- POISE-VAE was able to learn both modalities simultaneously
- Since, POISE-VAE has a private latent space for both modalities, lower dimensional dataset is also able to learn.



Comparision with other models

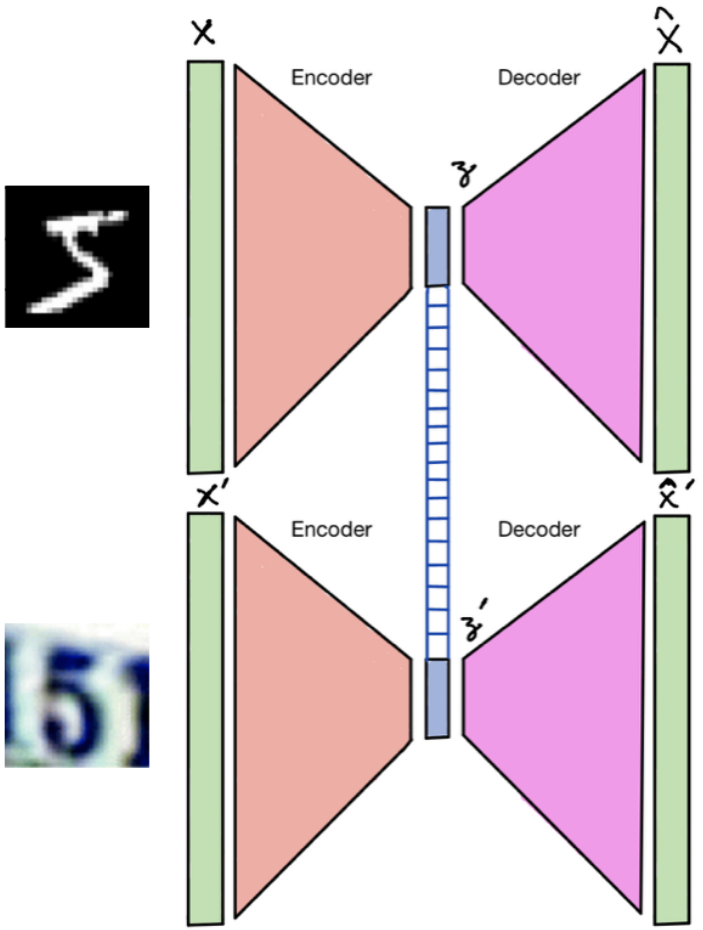
- At higher SNRs (large radius), all models roughly capture the lower dimensional mode's latent space
- At low SNR (small radius), most multimodal VAEs struggle to learn the eight clusters. POISE-VAE correctly learns even this comparatively small signal





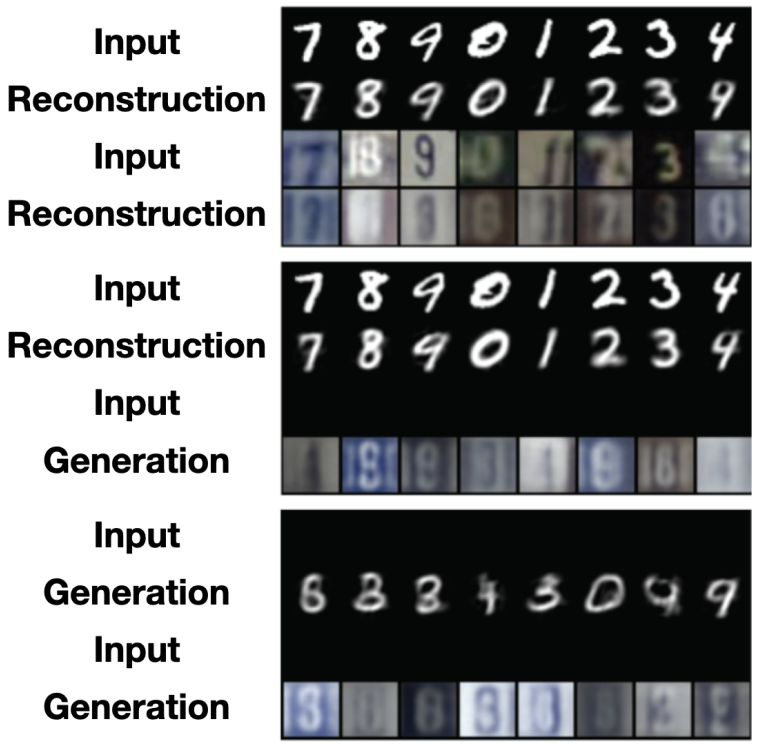
MNIST-SVHN Experiment

1. Input to POISE-VAE:
   - Modality 1: MNIST
   - Modality 2: SVHN
2. Methodology: Same digit class from the two modality is input to the VAE
3. Motivation: Can POISE-VAE learn to generate consistent digits in absence of one or both modalities?



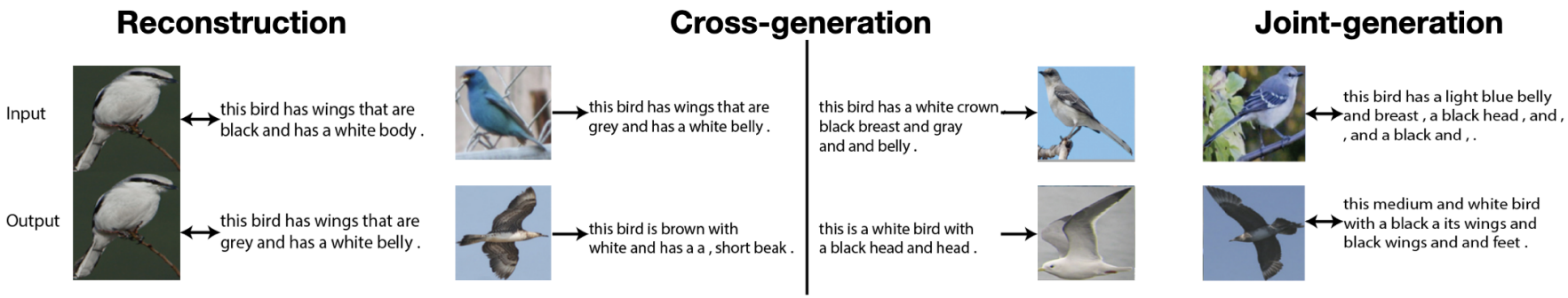
MNIST-SVHN Results

- Top: Sample reconstructions.
  Center: Cross generation (one modality absent).
  Bottom: Joint generation (both modalities absent).
- Implication for neuroscience: POISE-VAE trained on behavior video, neural recording. Trained network should be able to predict neural activity for novel behaviour



CUB-Caption Experiment

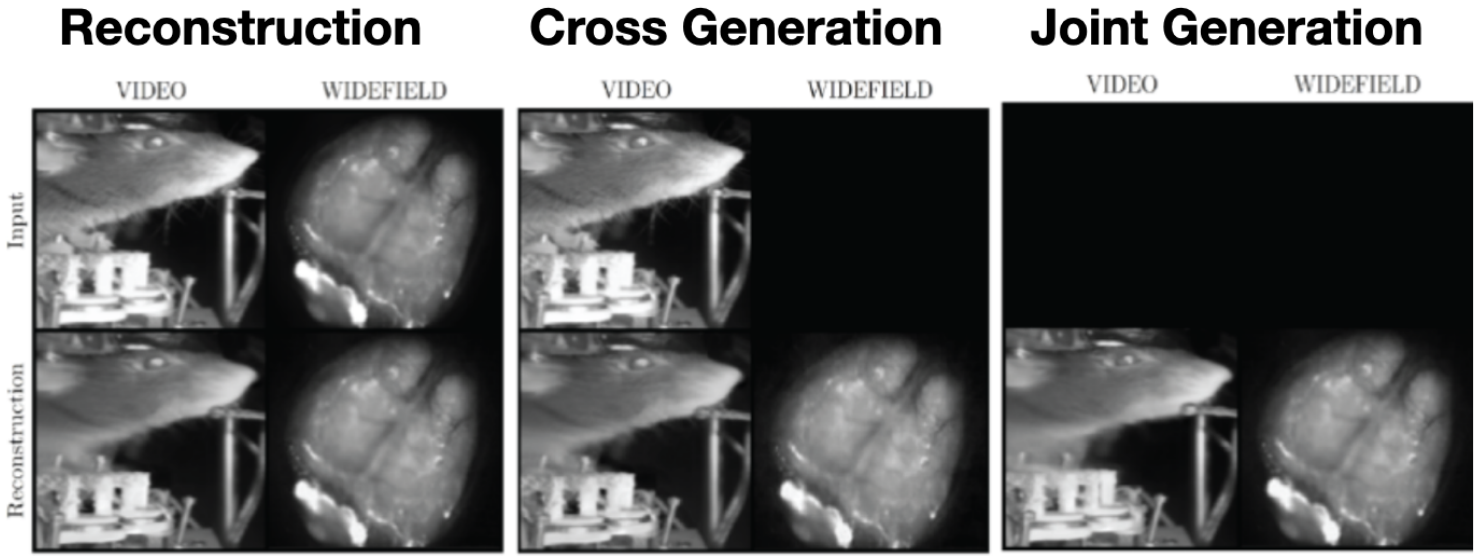
1. Input to POISE-VAE:
   - Modality 1: Bird Images
   - Modality 2: Image Caption
2. Motivation: Can POISE-VAE learn textual descriptions along with images?
3. Results:



Real neural data analysis

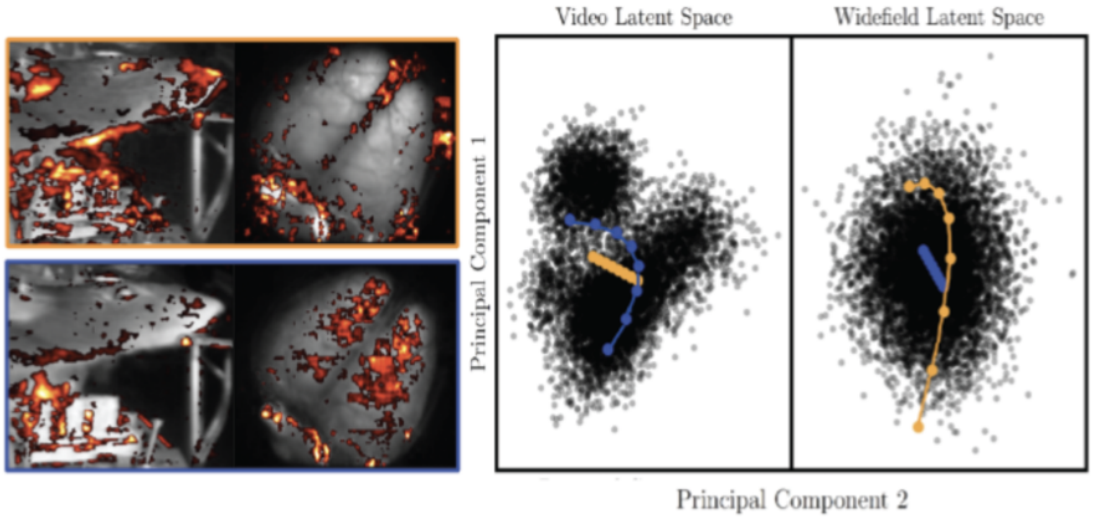
Data from Musall,*et. al.* 2019 paper was used for this experiment

1. Input to POISE-VAE:
   - Modality 1: Mouse video
   - Modality 2: Widefield calcium imaging
2. Motivation:
   - Can POISE-VAE generate neural activity from behavioral data and vice-versa?



Exploring the latent space

- Latent traversals capture shared relations between modalities.
- Linear paths through video latent space (orange) produce curvilinear trajectories through widefield space
- Linear paths through widefield latent space produce curvilinear trajectories through video space



Desiderata

Conclusion

How to infer underlying latent variables of high-dimensional data?

1. Project 1: Latent variable modelling of unimodal data (applications project)
    1. Found no difference in vocalization between hearing and deaf mice.
2. Project 2: Latent variable modelling of multimodal data (methods project)
    1. World Premiere of the first multi-modal, identifiable VAE
    2. Tested POISE VAE on various machine learning and neuroscience datasets. It is able to learn high dimensional and low dimensional data together. It is able to cross-generate and jointly generate new data, albeit not SOTA.

Conclusion

How to infer underlying latent variables of high-dimensional data?

1. Project 1: Latent variable modelling of unimodal data (applications project)
   1. Found no difference in vocalization between hearing and deaf mice.
2. Project 2: Latent variable modelling of multimodal data (methods project)
   1. World Premiere of the first multi-modal, identifiable VAE
   2. Tested POISE VAE on various machine learning and neuroscience datasets. It is able to learn high dimensional and low dimensional data together. It is able to cross-generate and jointly generate new data, albeit not SOTA.

Conclusion

How to infer underlying latent variables of high-dimensional data?

1. Project 1: Latent variable modelling of unimodal data (applications project)
    1. Found no difference in vocalization between hearing and deaf mice.
2. Project 2: Latent variable modelling of multimodal data (methods project)
    1. World Premiere of the first multi-modal, identifiable VAE
    2. Tested POISE VAE on various machine learning and neuroscience datasets. It is able to learn high dimensional and low dimensional data together. It is able to cross-generate and jointly generate new data, albeit not SOTA.

Acknowledgment

Lab Mates (Friends):

- Anne Draelos
- Seth Madlon-Kay
- Na Young Jun
- Pranjal Gupta
- Kevin O'Neill
- Daniela de Albuquerque
- Trevor Alston
- Raphael Geddert
- Miles Martinez
- Liz O'Gorman

Physics Cohort (Friends):

- Baran
- Alexey
- Adryanna
- James
- Nathan
- Jay
- Brodie
- Drew

Friends:

- Max
- Aghil
- Ziyi
- Tala
- Lalit & Swati
- Vani
- Shreya
- Miheer
- Anurag
- Alex
- Siddhi
- Atul

Family:

- Mom
- Dad
- Atulit
- Anushree
- Kushagra
- Atsew

Acknowledgment: Committee Members

1. Prof. Dan Scolnic: Thank you for for sparing time to be in my committee and for asking incisive questions during our meetings.
2. Prof. Richard Mooney: Thank you helping me develop collaboration with your lab and for always being supportive and generous with your time.
3. Prof. Nicolas Brunel: Thank you for mentoring and nurturing me for two years in graduate school. I really appreciate everything you've taught me.
4. Prof. Henry Greenside: Thank you for introducing me to so many fields in science. I think you see whole of science (nature) as one organic entity and you've shown me how beautiful she looks when admired in her entirety. I hope to carry that worldview with me for the rest of my life
5. Prof. John Pearson: My mom says, "He's been like a godfather to you." And I cannot agree more. You've taught me more than anyone else in my life. But more than knowledge, you've shown me how to be kind, empathetic and caring person. Thank you for that.

Acknowledgment: Committee Members

1. Prof. Dan Scolnic: Thank you for for sparing time to be in my committee and for asking incisive questions during our meetings.
2. Prof. Richard Mooney: Thank you helping me develop collaboration with your lab and for always being supportive and generous with your time.
3. Prof. Nicolas Brunel: Thank you for mentoring and nurturing me for two years in graduate school. I really appreciate everything you've taught me.
4. Prof. Henry Greenside: Thank you for introducing me to so many fields in science. I think you see whole of science (nature) as one organic entity and you've shown me how beautiful she looks when admired in her entirety. I hope to carry that worldview with me for the rest of my life
5. Prof. John Pearson: My mom says, "He's been like a godfather to you." And I cannot agree more. You've taught me more than anyone else in my life. But more than knowledge, you've shown me how to be kind, empathetic and caring person. Thank you for that.

Acknowledgment: Committee Members

1. Prof. Dan Scolnic: Thank you for for sparing time to be in my committee and for asking incisive questions during our meetings.
2. Prof. Richard Mooney: Thank you helping me develop collaboration with your lab and for always being supportive and generous with your time.
3. Prof. Nicolas Brunel: Thank you for mentoring and nurturing me for two years in graduate school. I really appreciate everything you've taught me.
4. Prof. Henry Greenside: Thank you for introducing me to so many fields in science. I think you see whole of science (nature) as one organic entity and you've shown me how beautiful she looks when admired in her entirety. I hope to carry that worldview with me for the rest of my life
5. Prof. John Pearson: My mom says, "He's been like a godfather to you." And I cannot agree more. You've taught me more than anyone else in my life. But more than knowledge, you've shown me how to be kind, empathetic and caring person. Thank you for that.

Acknowledgment: Committee Members

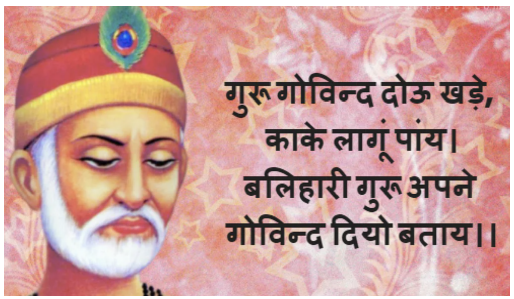
1. Prof. Dan Scolnic: Thank you for for sparing time to be in my committee and for asking incisive questions during our meetings.
2. Prof. Richard Mooney: Thank you helping me develop collaboration with your lab and for always being supportive and generous with your time.
3. Prof. Nicolas Brunel: Thank you for mentoring and nurturing me for two years in graduate school. I really appreciate everything you've taught me.
4. Prof. Henry Greenside: Thank you for introducing me to so many fields in science. I think you see whole of science (nature) as one organic entity and you've shown me how beautiful she looks when admired in her entirety. I hope to carry that worldview with me for the rest of my life
5. Prof. John Pearson: My mom says, "He's been like a godfather to you." And I cannot agree more. You've taught me more than anyone else in my life. But more than knowledge, you've shown me how to be kind, empathetic and caring person. Thank you for that.

Acknowledgment: Committee Members

1. Prof. Dan Scolnic: Thank you for for sparing time to be in my committee and for asking incisive questions during our meetings.
2. Prof. Richard Mooney: Thank you helping me develop collaboration with your lab and for always being supportive and generous with your time.
3. Prof. Nicolas Brunel: Thank you for mentoring and nurturing me for two years in graduate school. I really appreciate everything you've taught me.
4. Prof. Henry Greenside: Thank you for introducing me to so many fields in science. I think you see whole of science (nature) as one organic entity and you've shown me how beautiful she looks when admired in her entirety. I hope to carry that worldview with me for the rest of my life
5. Prof. John Pearson: My mom says, "He's been like a godfather to you." And I cannot agree more. You've taught me more than anyone else in my life. But more than knowledge, you've shown me how to be kind, empathetic and caring person. Thank you for that.

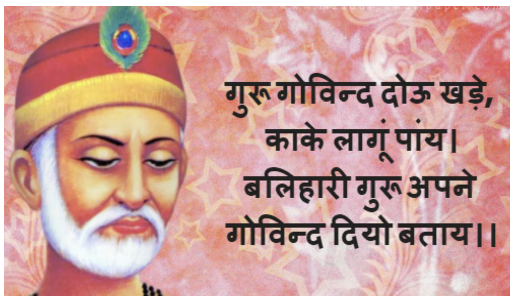
Acknowledgment: Committee Members

1. Prof. Dan Scolnic: Thank you for for sparing time to be in my committee and for asking incisive questions during our meetings.
2. Prof. Richard Mooney: Thank you helping me develop collaboration with your lab and for always being supportive and generous with your time.
3. Prof. Nicolas Brunel: Thank you for mentoring and nurturing me for two years in graduate school. I really appreciate everything you've taught me.
4. Prof. Henry Greenside: Thank you for introducing me to so many fields in science. I think you see whole of science (nature) as one organic entity and you've shown me how beautiful she looks when admired in her entirety. I hope to carry that worldview with me for the rest of my life
5. Prof. John Pearson: My mom says, "He's been like a godfather to you." And I cannot agree more. You've taught me more than anyone else in my life. But more than knowledge, you've shown me how to be kind, empathetic and caring person. Thank you for that.



Acknowledgment: Committee Members

1. Prof. Dan Scolnic: Thank you for for sparing time to be in my committee and for asking incisive questions during our meetings.
2. Prof. Richard Mooney: Thank you helping me develop collaboration with your lab and for always being supportive and generous with your time.
3. Prof. Nicolas Brunel: Thank you for mentoring and nurturing me for two years in graduate school. I really appreciate everything you've taught me.
4. Prof. Henry Greenside: Thank you for introducing me to so many fields in science. I think you see whole of science (nature) as one organic entity and you've shown me how beautiful she looks when admired in her entirety. I hope to carry that worldview with me for the rest of my life
5. Prof. John Pearson: My mom says, "He's been like a godfather to you." And I cannot agree more. You've taught me more than anyone else in my life. But more than knowledge, you've shown me how to be kind, empathetic and caring person. Thank you for that.



You've all brought me closer to the Gods. Thank you for that! *Fini!*