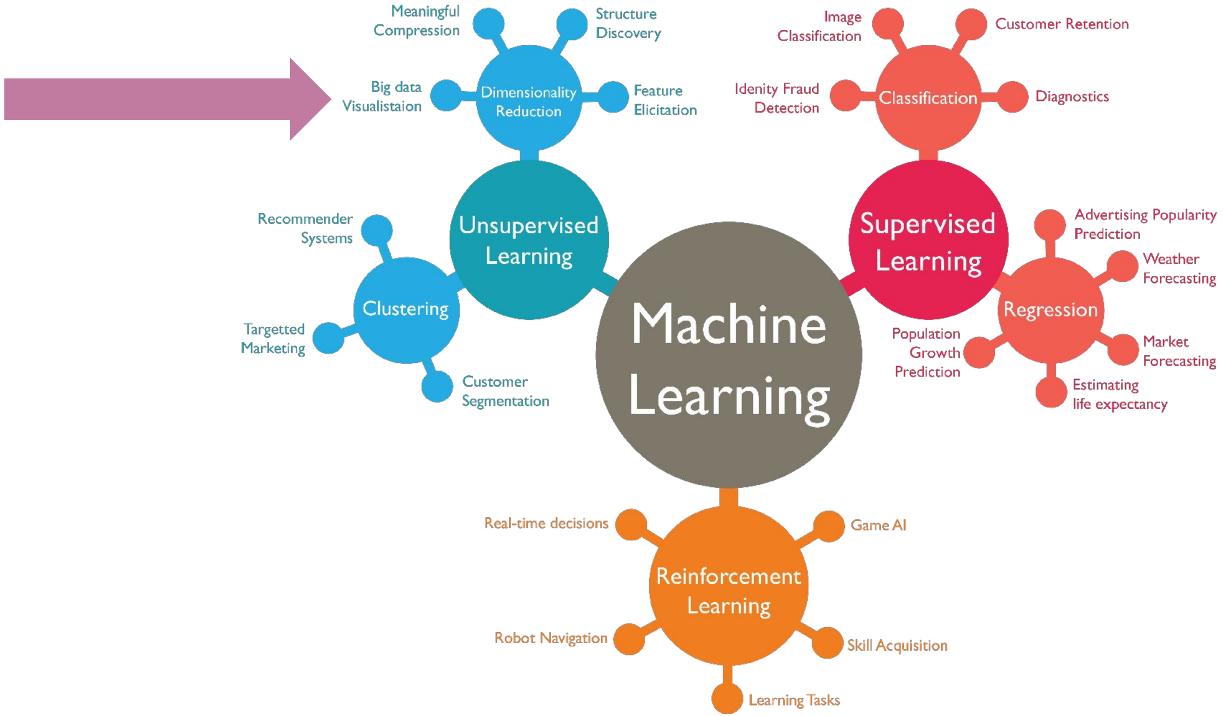
Autoencoders

CompCancer deep learning workshop

Jonathan Ronen

The machine learning landscape



https://towardsdatascience.com/machine-learning-algorithms-in-laymans-terms-part-1-d0368d769a7b

Unsupervised representation learning

Why dimensionality reduction?

- The curse of dimensionality
  - Most downstream analysis benefits from lower-dimension data
- Reduce multicollinearity
  - Most downstream algorithms you'll use assume some sort of independence
- Pattern recognition
  - E.g. discover biomarkers
  - Data compression
- Visualization
  - If 2D/3D...

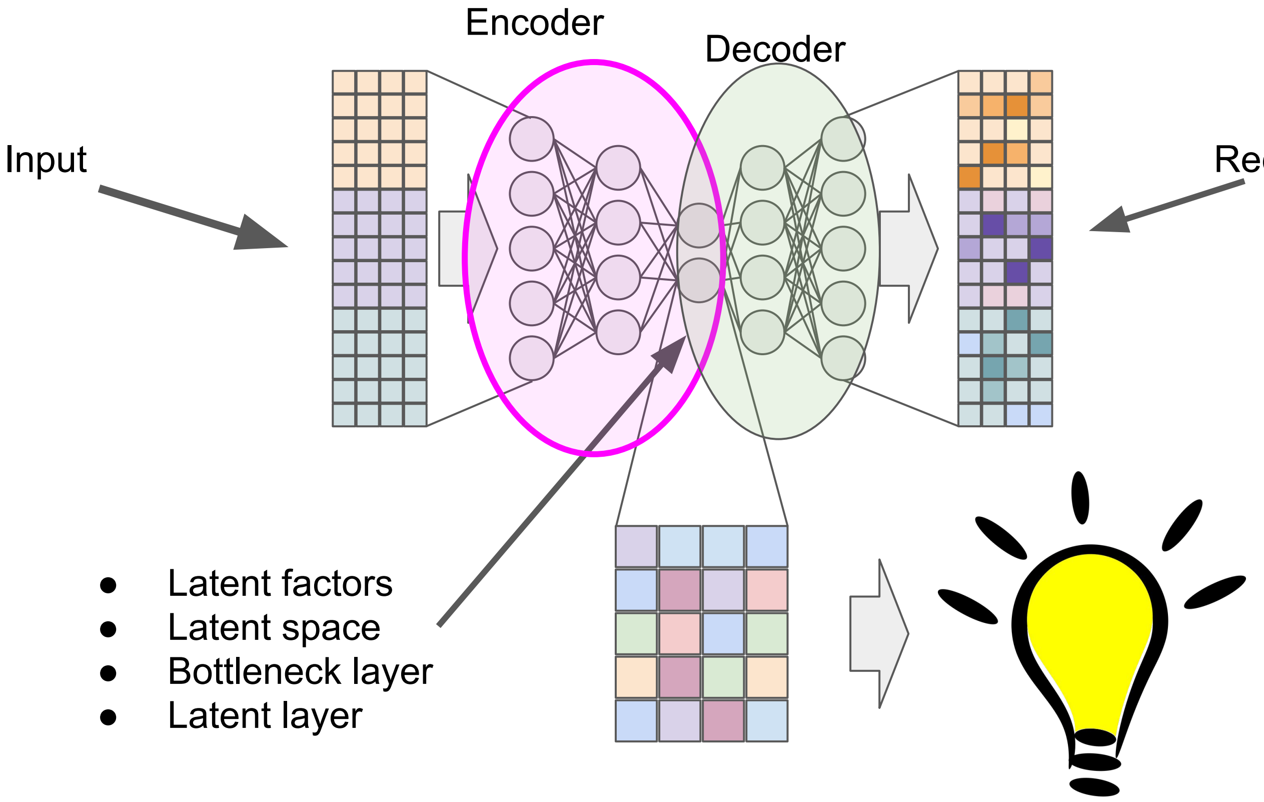
Dimensionality reduction

Other classes of dimensionality reduction

- Feature selection (variance filtering)
- Matrix factorization (PCA, NMF)
- Graph layout (tSNE, UMAP)

Autoencoders are different, but can do the same things sometimes

Autoencoders look-ahead & nomenclature

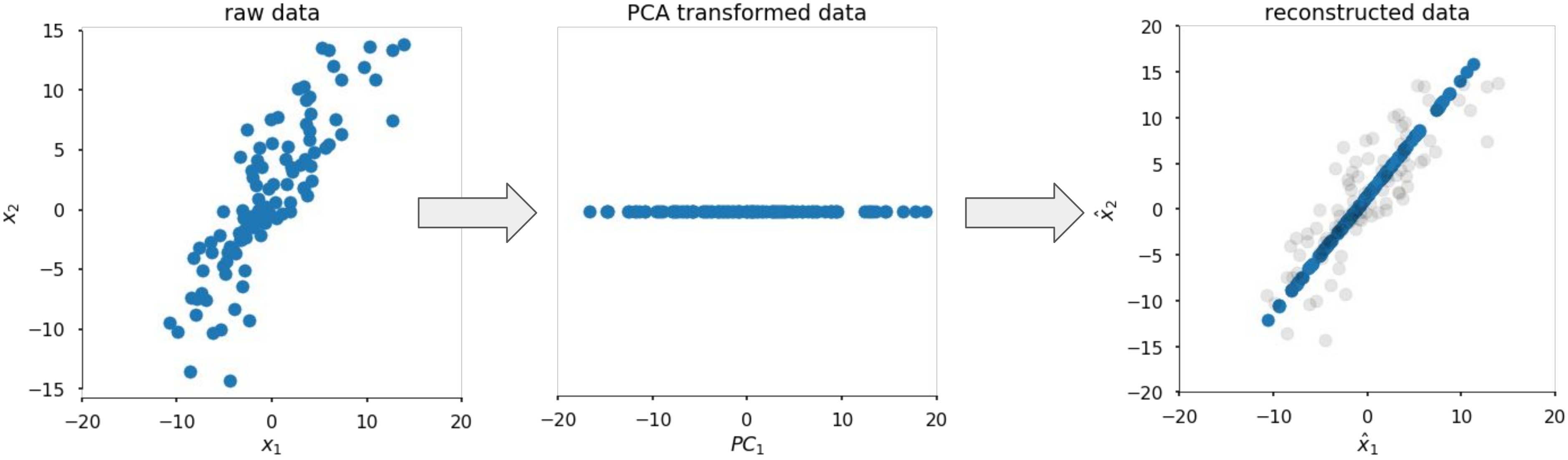


- Latent factors
- Latent space
- Bottleneck layer
- Latent layer

Objective

$$\underset{W}{\operatorname{argmin}} ||x - \hat{x}||^2$$

PCA for dimensionality reduction



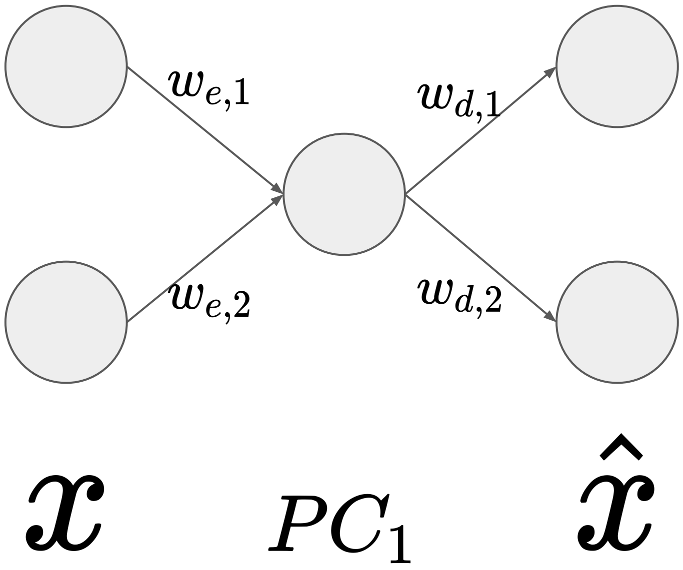
$$PC_1 = XU \qquad \hat{x} = PC_1 U^T$$

PCA for dimensionality reduction

- The U that maximizes the variance of PC1 $|PC_1|^2$
- also minimizes the reconstruction error $|x - \hat{x}|^2$
    - Note: this is not the same as OLS,
      which minimizes $|y - \hat{y}|^2$

There are efficient solvers for this, but we **could** also use **backpropagation**

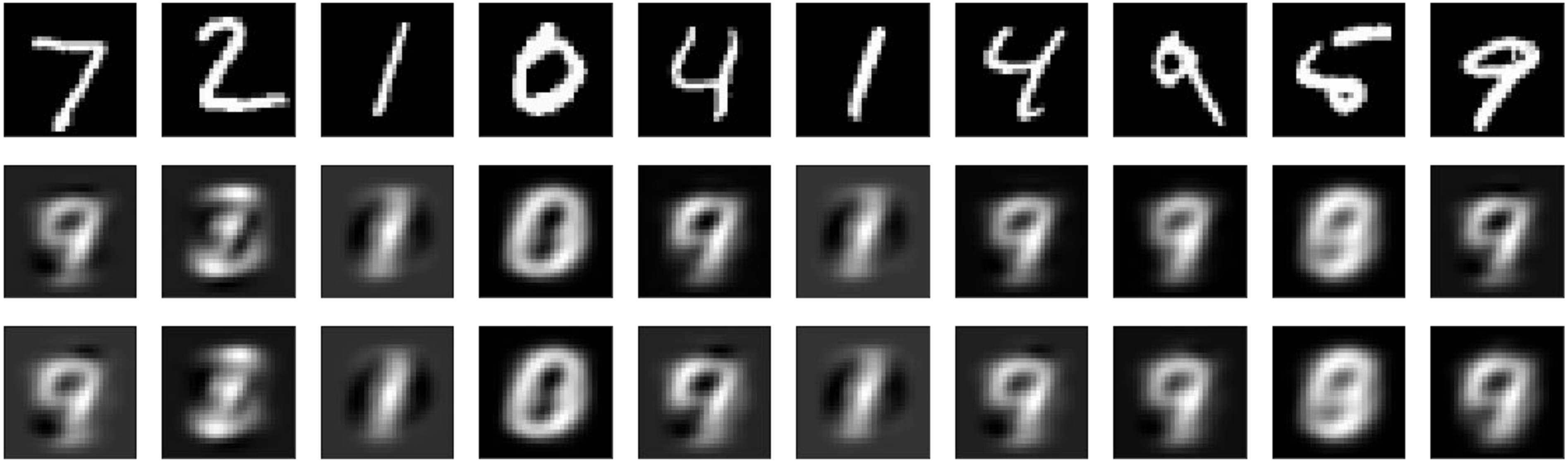
PCA through backpropagation



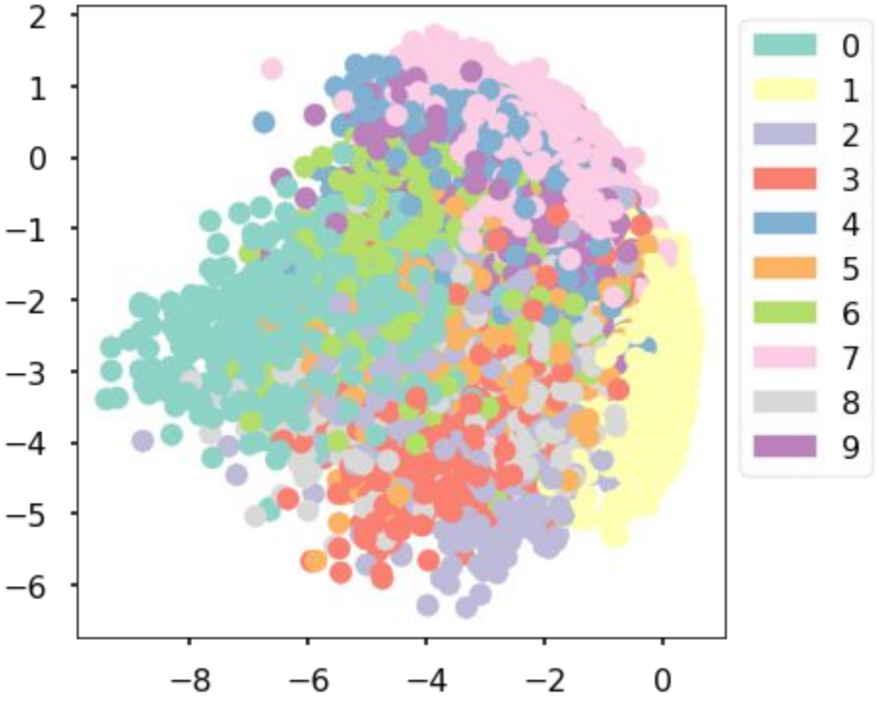
- $$\underset{W}{\mathbf{argmin}} |x - \hat{x}|^2$$

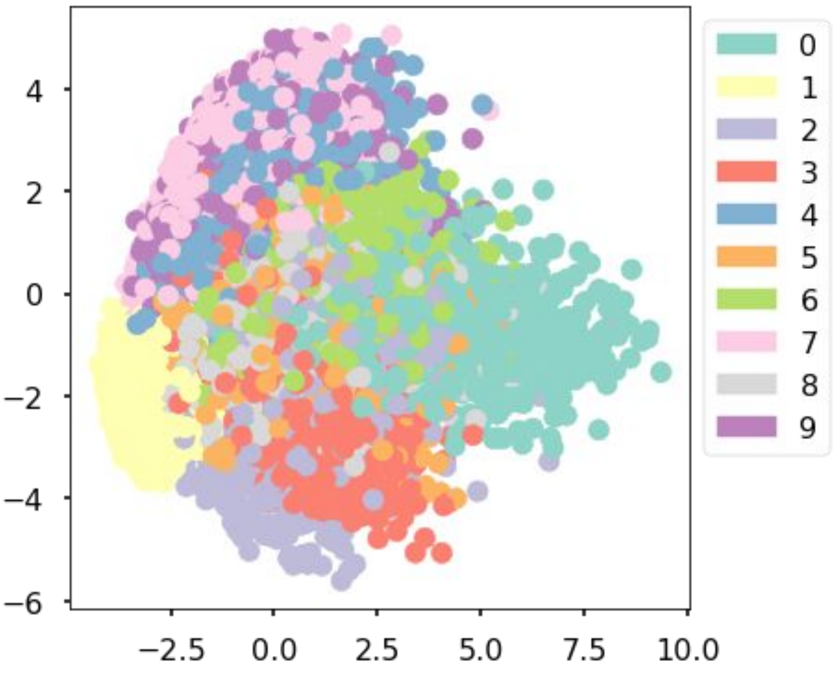
- This is an autoencoder
- If the neurons are linear, it is similar to PCA
  - Caveat: PCs are orthogonal, autoencoded components are not - but they will span the same space

PCA vs linear autoencoders for MNIST



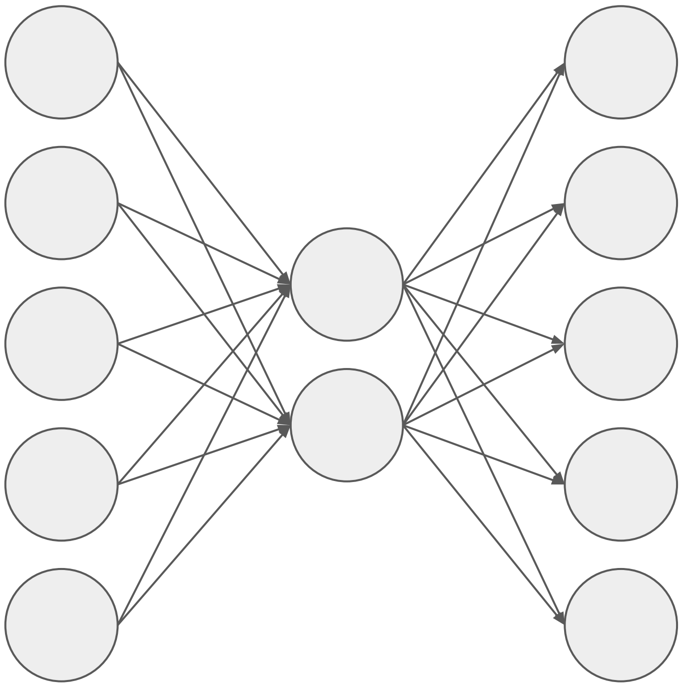
PCA vs linear autoencoders for MNIST





Autoencoders can be nonlinear

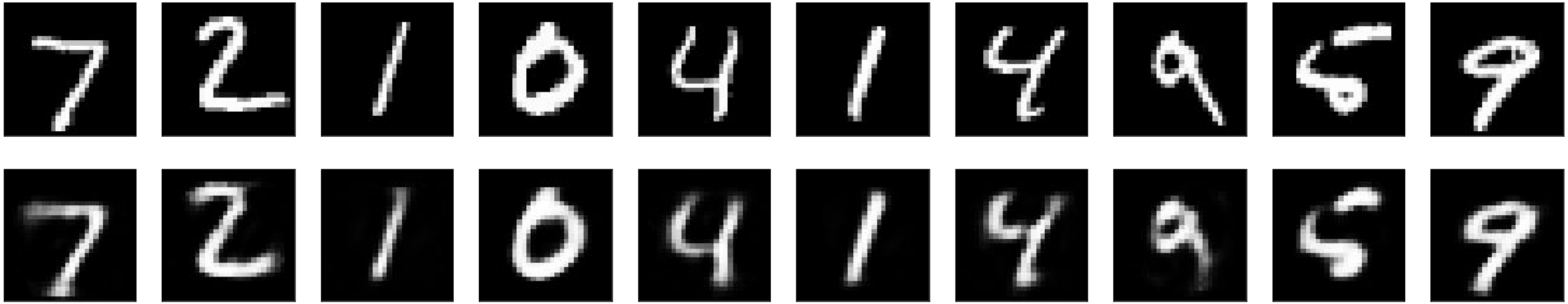
$$x$$



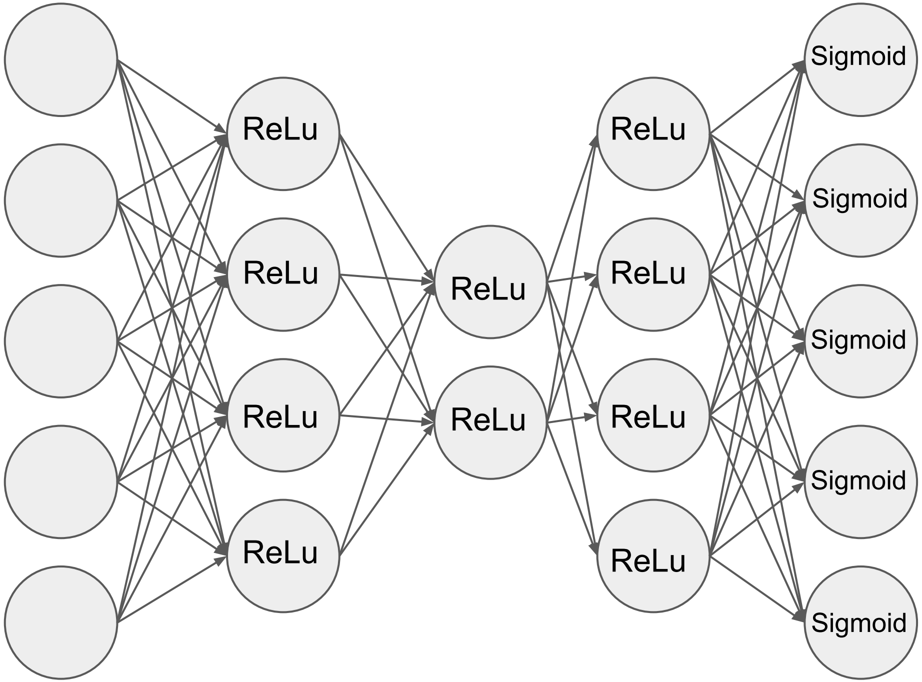
$$\hat{x} = sigmoid(Wz)$$

$$z = relu(Wx)$$

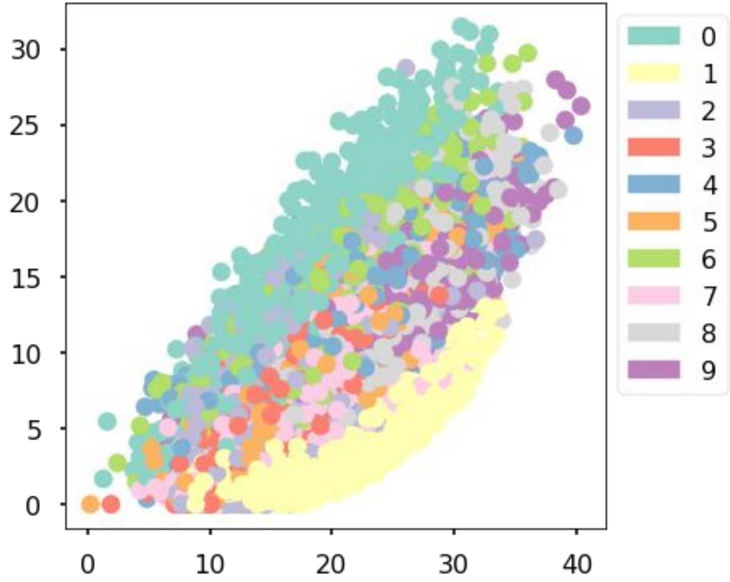
Nonlinear autoencoder with 32 hidden neurons

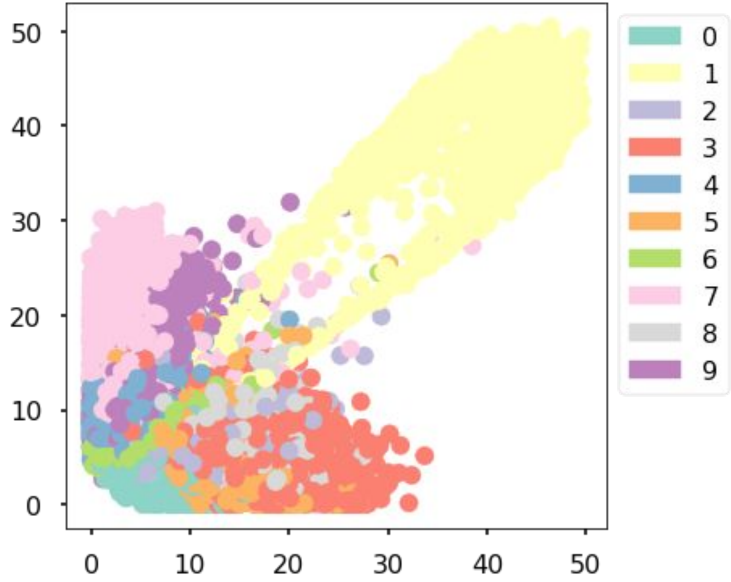


Autoencoders can be deep



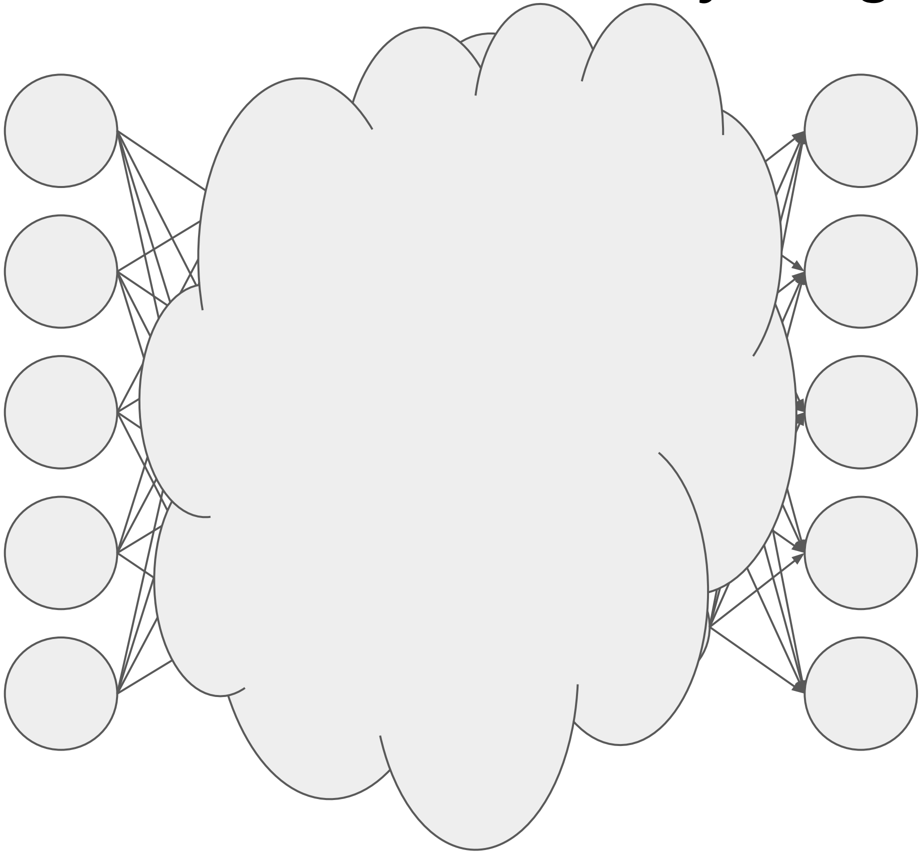
Deep autoencoder (bottleneck of 2)





Guess which one is deep (has intermediate layer)?

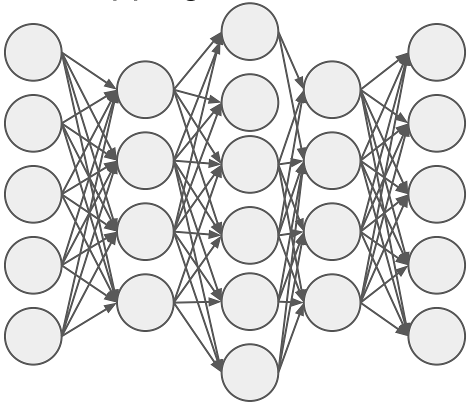
Autoencoders can be almost *anything*

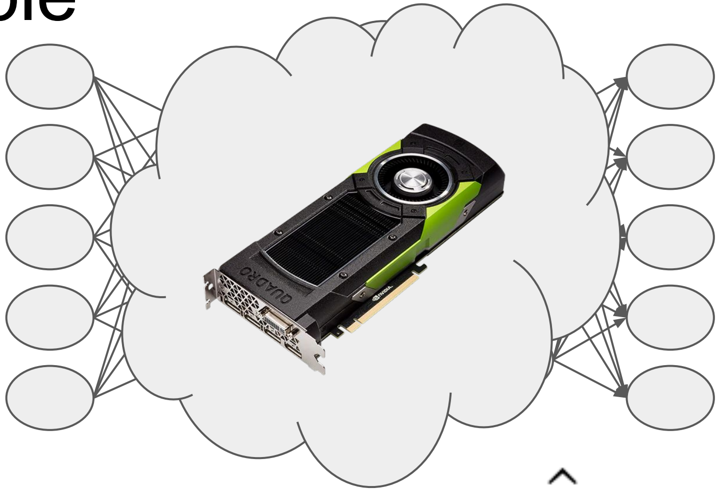


Deep nets can be very flexible

Great power to do stupid things

Index mapping





$$\hat{x} = Ix$$

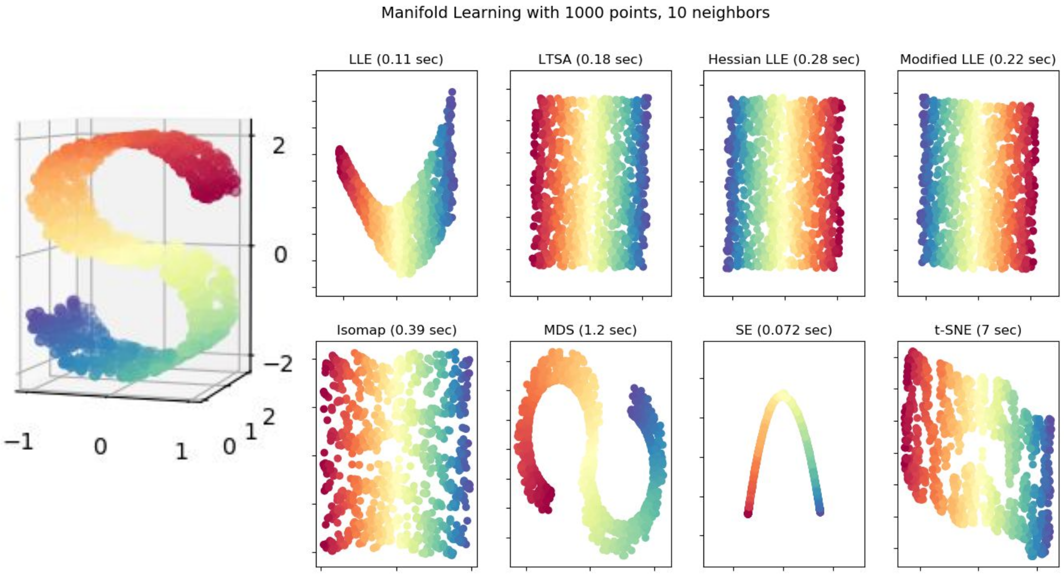
A GPU powered identity function

Basically, overfitting

Manifolds

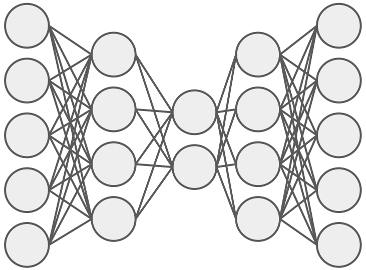
“Locally euclidean subspaces”

- Is your 3D data really on a 2D surface (manifold)
- Is your 10,000D data really on a 100D manifold?



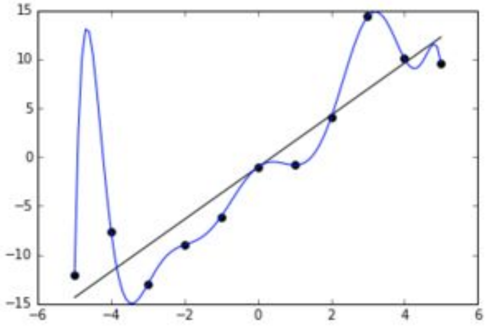
From the scikit-learn documentation

Manifolds and latent spaces



$$\underset{W}{\operatorname{argmin}} \|x - \hat{x}\|^2$$

Think overfitting - manifolds are how it generalizes well



Better generalization?

- Try regularization!

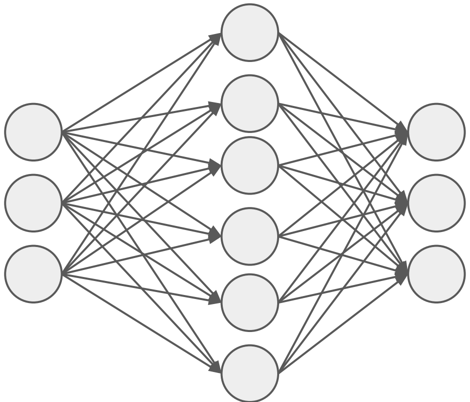
How to learn manifolds?

regularization

- bottleneck is one way to regularize

- L1 (laplacian) another way to regularize (this is called Sparse Autoencoders)

$$\underset{W}{\operatorname{argmin}} ||x - \hat{x}||^2 + \lambda ||W||_1$$



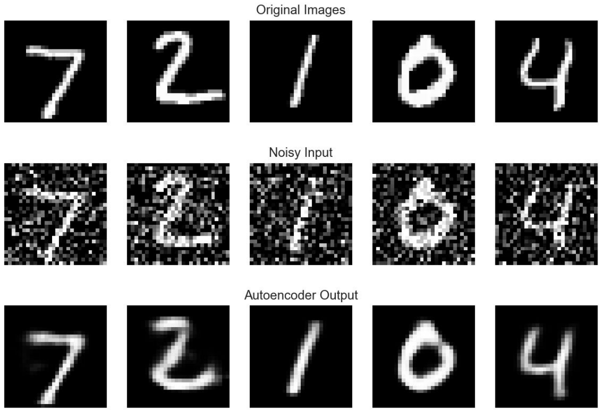
Denoising Autoencoders (DAE)

We want:

- Similar in *input space* => similar in *latent space* => similar *reconstruction*

Idea:

- Add noise to *input*, but not to *reconstruction target*



$$\underset{W}{\operatorname{argmin}} ||x - \hat{x}||^2$$

Contractive Autoencoders (CAE)

We want:

- Similar in *input space* => similar *latent space*

Idea:

- Add a penalty for the sensitivity of the latent space to perturbations in the input

$$\underset{W}{\operatorname{argmin}} \|\hat{x} - x\|^2 + \lambda \|J_z(x)\|_F^2$$

$$J_z(x) = \left[\frac{\partial z_i}{\partial x_j} \right]$$

CAE == DAE

Journal of Machine Learning Research 15 (2014) 3743-3773 Submitted 6/13; Published 11/14

What Regularized Auto-Encoders Learn from the Data-Generating Distribution

Guillaume Alain GUILLAUME.ALAIN@UMONTREAL.CA
Yoshua Bengio YOSHUA.BENGIO@UMONTREAL.CA
Department of Computer Science and Operations Research
University of Montreal
Montreal, H3C 3J7, Quebec, Canada

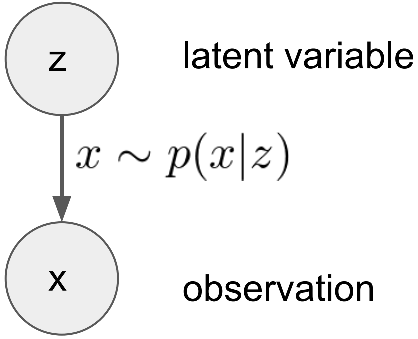
Variational autoencoders

Variational Inference

+

autoencoders

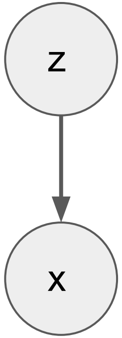
Generative model:



The inference problem:

$$p(z|x)$$

Variational Inference (quick overview)

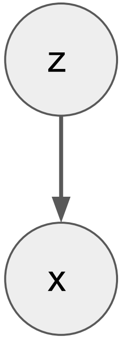


latent variable

observation

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

Variational Inference (quick overview)



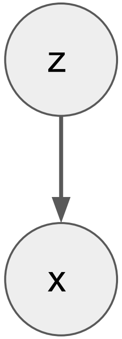
latent variable

observation

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

problematic...

Variational Inference (quick overview)



latent variable

observation

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

← problematic...

Variational Inference Solution:

$$p(z|x) \approx q(z|x)$$

← Chosen to be a distribution we can work with

Side note on $p(z|x) \approx q(z|x)$

- Information
  - "How many bits do we **need** to represent event *x* if we optimized for *p(x)*?"

$$I = -log\, p(x)$$

- Entropy
  - "What is the **expected amount of information** in each event drawn from *p(x)*?" (how many bits?)

$$H = -\sum p(x) log\, p(x)$$

- Cross-entropy
  - "What is the **expected amount of information** in *p(x)* if we **optimized for *q(x)***?" (how many bits?)

$$H\big(p(x), q(x)\big) = -\sum p(x) log\, q(x)$$

- **K**ullback-**L**eibler divergence: "cross-entropy - entropy"
  - "How many **more bits** will we **need** to represent events from *p(x)* if we optimized for *q(x)*?

$$D_{KL}\big(p(x)||q(x)\big) = -\sum p(x) log\, \frac{q(x)}{p(x)}$$

Variational Inference (quick overview)
skipping the math...

$$min\, D_{KL}\left(q_\theta(z|x)\;||\;p_\phi(z|x)\right)$$

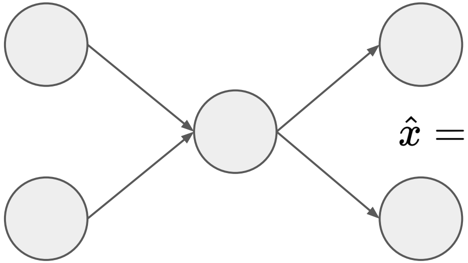
Maximizing the Evidence LOwer Bound (ELBO)

$$\mathcal{L} = \mathbf{E}_{z\sim q(z|x)}\left[log(p_\phi(x|z)\right] - D_{KL}\left(q_\theta(z|x)\;||\;p_\phi(z)\right)$$

Variational inference is methods to maximize ELBO

How does it fit in with autoencoders?

What if autoencoders were **probabilistic**?

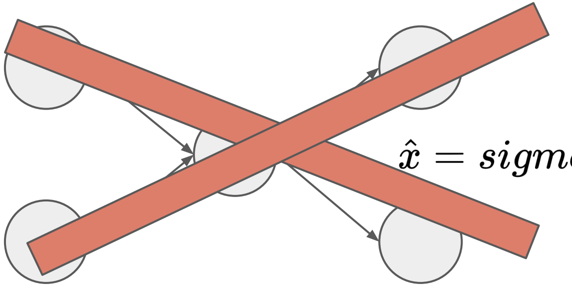


$$\hat{x} = sigmoid(Wz)$$

$$z = relu(Wx)$$

What if autoencoders were **probabilistic**?

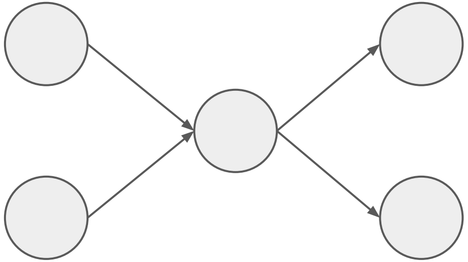
Regular autoencoder



$$\hat{x} = sigmoid(Wz)$$

$$z = relu(Wx)$$

Variational autoencoder

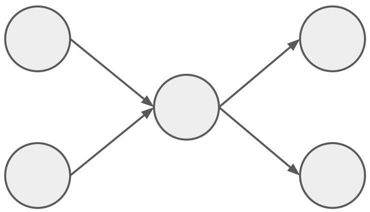


$$\hat{x} \sim p_\phi(x|z)$$

$$z \sim q_\theta(z|x)$$

Variational Autoencoder loss - negative ELBO

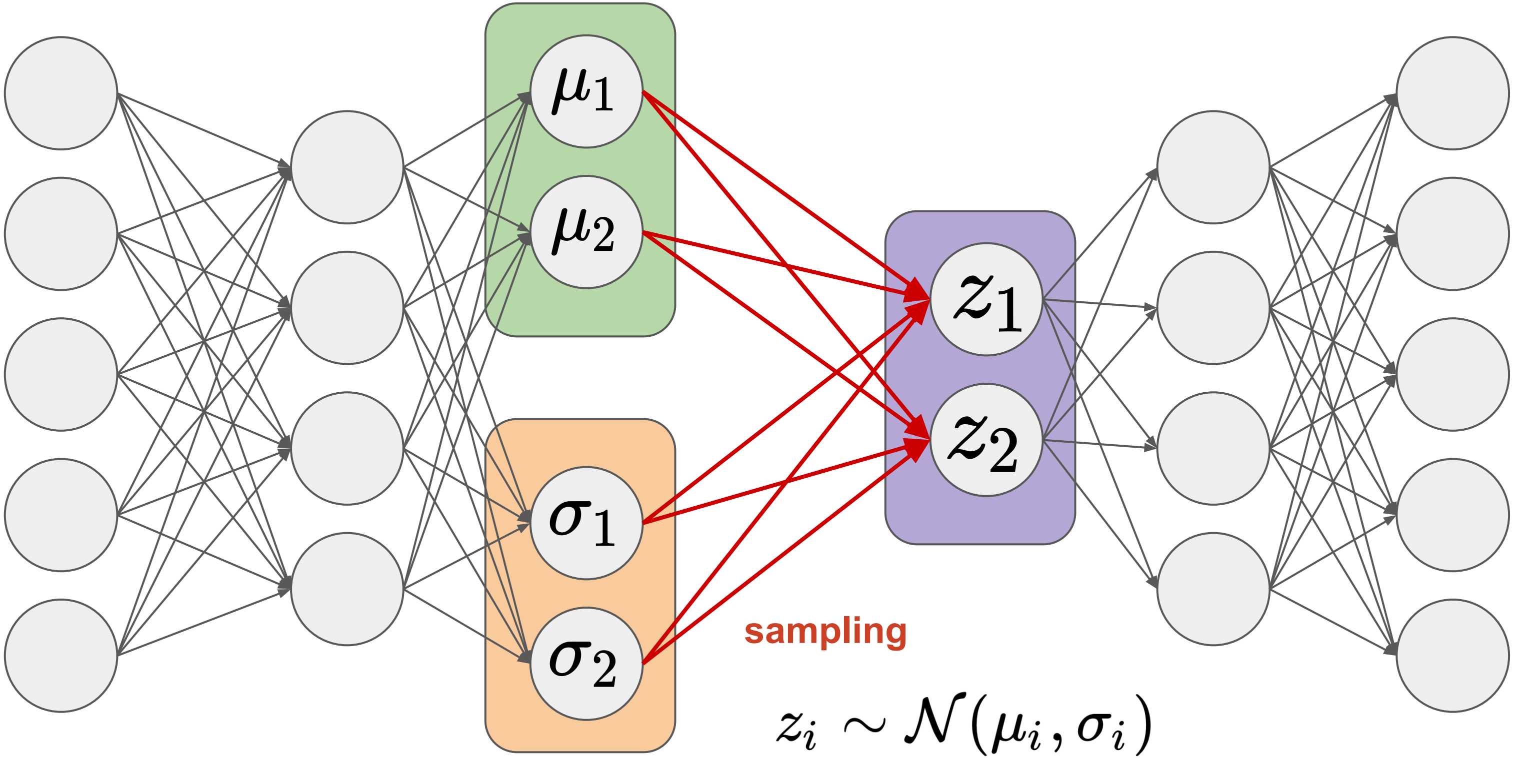
$$l = -\underbrace{\mathbf{E}_{z\sim q(z|x)}\left[log(p_\phi(x|z)\right]}_{\text{reconstruction error}} + \underbrace{D_{KL}\left(q_\theta(z|x)\,||\,p_\phi(z)\right)}_{\text{divergence from prior}}$$



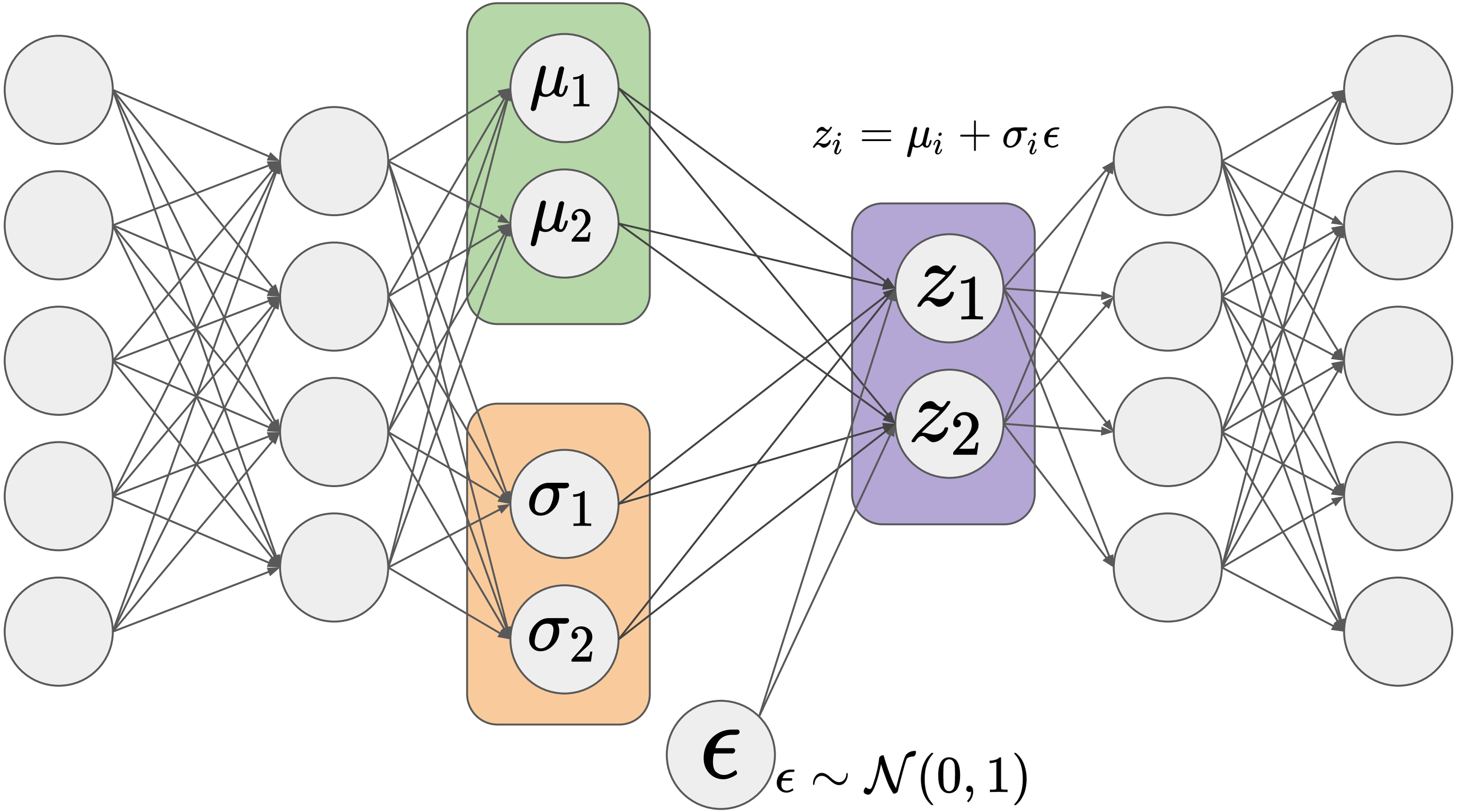
$\hat{x} \sim p_\phi(x|z)$

$z \sim q_\theta(z|x)$

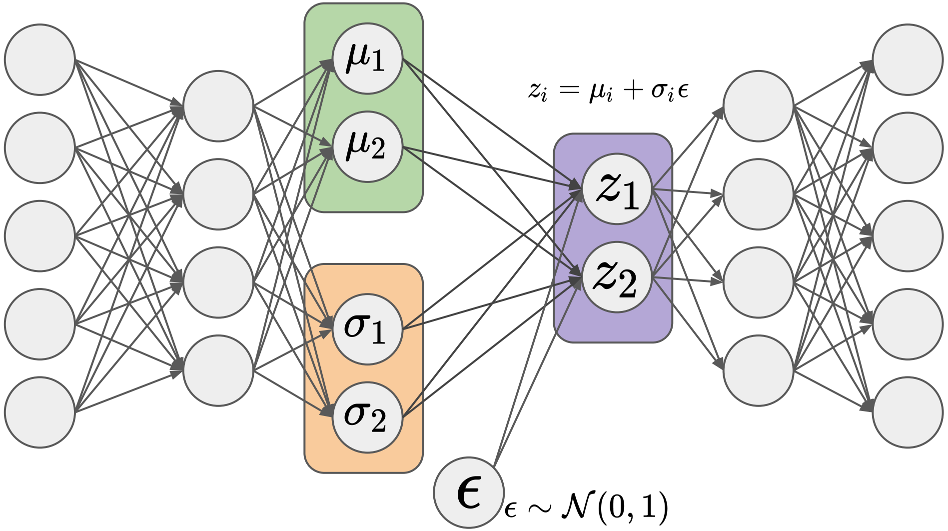
Backpropagation through VAEs



Backpropagation through VAEs - reparameterizing

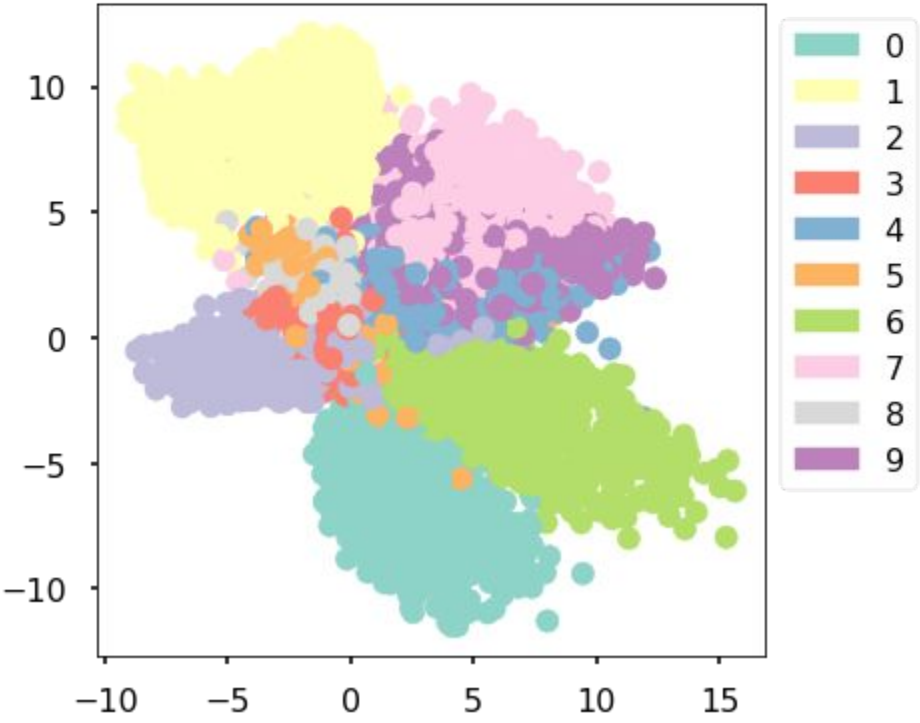


Backpropagation through VAEs - reparameterizing

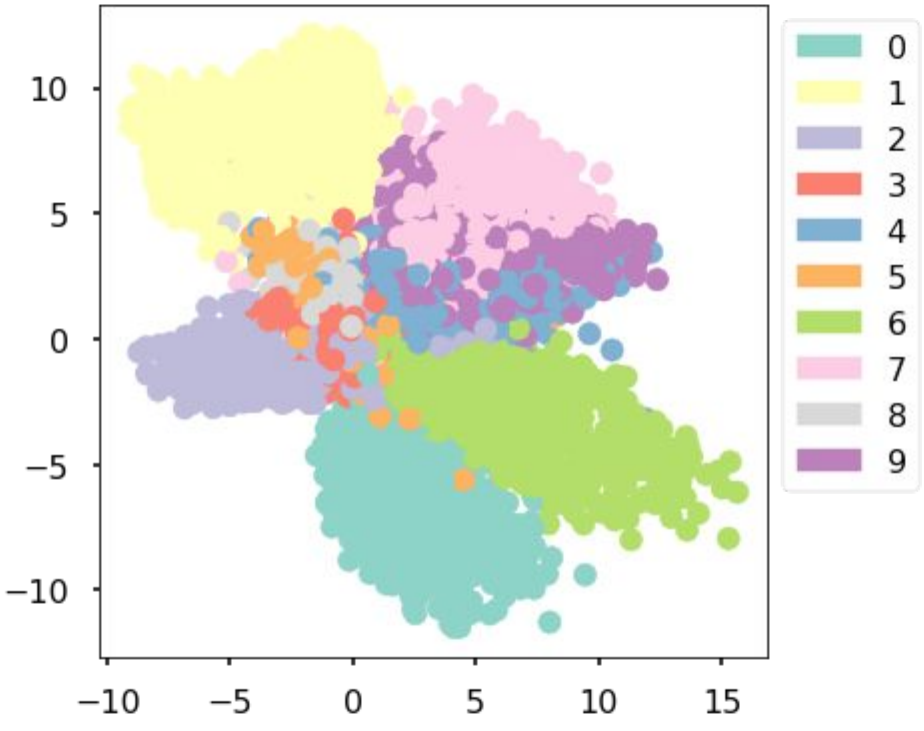


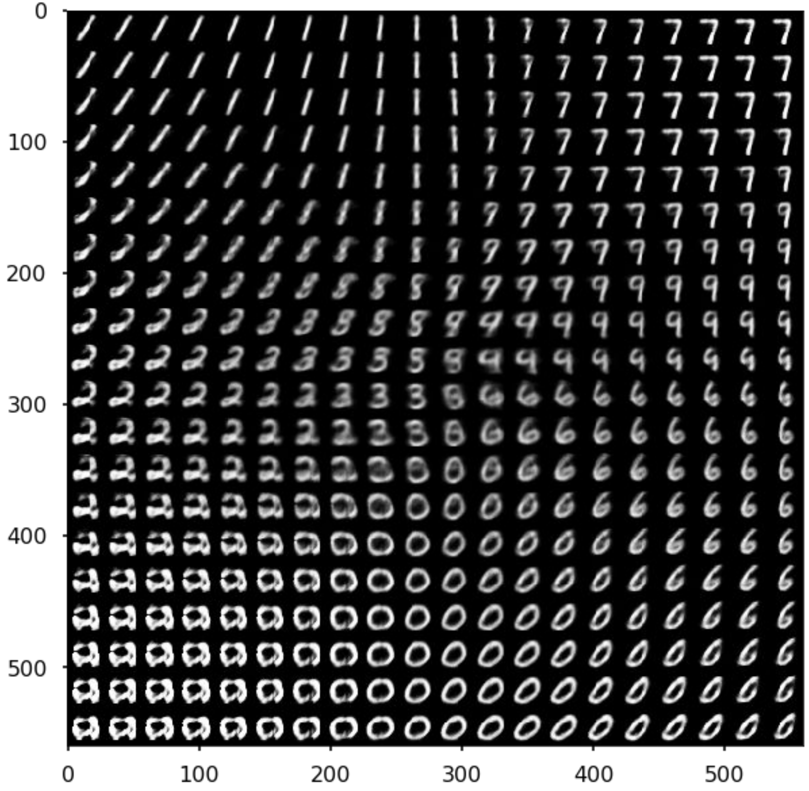
$$l = -\underbrace{\mathbf{E}_{z \sim q(z|x)}\left[log(p_\phi(x|z)\right]}_{\text{reconstruction error}} + \underbrace{D_{KL}\left(q_\theta(z|x) \,||\, p_\phi(z)\right)}_{\text{divergence from prior}}$$

VAE 2d embedding

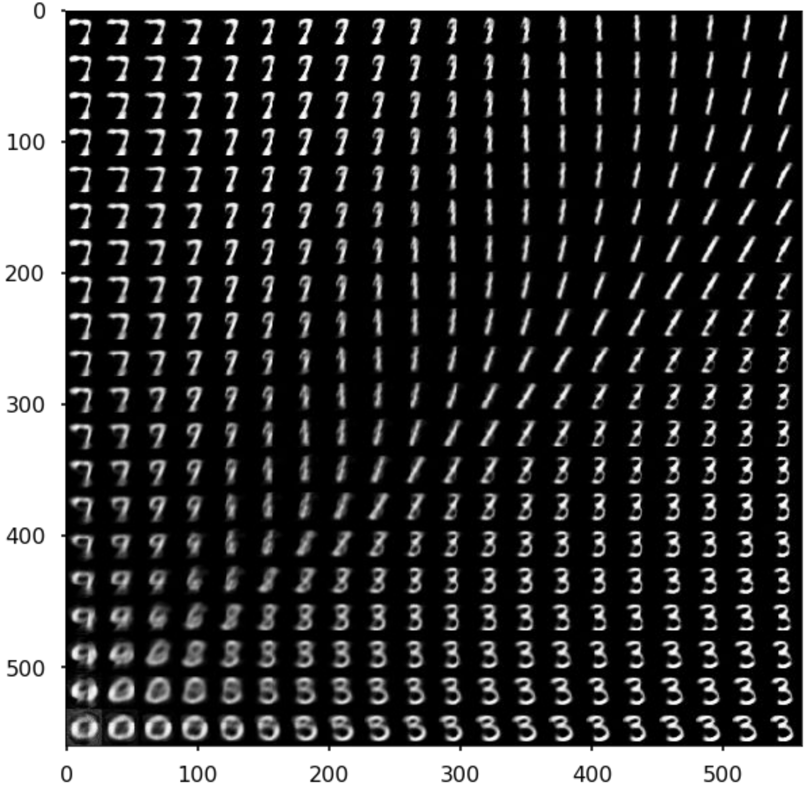


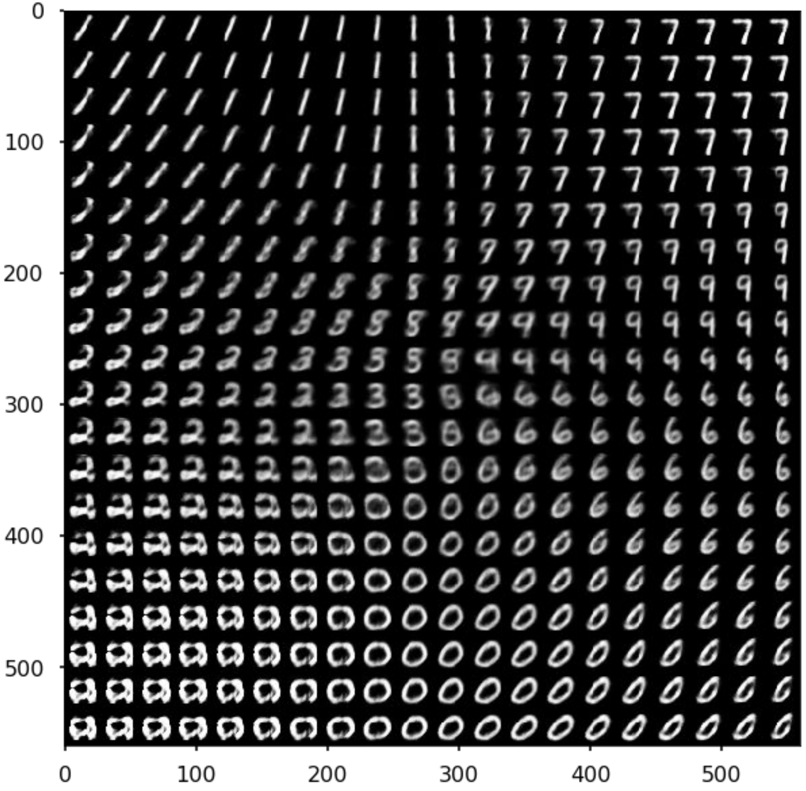
VAEs are a generative model



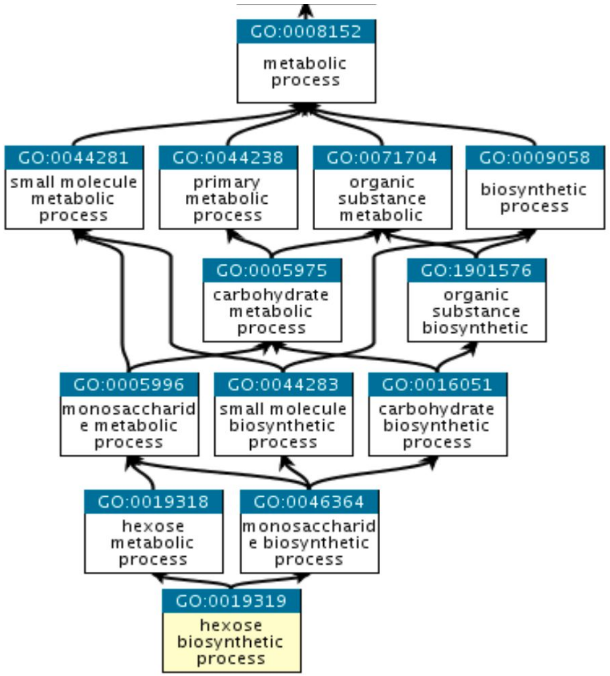


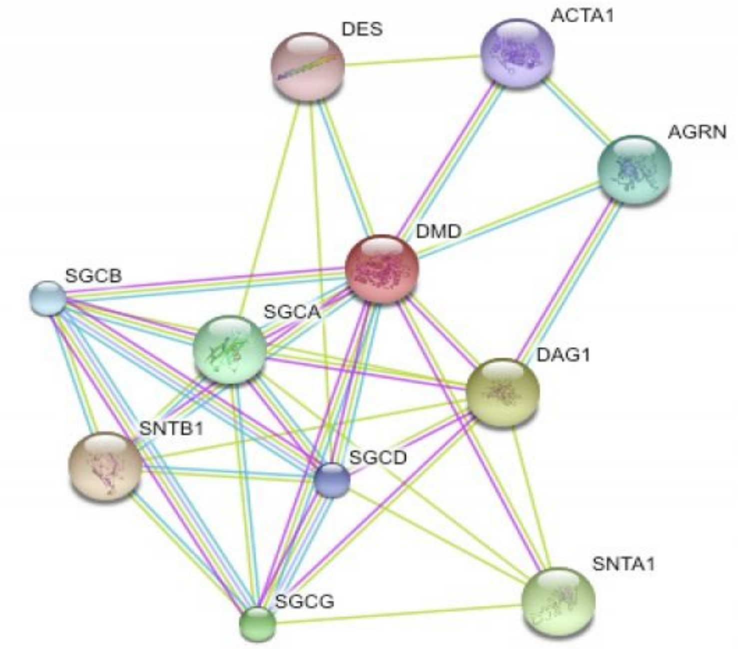
Regular autoencoder as a generative model?



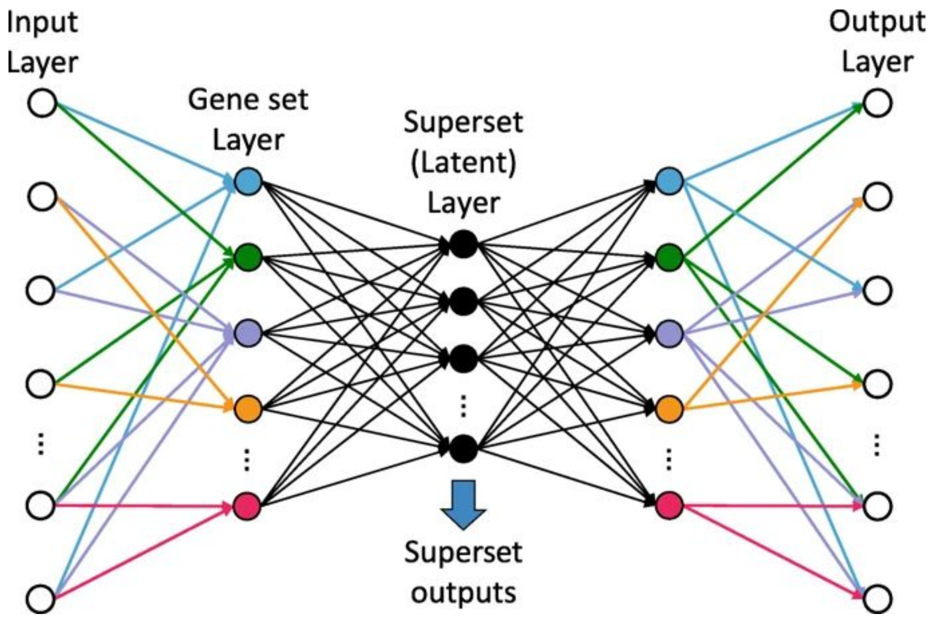


Biologically regularized autoencoders



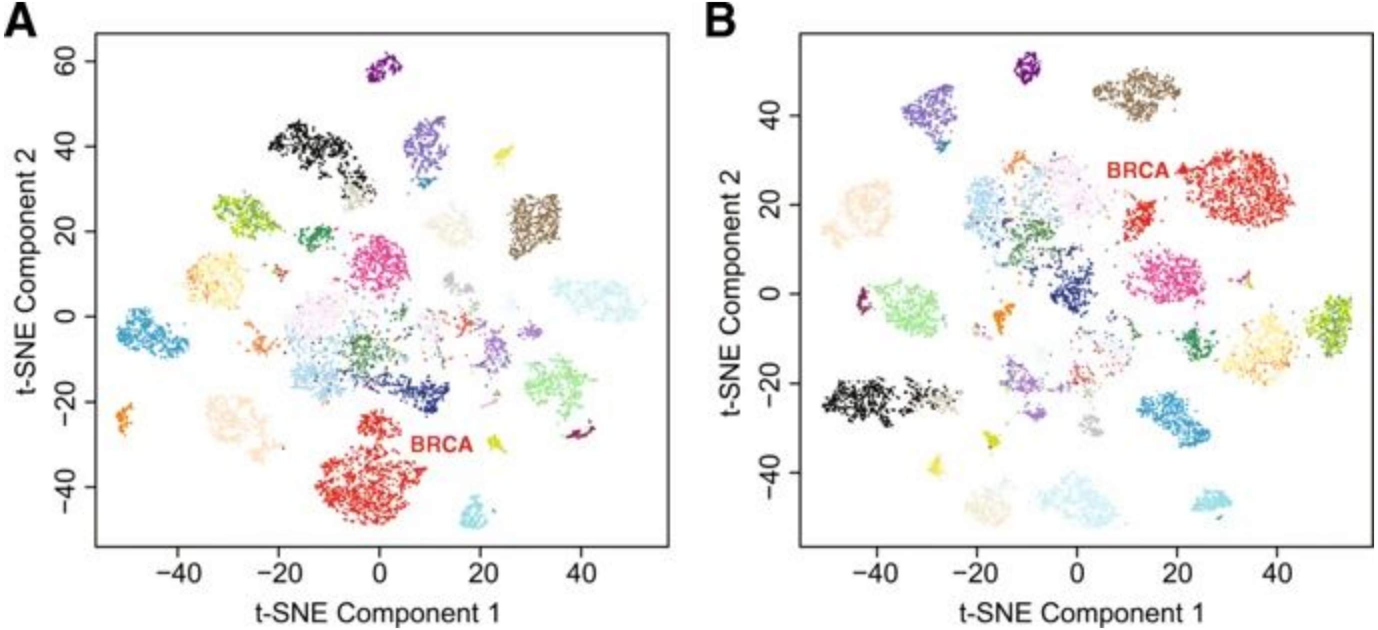


Gene Superset Autoencoder (GSEA)



Chen, H.H., Chiu, Y., Zhang, T. *et al.* GSAE: an autoencoder with embedded gene-set nodes for genomics functional characterization. *BMC Syst Biol* **12,** 142 (2018)

Gene Superset Autoencoder (GSEA)



Chen, H.H., Chiu, Y., Zhang, T. *et al.* GSAE: an autoencoder with embedded gene-set nodes for genomics functional characterization. *BMC Syst Biol* **12,** 142 (2018)

Questions?

Hands-on session

- Jupyter notebook
- Omics data from TCGA colorectal cancers
- Implement autoencoders in TensorFlow 2.0 keras
- The notebook has a demo and exercises
- Pick the exercises that you want to work on, we’ll help you out

Further reading

- Autoencoders: http://www.deeplearningbook.org/contents/autoencoders.html
- GSEA: https://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-018-0642-2
- VAE: https://arxiv.org/abs/1312.6114
  - https://www.jeremyjordan.me/variational-autoencoders/
  - https://www.youtube.com/watch?v=uaaqyVS9-rM
- CAE: https://arxiv.org/abs/1305.4076
  - http://jmlr.csail.mit.edu/papers/volume15/alain14a/alain14a.pdf
- VAE for cancer: https://www.life-science-alliance.org/content/2/6/e201900517