

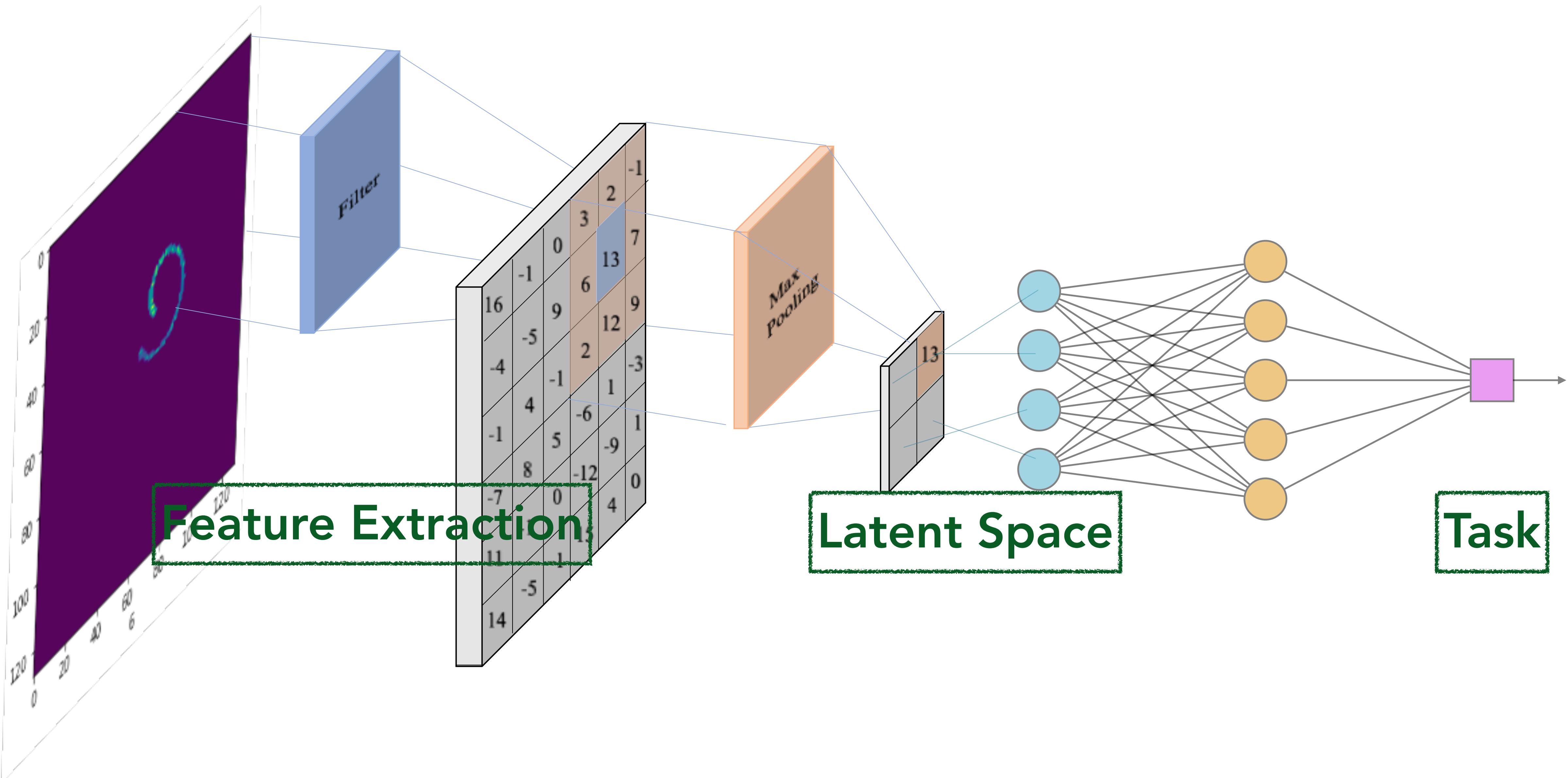
# UNSUPERVISED AND SELF-SUPERVISED LEARNING

---

MICHELLE KUCHERA  
DAVIDSON COLLEGE

CPS-FR  
MIT  
22 AUGUST 2024

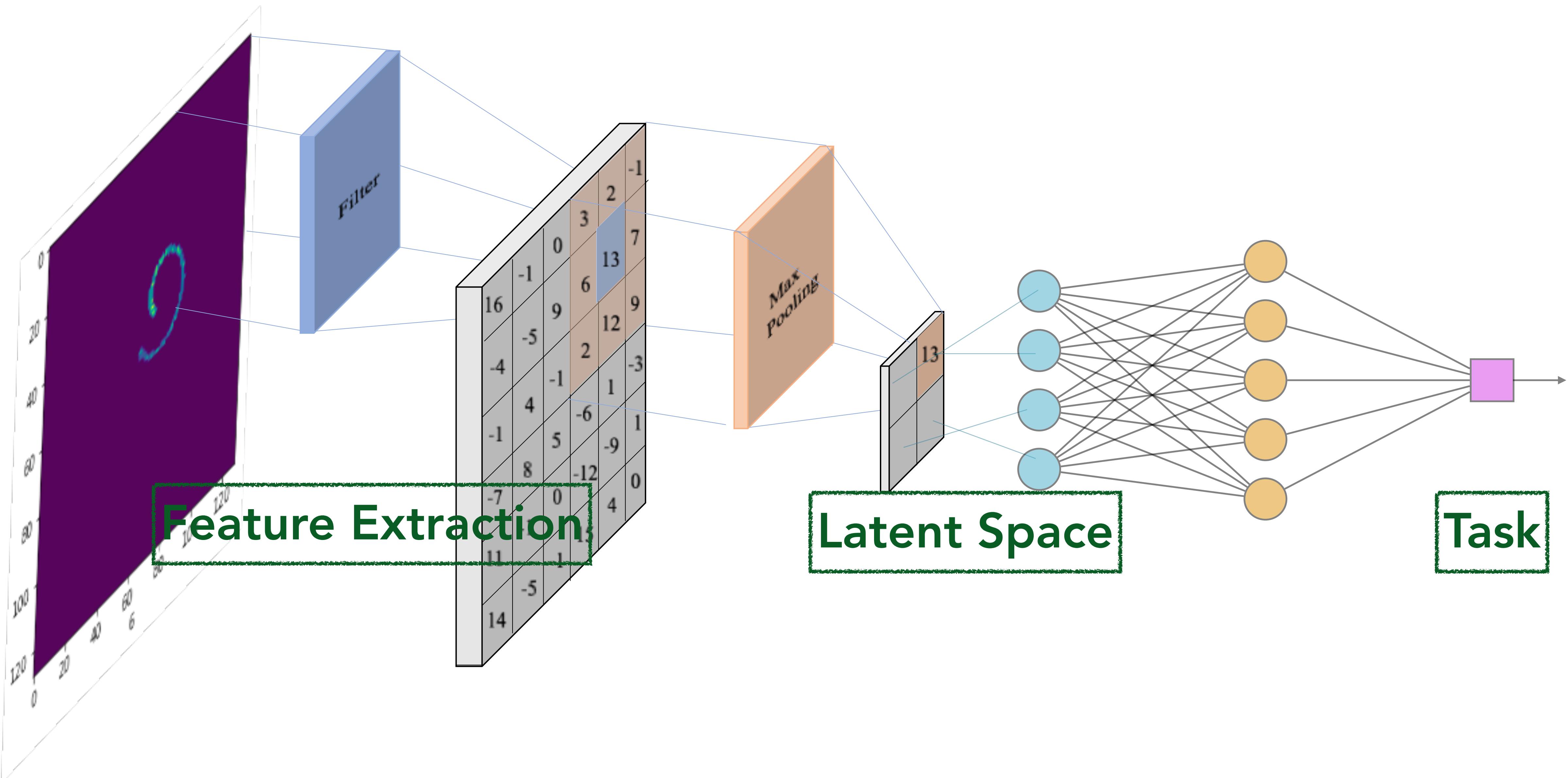
# CONVOLUTIONAL NEURAL NETWORKS



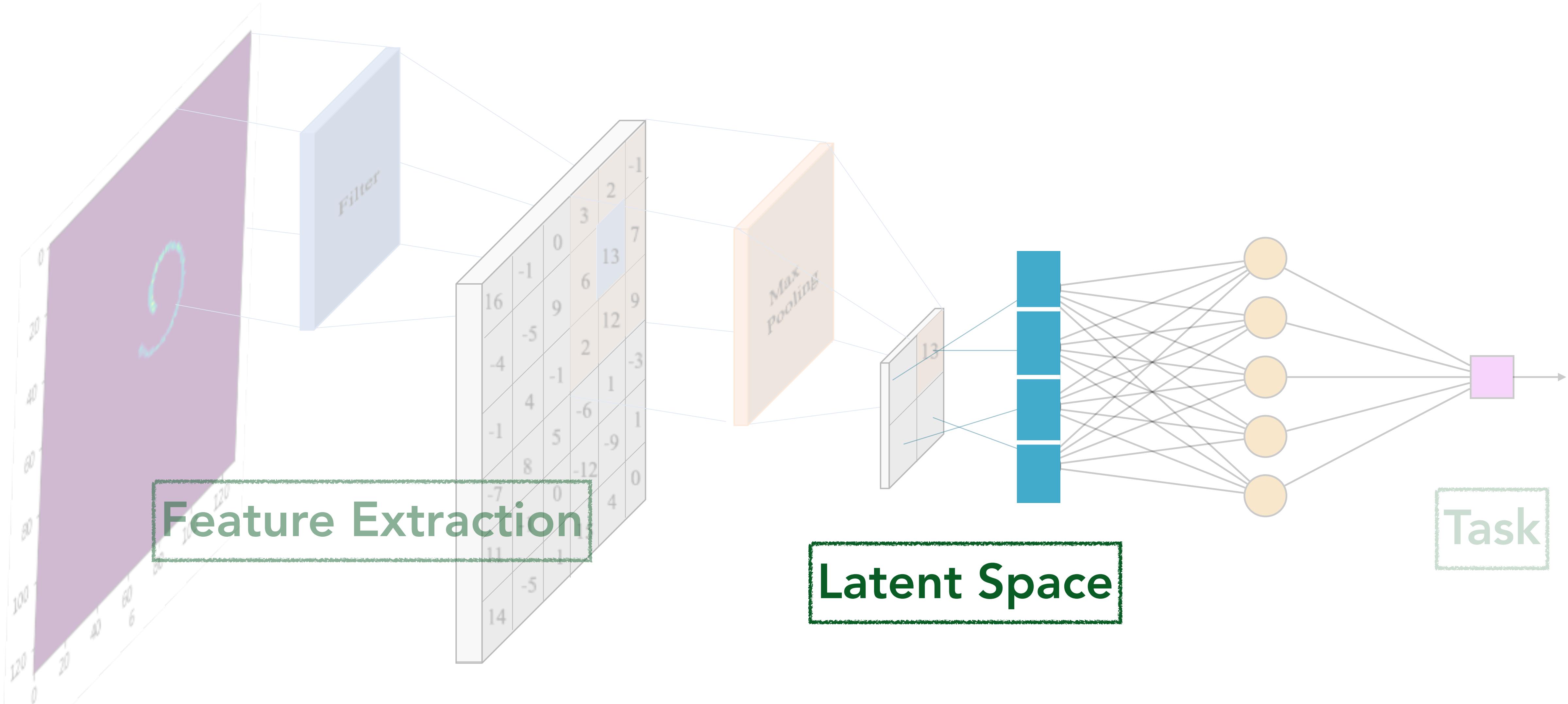
# PRETRAINED MODELS



# CONVOLUTIONAL NEURAL NETWORKS



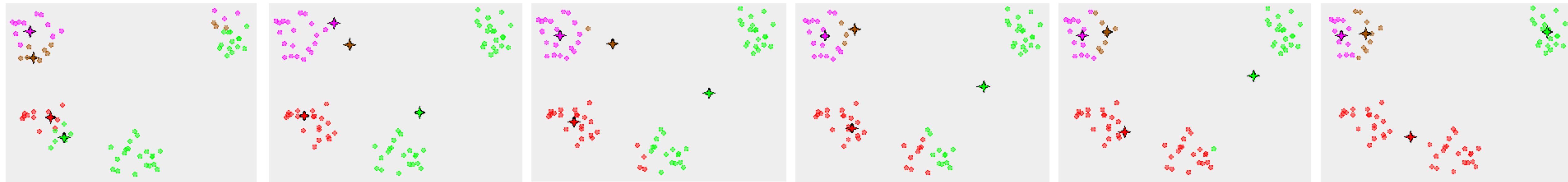
# CONVOLUTIONAL NEURAL NETWORKS



# CLUSTERING — KMEANS

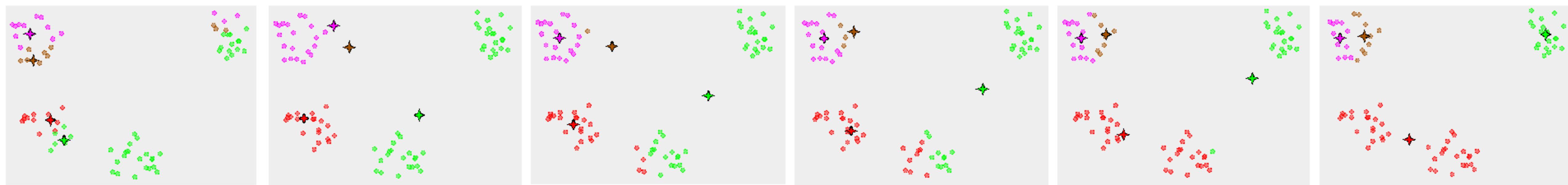
Goal: minimize pairwise distances between points in same cluster

$$\min \sum_{i=1}^k \frac{1}{2N} \sum_{x,y,x \neq y} (\vec{x} - \vec{y})^2$$

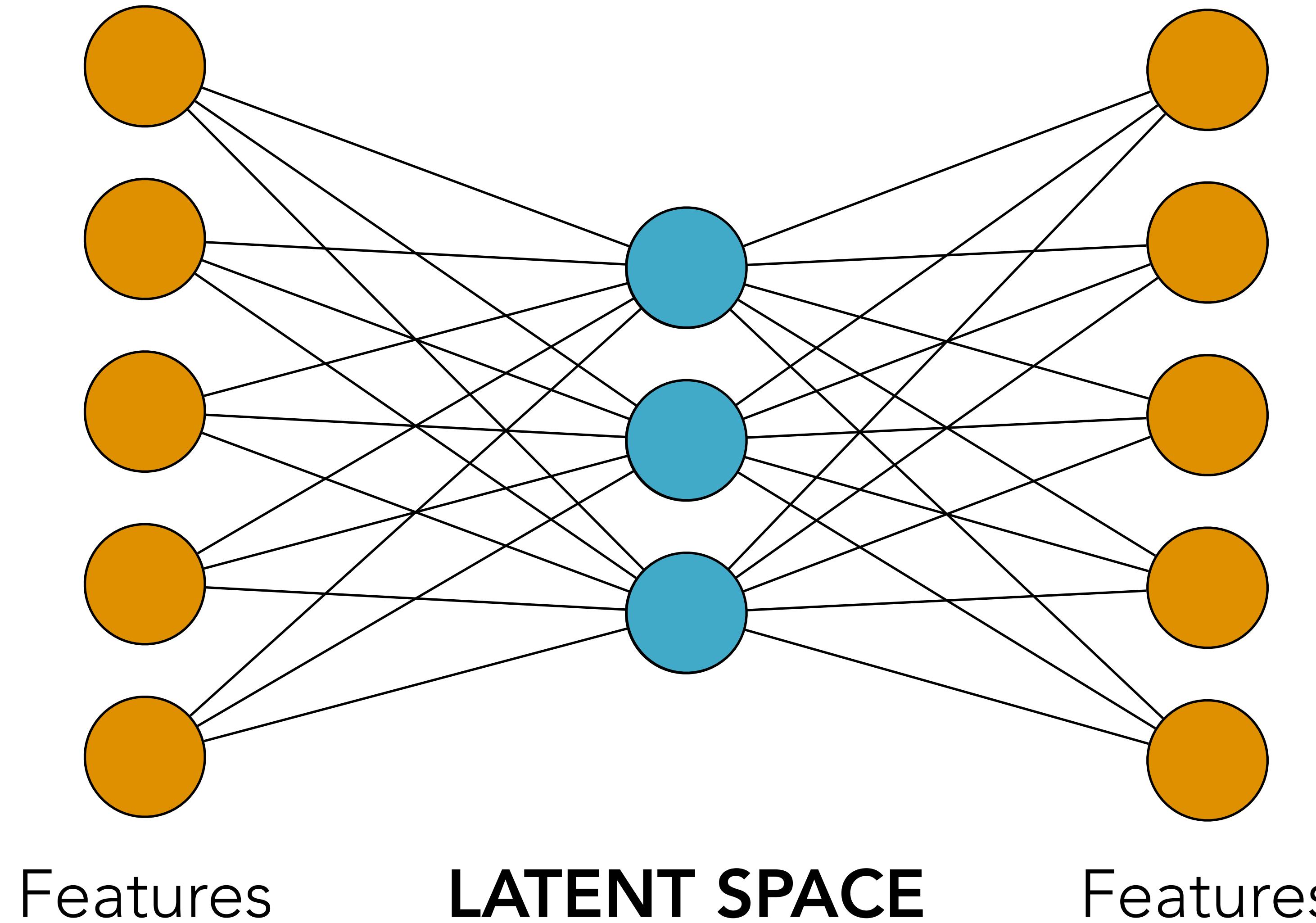


Goal: maximize pairwise distances between points in *different* clusters

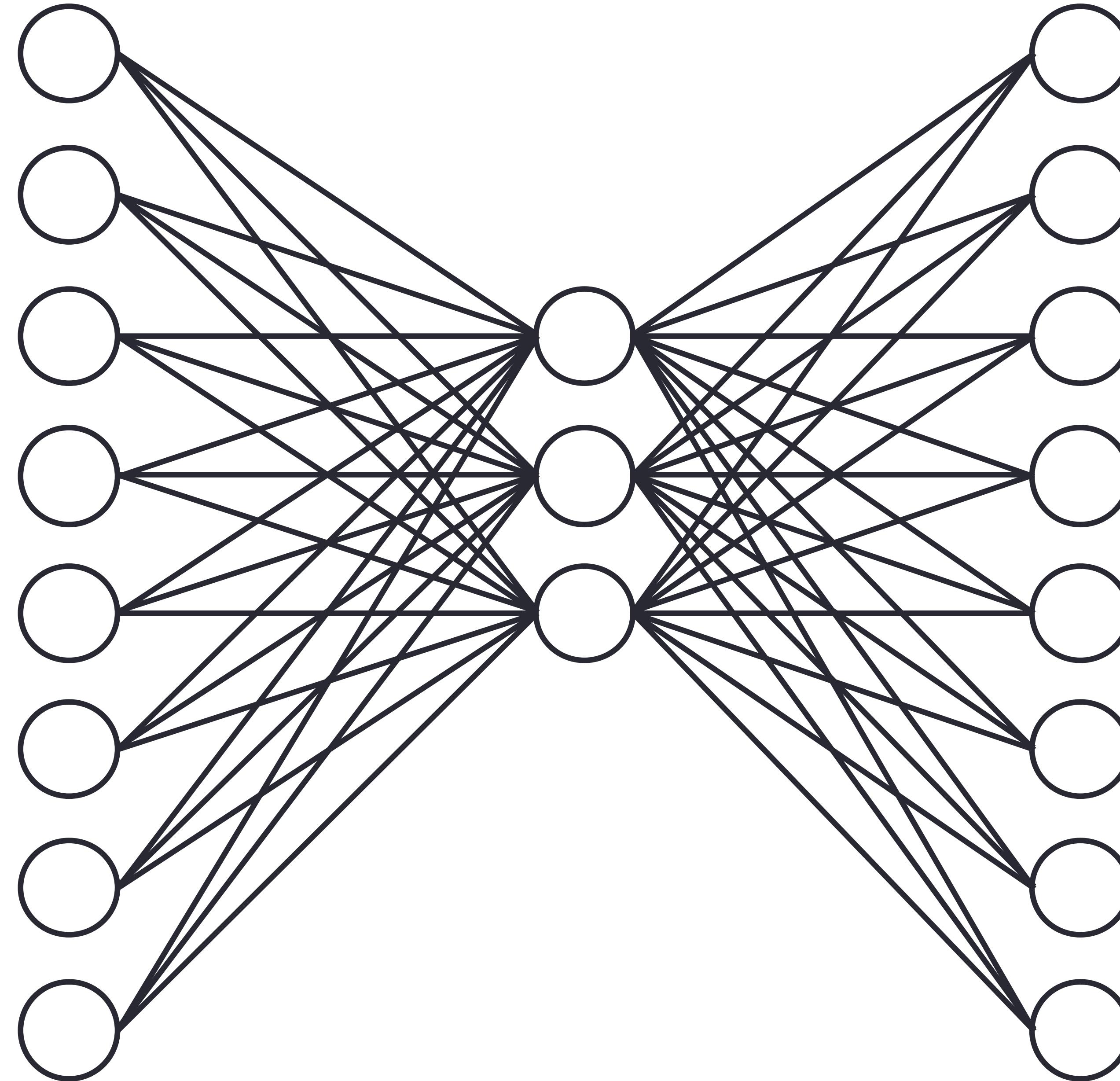
# CLUSTERING — KMEANS



# AUTOENCODER

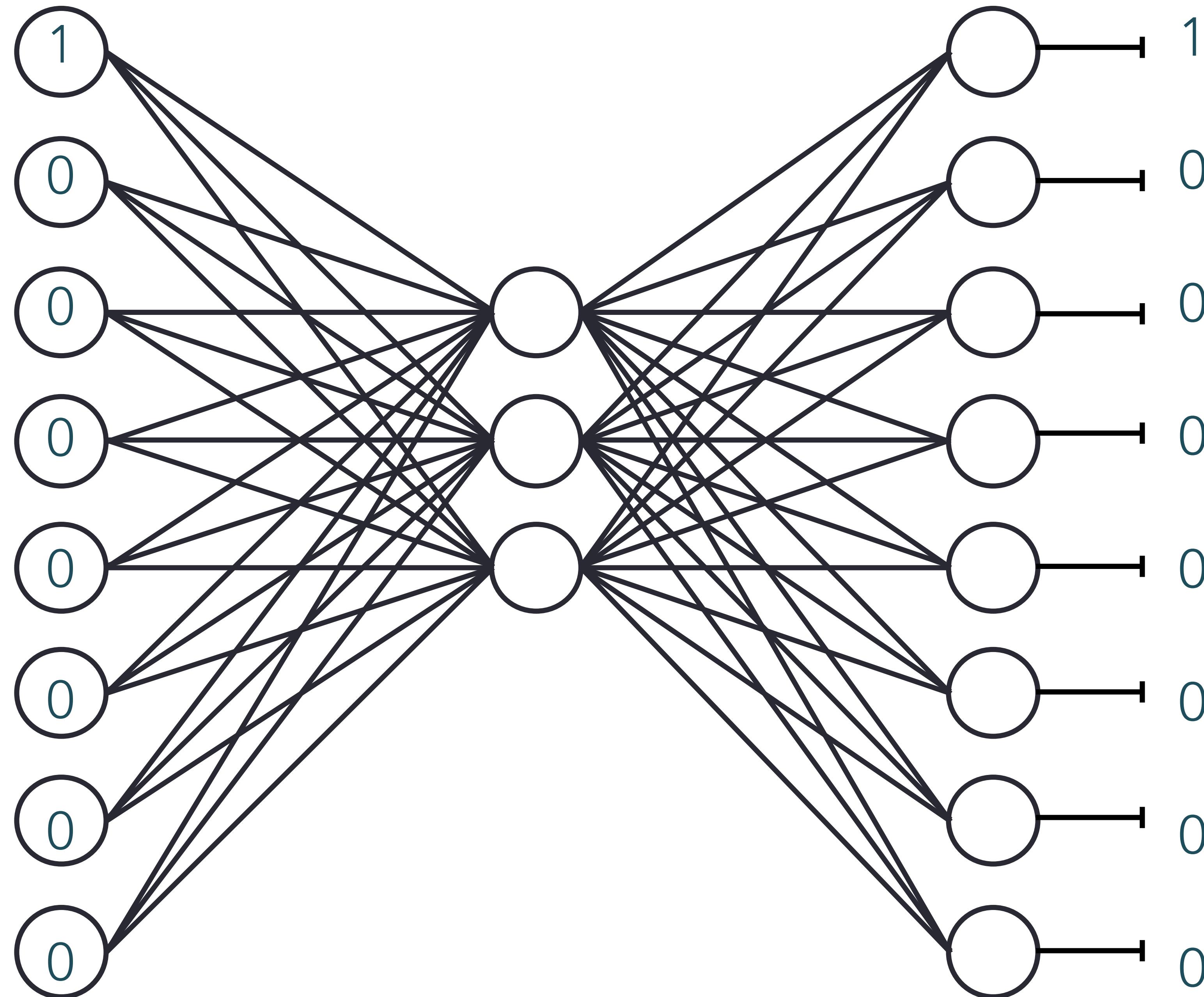


# AUTOENCODER ARCHITECTURE



# SELF-SUPERVISED LEARNING

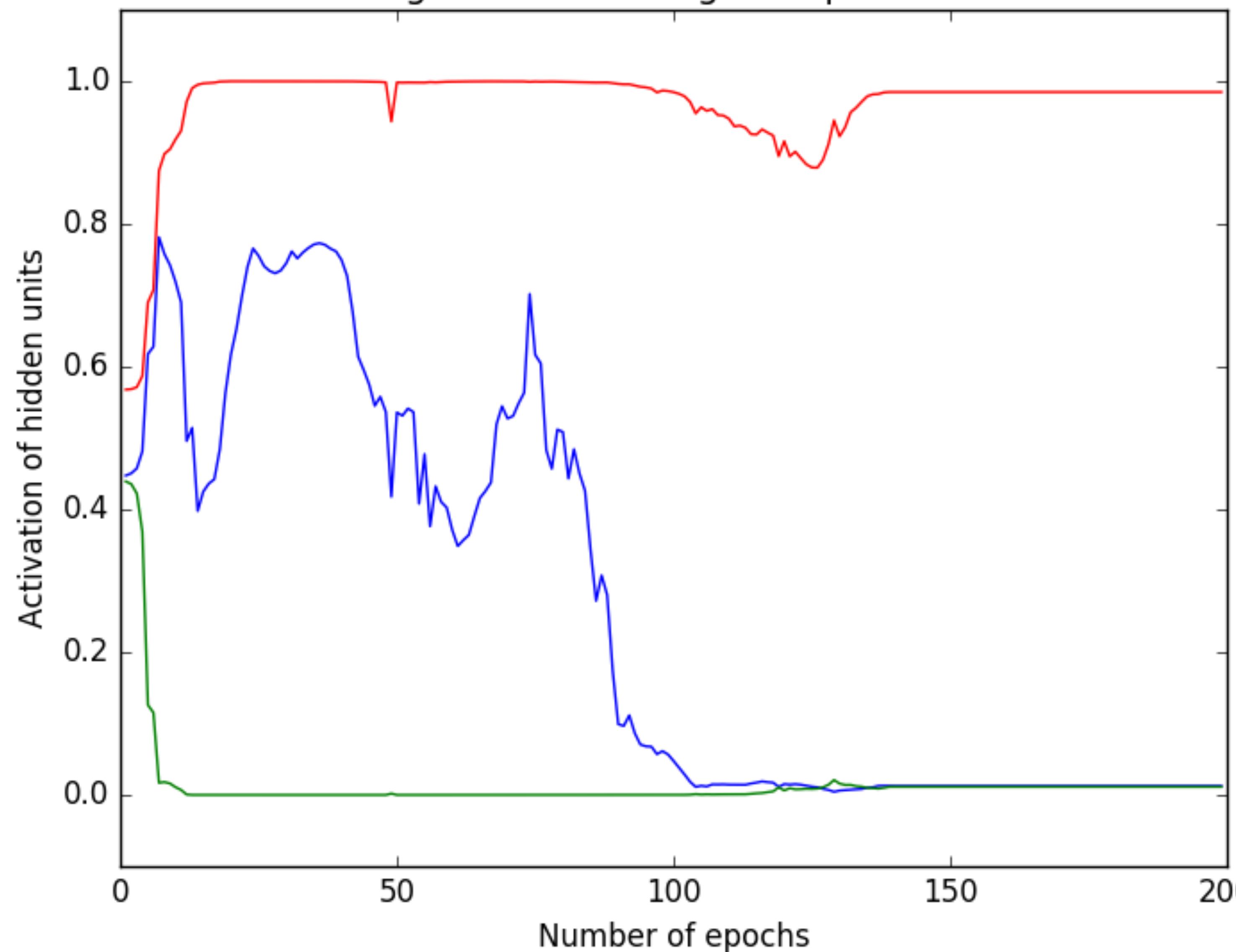
Input	Output
10000000	10000000
01000000	01000000
00100000	00100000
00010000	00010000
00001000	00001000
00000100	00000100



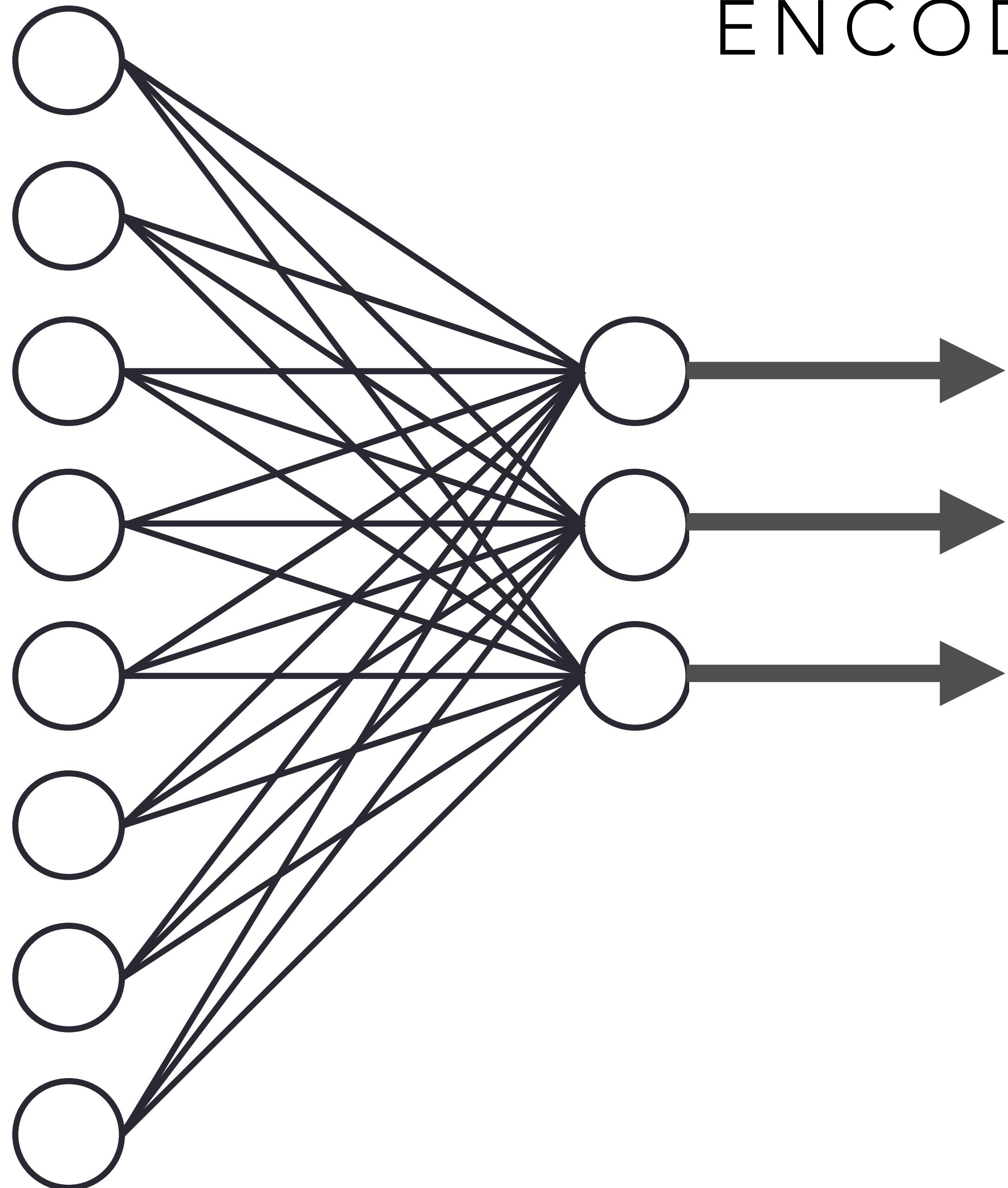
Input	Output
10000000	10000000
01000000	01000000
00100000	00100000
00010000	00010000
00001000	00001000
00000100	00000100

Input	A1	A2	A3	Output
10000000	0.9911	0.9869	0.0093	10000000
01000000	0.9892	0.0095	0.0124	01000000
00100000	0.0094	0.0283	0.0122	00100000
00010000	0.9840	0.9836	0.9900	00010000
00001000	0.0139	0.9904	0.0186	00001000
00000100	0.0128	0.9805	0.9868	00000100

Learning of the encoding for input 00000010

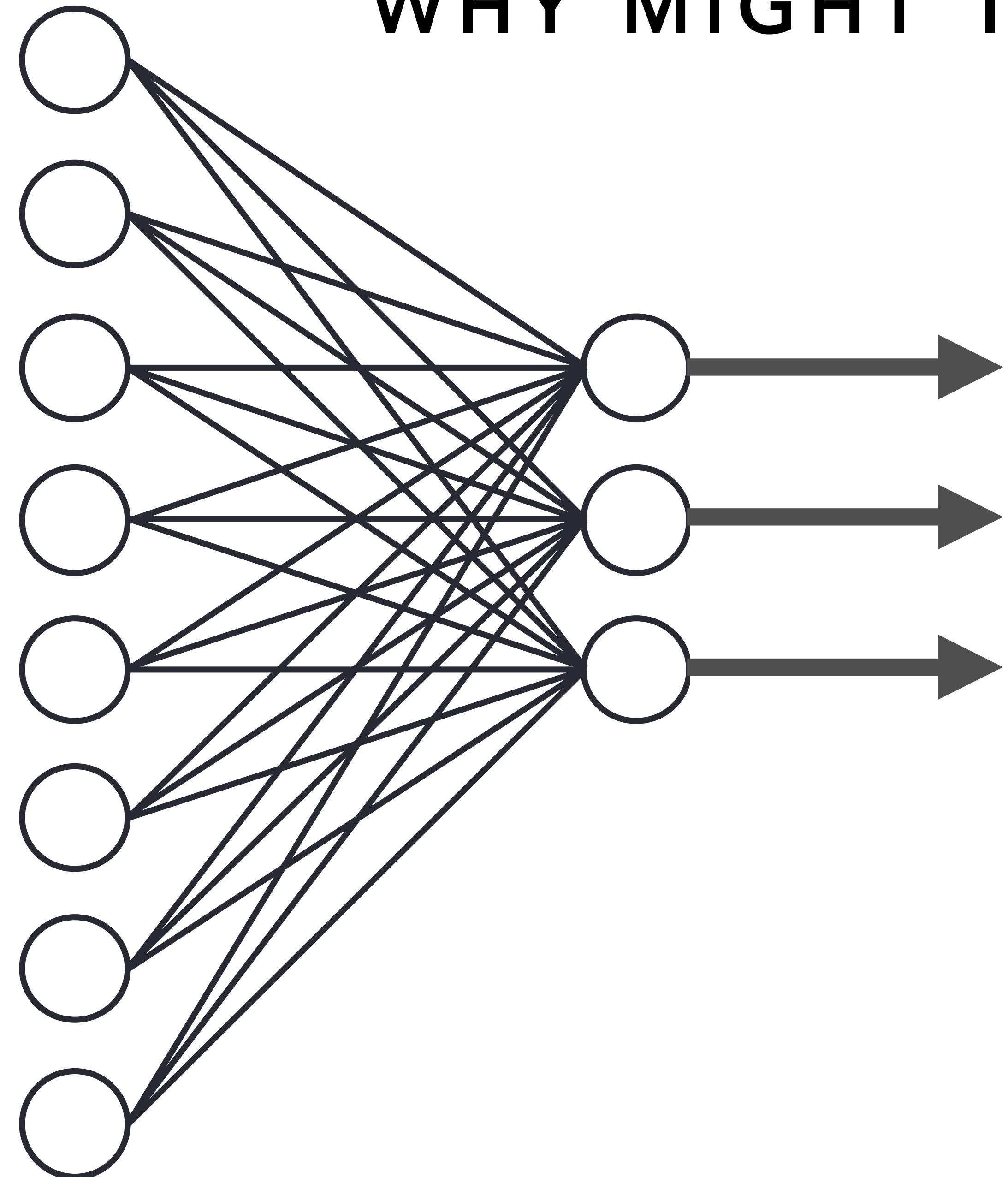


# ENCODER

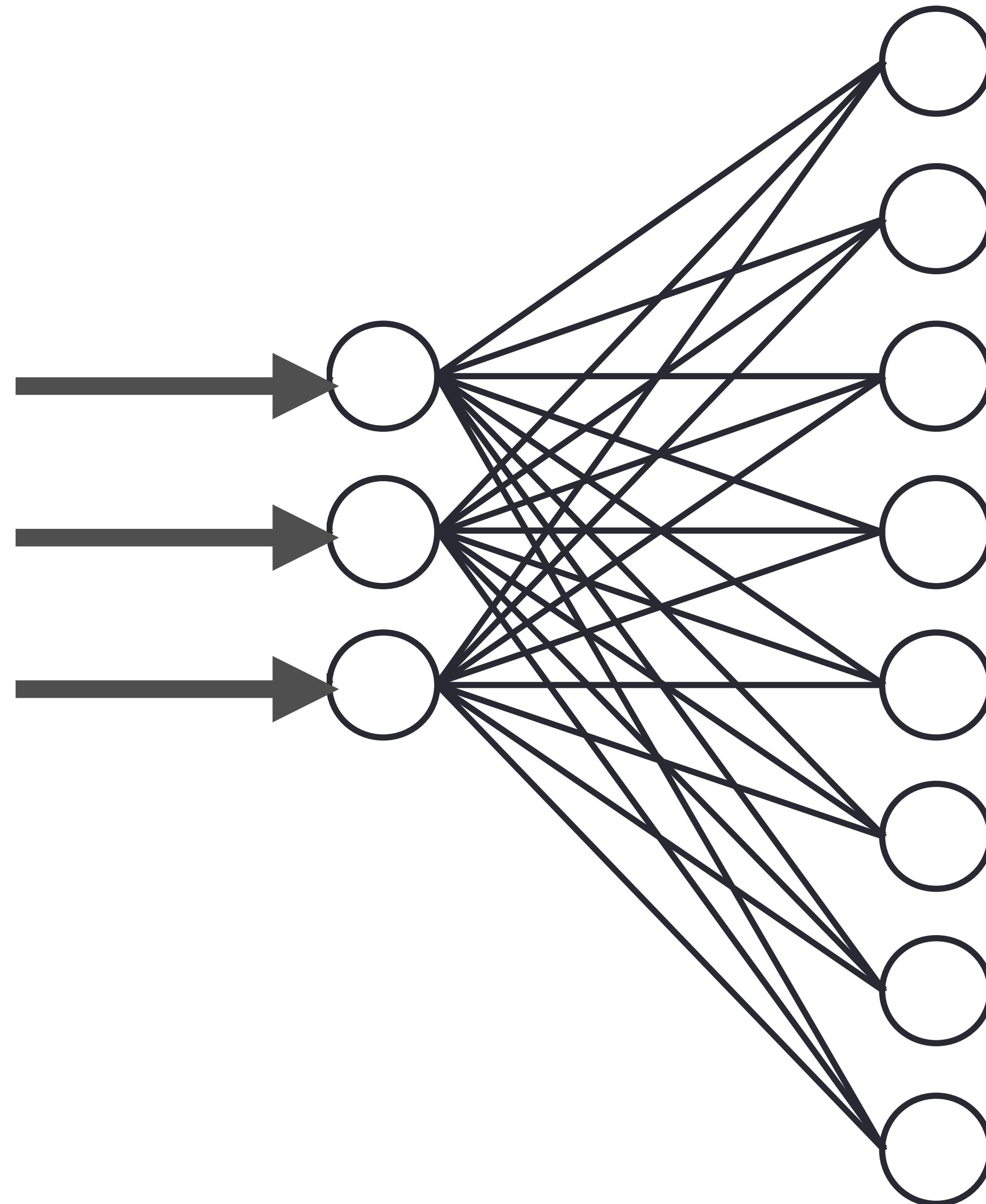


ENCODER

WHY MIGHT THIS BE USEFUL?

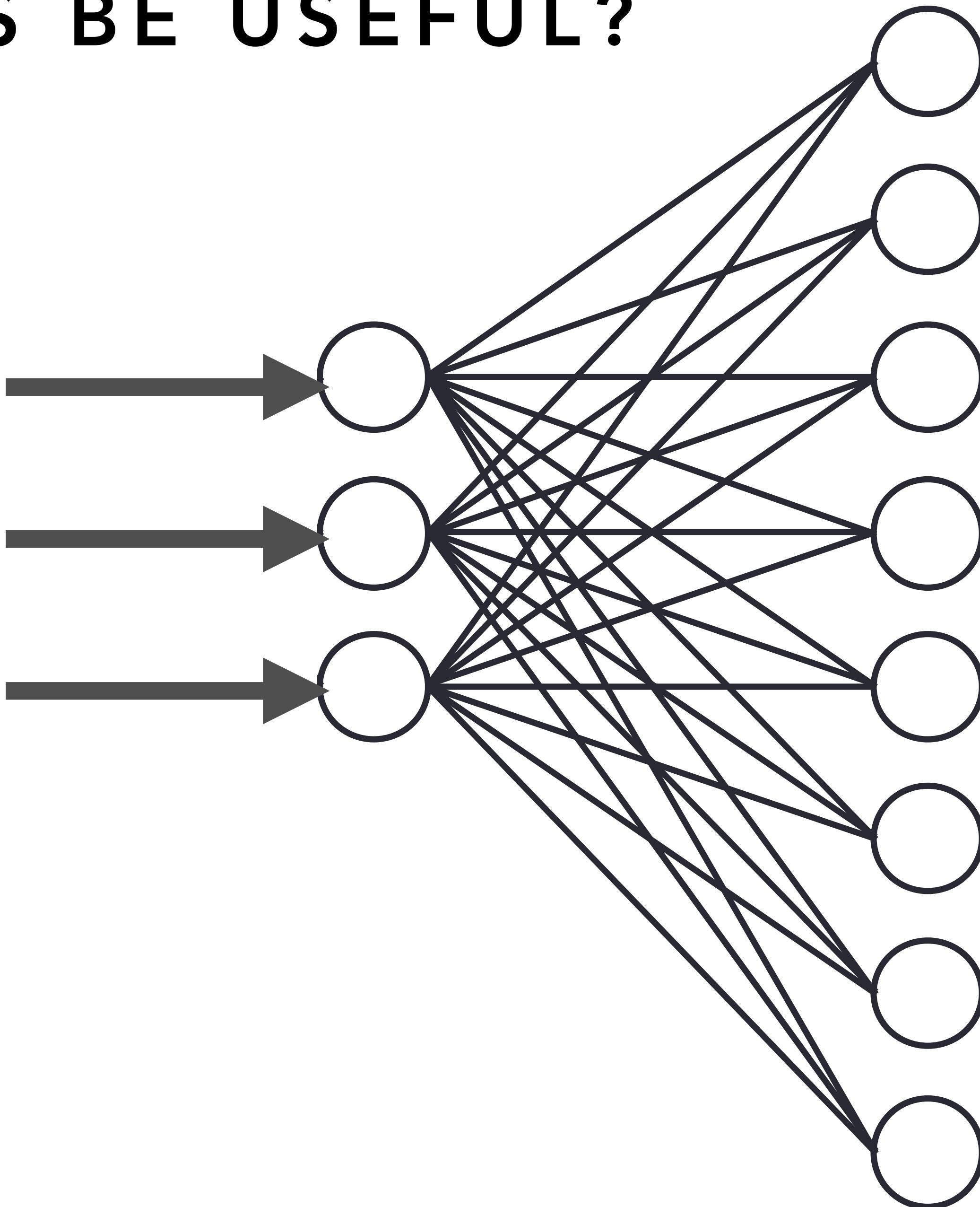


# DECODER



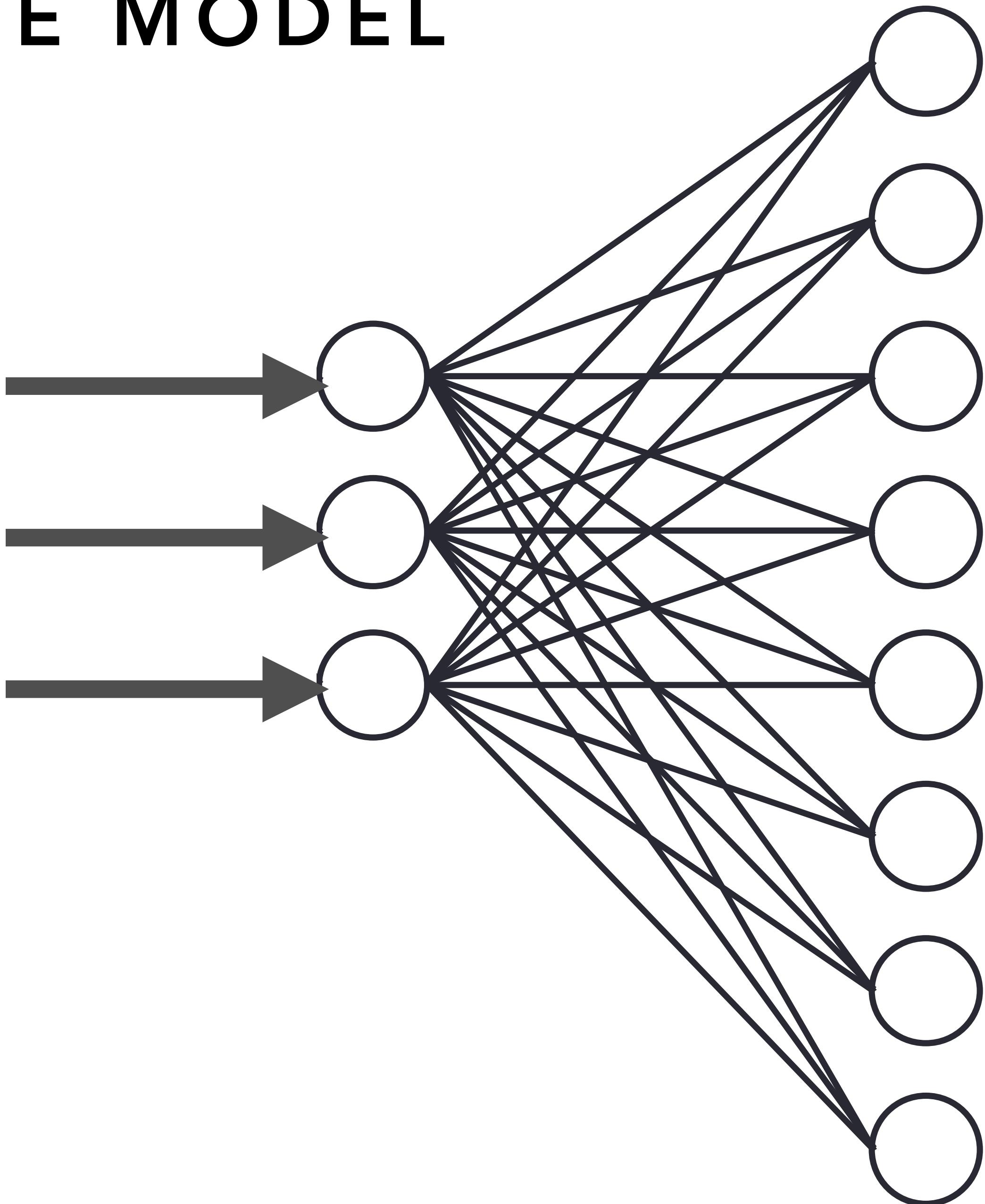
# WHY MIGHT THIS BE USEFUL?

# DECODER

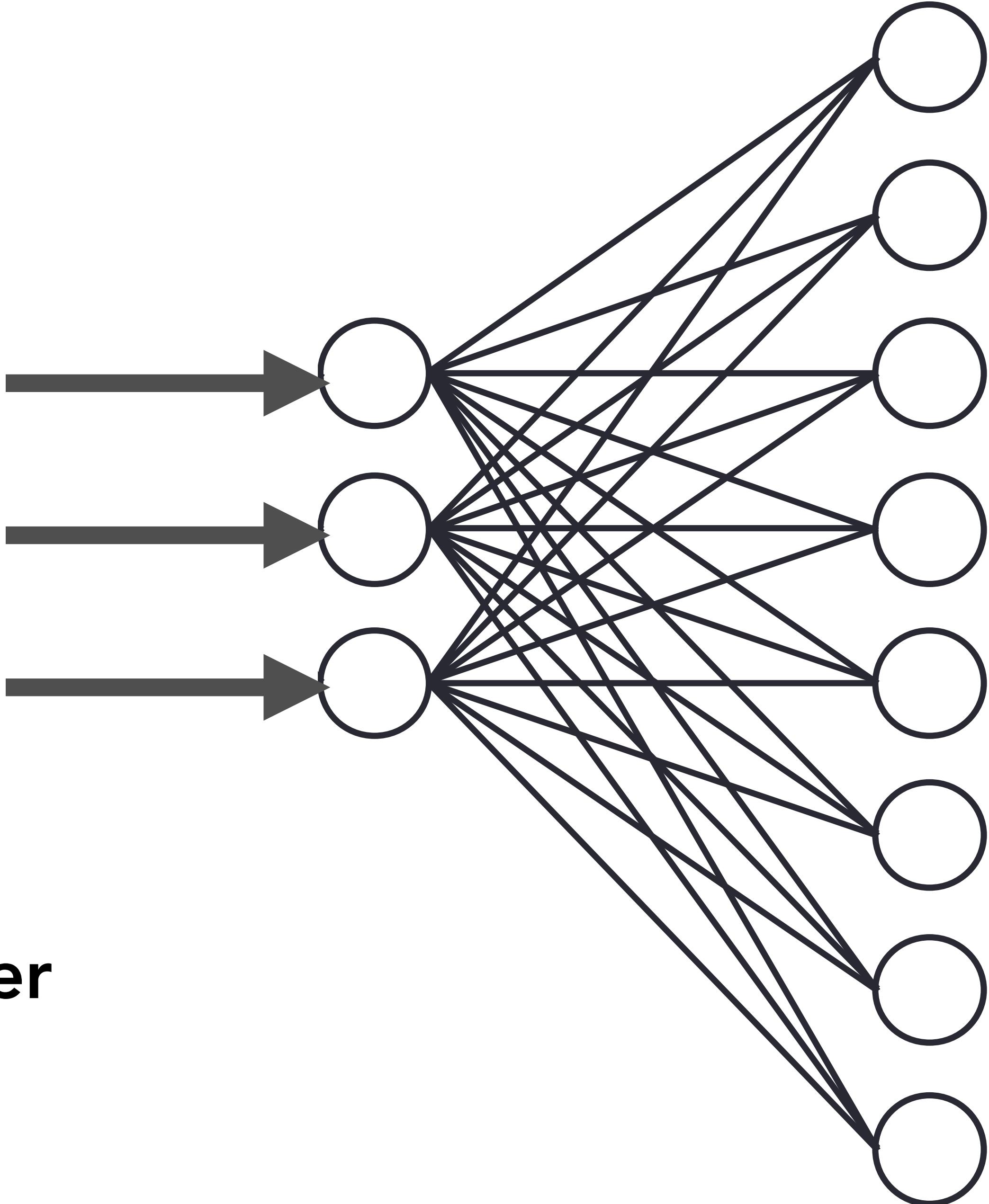


AS A GENERATIVE MODEL

DECODER

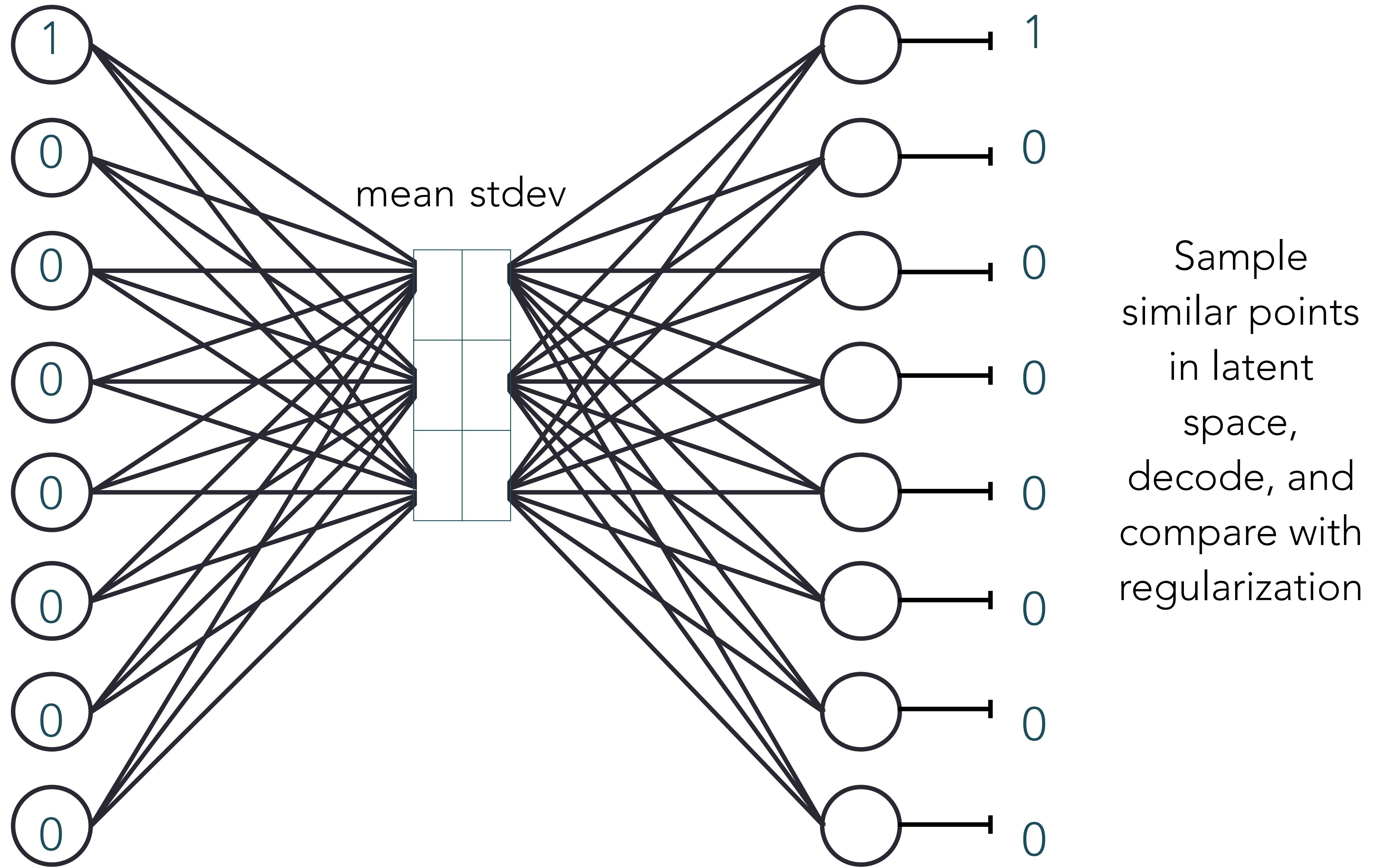


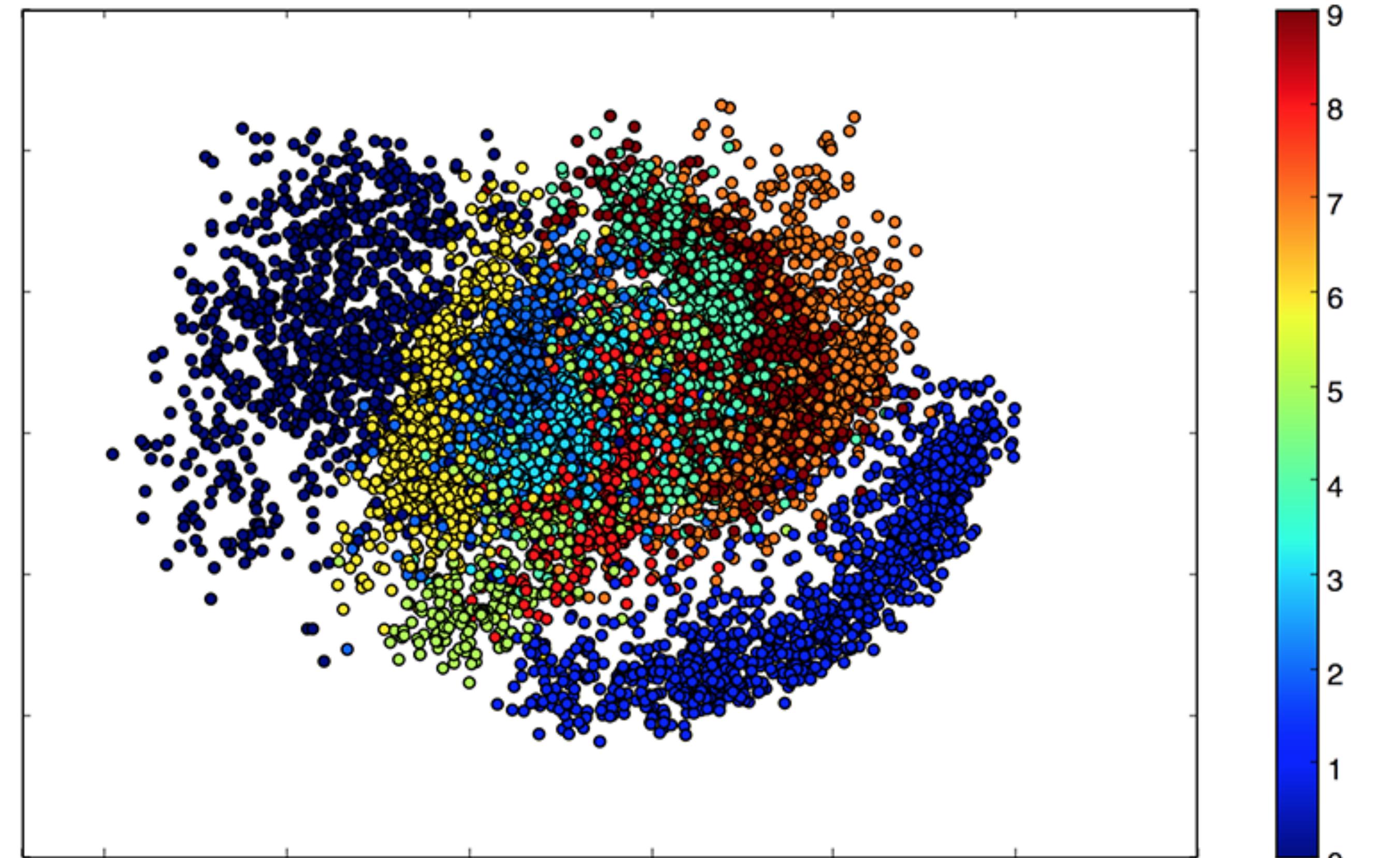
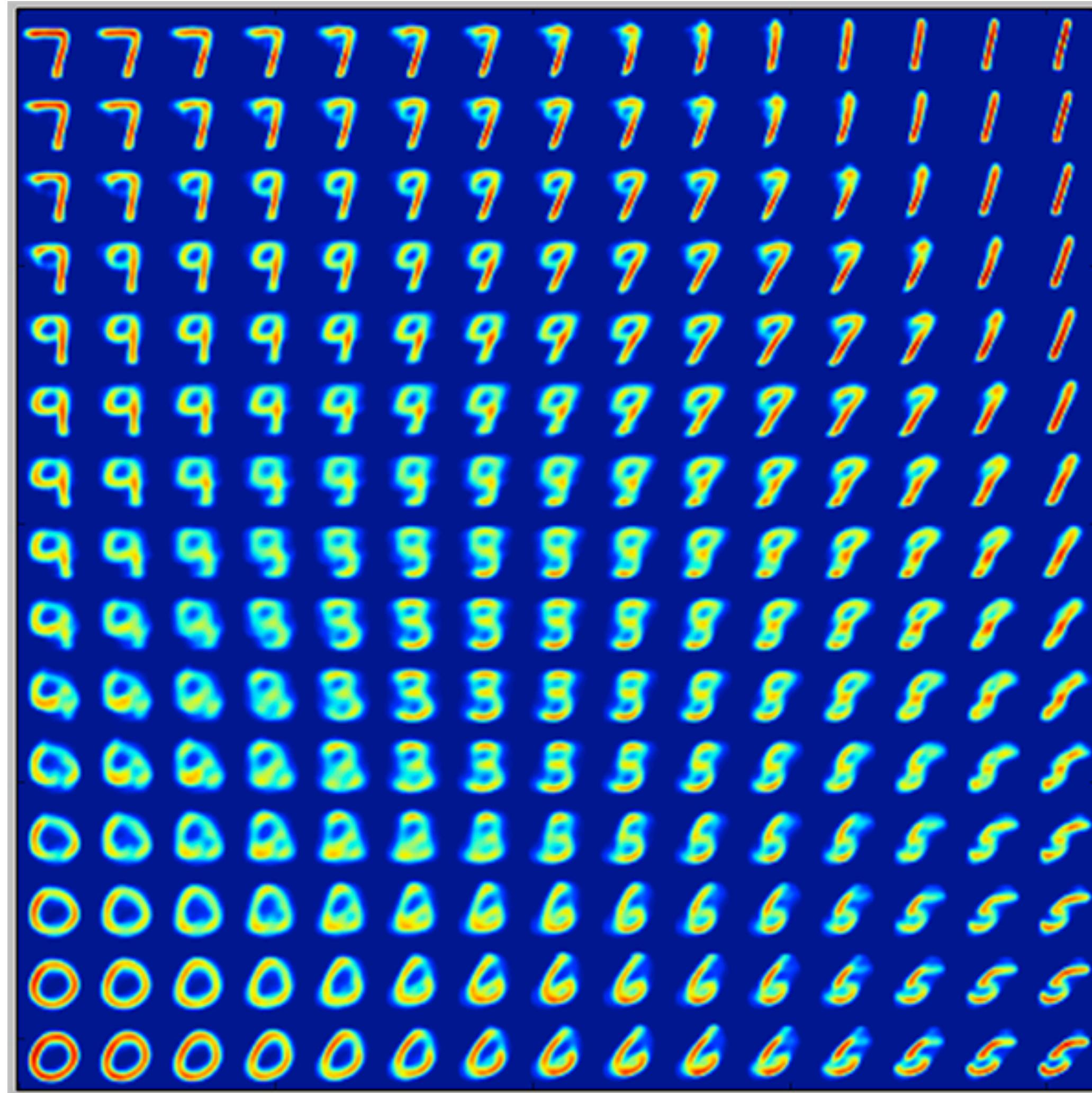
How do we know that  
we are providing a  
latent vector that  
represents those seen  
in training?



## Variational Autoencoder

Encode to  
two outputs  
for each  
latent  
dimension:  
mean and  
stdev

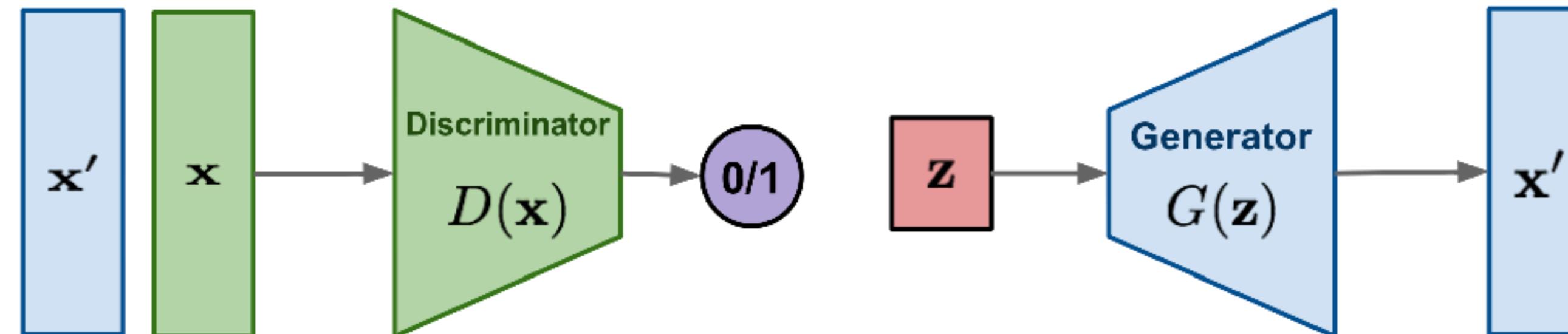




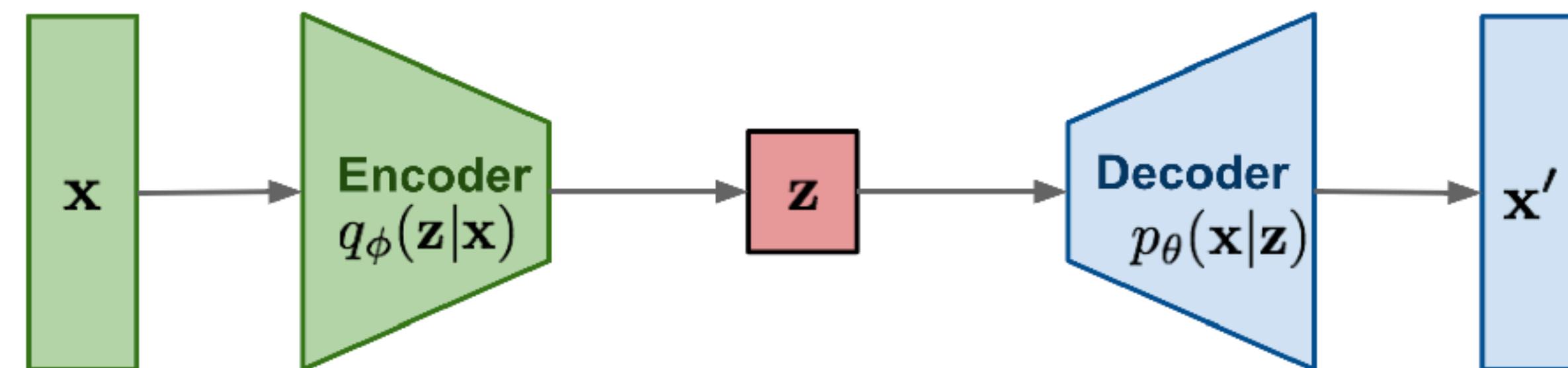
<https://blog.keras.io/building-autoencoders-in-keras.html>

# DISTRIBUTIONS

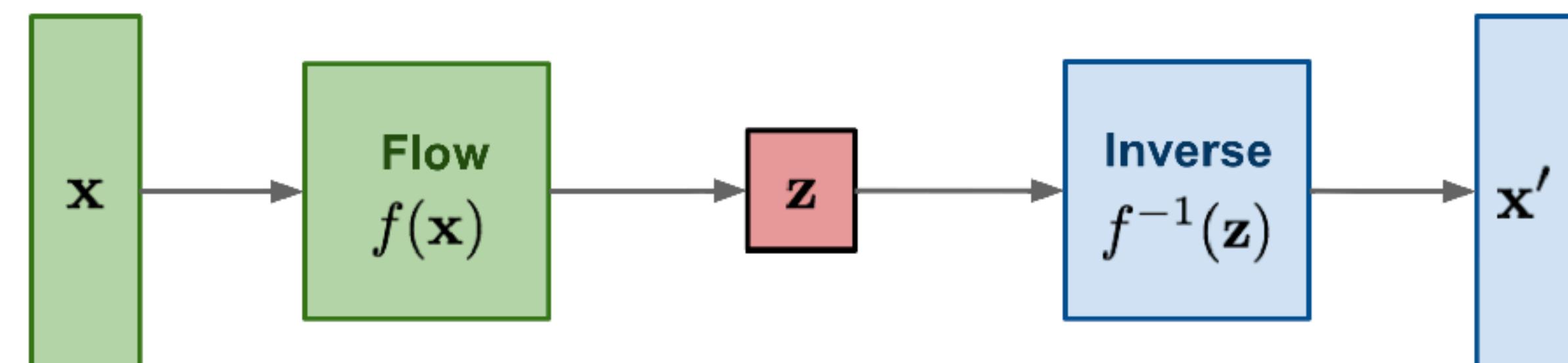
**GAN:** minimize the classification error loss.



**VAE:** maximize ELBO.



**Flow-based generative models:** minimize the negative log-likelihood



# BEST PRACTICES

## LITERATURE

# ETHICS

Large, pertained models

- power (cost and authority)

- bias

image models

language models

# MODEL CARDS

## Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru  
{mmitchellai,simonewu,andyzaldivar,parkerbarnes,lucyvasserman,benhutch,espitzer,tgebru}@google.com  
deborah.raji@mail.utoronto.ca

### Model Card

- **Model Details.** Basic information about the model.
  - Person or organization developing model
  - Model date
  - Model version
  - Model type
  - Information about training algorithms, parameters, fairness constraints or other applied approaches, and features
  - Paper or other resource for more information
  - Citation details
  - License
  - Where to send questions or comments about the model
- **Intended Use.** Use cases that were envisioned during development.
  - Primary intended uses
  - Primary intended users
  - Out-of-scope use cases
- **Factors.** Factors could include demographic or phenotypic groups, environmental conditions, technical attributes, or others listed in Section 4.3.
  - Relevant factors
  - Evaluation factors
- **Metrics.** Metrics should be chosen to reflect potential real-world impacts of the model.
  - Model performance measures
  - Decision thresholds
  - Variation approaches
- **Evaluation Data.** Details on the dataset(s) used for the quantitative analyses in the card.
  - Datasets
  - Motivation
  - Preprocessing
- **Training Data.** May not be possible to provide in practice. When possible, this section should mirror Evaluation Data. If such detail is not possible, minimal allowable information should be provided here, such as details of the distribution over various factors in the training datasets.
- **Quantitative Analyses**
  - Unitary results
  - Intersectional results
- **Ethical Considerations**
- **Caveats and Recommendations**

# MODEL CARDS

## Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru  
{mmitchellai,simonewu,andyzaldivar,parkerbarnes,lucyvasserman,benhutch,espitzer,tgebru}@google.com  
deborah.raji@mail.utoronto.ca

### Model Card

- **Model Details.** Basic information about the model.
  - Person or organization developing model
  - Model date
  - Model version
  - Model type
  - Information about training algorithms, parameters, fairness constraints or other applied approaches, and features
  - Paper or other resource for more information
  - Citation details
  - License
  - Where to send questions or comments about the model
- **Intended Use.** Use cases that were envisioned during development.
  - Primary intended uses
  - Primary intended users
  - Out-of-scope use cases
- **Factors.** Factors could include demographic or phenotypic groups, environmental conditions, technical attributes, or others listed in Section 4.3.
  - Relevant factors
  - Evaluation factors

- **Metrics.** Metrics should be chosen to reflect potential real-world impacts of the model.
  - Model performance measures
  - Decision thresholds
  - Variation approaches
- **Evaluation Data.** Details on the dataset(s) used for the quantitative analyses in the card.
  - Datasets
  - Motivation
  - Preprocessing
- **Training Data.** May not be possible to provide in practice. When possible, this section should mirror Evaluation Data. If such detail is not possible, minimal allowable information should be provided here, such as details of the distribution over various factors in the training datasets.
- **Quantitative Analyses**
  - Unitary results
  - Intersectional results
- **Ethical Considerations**
- **Caveats and Recommendations**

# PUBLICATIONS

**Reproducibility:**

Transparent

Robust

# NEW RESEARCH

## Conference papers

NeurIPS: Neural Information Processing Systems

ICML: International Conference on Machine Learning

IJCAI: International Joint Conference on Artificial Intelligence

FAccT: Fairness, Accountability, and Transparency

NeurIPS: Machine Learning and the Physical Sciences Workshop

# COMMUNITY

- Each of you arrived here with your own backgrounds, specialty, and path in life
- Your experience and expertise are valuable here, no matter what it is
- If the activity is within your background, help others!
- If you are totally (or a little) lost, ask for help!
- It is our shared goal to have **each** of us leave with some new skill/knowledge/understanding

# THANK YOU!

To YOU

Especially VECC, IRIS-HEP, and organizers

ALPhA

