# TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

The Case for Symbols

## Vector Quantization (Emergent Symbols)

Vector quantization represents a distribution (or density) on vectors with a discrete set of embedded symbols.

Vector quantization optimizes a rate-distortion tradeoff for vector compression.

The VQ-VAE uses vector quantization to construct a discrete representation of images and hence a measurable image compression rate-distortion trade-off.

## Symbols: A Better Learning Bias

Do the objects of reality fall into categories?

If so, shouldn't a learning architecture be designed to categorize?

Whole image symbols would yield emergent whole image classification.

## Symbols: Improved Interpretability

Vector quantization shifts interpretation from linear threshold units to the emergent symbols.

This seems related to the use of t-SNE as a tool in interpretation.

## Symbols: Unifying Vision and Language

Modern language models use word vectors.

Word vectors are embedded symbols.

Vector quantization also results in models based on embedded symbols.

## Symbols: Addressing the "Forgetting" Problem

When we learn to ski we do not forget how to ride a bicycle.

However, when a model is trained on a first task, retraining on a second tasks degrades performance on the first (the model "forgets").

But embedded symbols can be task specific.

The embedding of a task-specific symbol will not change when training on a different task.

# Symbols: Improved Transfer Learning.

Embedded symbols can be domain specific.

Separating domain-general parameters from domain-specific parameters may improve transfer between domains.

**END**