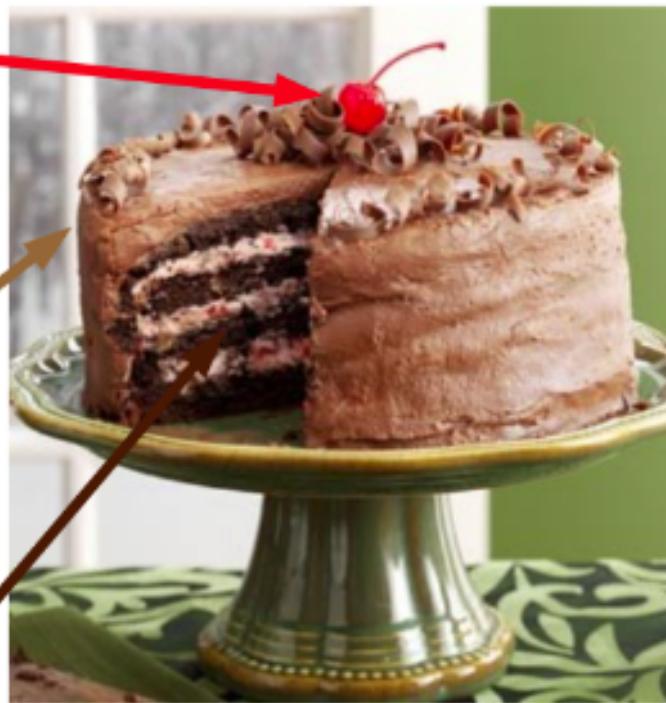# TTIC 31230, Fundamentals of Deep Learning

David McAllester, April 2017

Unsupervised and Predictive Learning

"Pure" Reinforcement Learning (cherry)
- The machine predicts a scalar reward given once in a while.
- **A few bits for some samples**

Supervised Learning (icing)
- The machine predicts a category or a few numbers for each input
- Predicting human-supplied data
- **10→10,000 bits per sample**

Unsupervised/Predictive Learning (cake)
- The machine predicts any part of its input for any observed part.
- Predicts future frames in videos
- **Millions of bits per sample**

(Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

[LeCun]

Unsupervised/Predictive Learning

Can we formally define unsupervised or predictive learning?

Learning from raw text. We think of the next word as a "label" to be predicted from the previous words. Mathematically a labeling problem.

Learning from a corpus of raw images. We can try to predict the color map from the grey-scale version of the image. Again a labeling problem.

Learning from raw video. Can we predict the next frame from the preceding frames. Again a labeling problem.

Unsupervised Learning

By "unsupervised learning" we will mean learning from **massively available** data. This is not a mathematical definition.

Massive: images, audio, text, video, click-through data.

Less Massive: car control data, stereo image pairs, closed captioned video, captioned images.

Big: Manually annotated images or audio.

Small: manually annotated text — parse trees, named entities, semantic roles, coreference, entailment.

Smallest: Manually annotated text in an obscure language.

Colorization

We have massive data for colorization (by decolorizing color images).

But any particular colorization is still a guess.

For colorization we might evaluate by measuring the best match over k guesses.

But in complex images even k guesses will not include a good match.

Modeling Distributions

We want to construct a model $P_\Theta(x)$ or $P_\Theta(y|x)$ of a (natural) data distribution given massive samples of x or massive samples of (x, y).

In cases where massive data is available (the unsupervised or predictive case) we typically want to model distributions $P(x)$ or $P(y|x)$ that are sufficiently broad that any guess of x or any guess of y given x is almost sure to be wrong.

Summary of Unsupervised and Predictive Learning

There is no mathematically well defined problem of unsupervised or predictive learning.

Unsupervised and predictive learning should simply be interpreted to mean that we have **massive** samples of x that can can be used to model $P(x)$ or **massive** samples of (x, y) that can be used to model $P(y|x)$.

END