Deep Generative Models

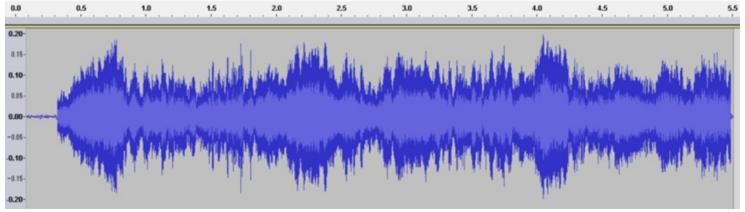
Mostafa Tavassolipour and MohammadAmin Sadeghi

Fall 2023

Introduction

Challenge: understand complex, unstructured inputs

Computer Vision



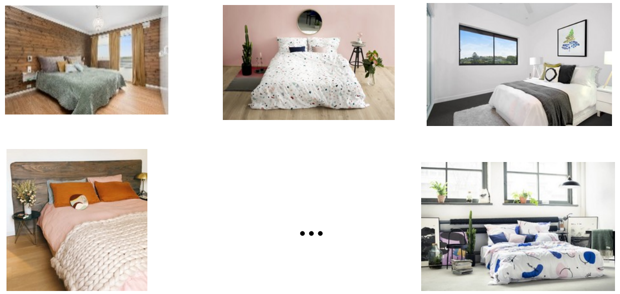
Computational Speech

Natural Language Processing

Robotics

Statistical Generative Models

Statistical generative models are **learned from data**



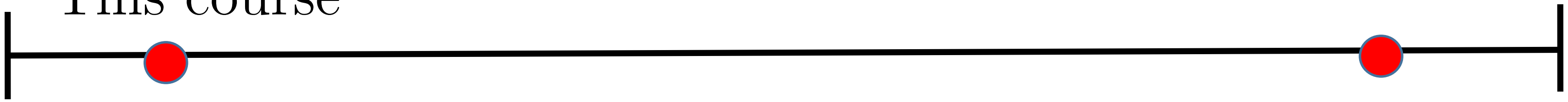
Data
(e.g., images of bedrooms)

+

Prior Knowledge
(e.g., physics, materials, ..)

Priors are always necessary, but there is a spectrum

This course

Data  Prior Knowledge

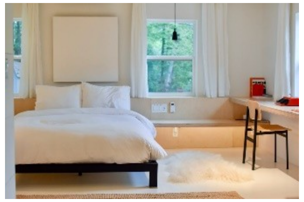
Statistical Generative Models

A statistical generative model is a **probability distribution** p(x)

- **Data:** samples (e.g., images of bedrooms)
- **Prior knowledge:** parametric form (e.g., Gaussian?), loss function (e.g., maximum likelihood?), optimization algorithm, etc.

Image x  → A probability distribution p(x) → probability p(x)

It is generative because **sampling from p(x) generates new images**

 … 

Discriminative vs. generative

Discriminative: classify bedroom vs. dining room

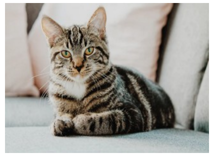
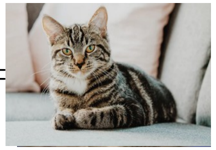
The image X is given. **Goal**: decision boundary, via **conditional distribution over label Y**

P(Y = Bedroom | X= ) = 0.0001

Ex: logistic regression, convolutional net, etc.

Generative: generate X

Y=B , X=  Y=D , X= 

Y=B , X=  Y=D , X= 

... ...

Y=B , X=  Y=D , X= 

The input X is **not** given. Requires a model of the **joint distribution over both X and Y**

P(Y = Bedroom , X= ) = 0.0002

Discriminative vs. generative

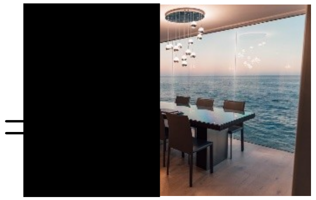
Joint and conditional are related via **Bayes Rule**:

$P(Y = \text{Bedroom} \mid X=$  $) =$

$P(Y = \text{Bedroom},\ X=$  $)$

$P(\ X =$  $)$

Discriminative: Y is simple; X is always given, so not need to model $P(X=$ 

Therefore it cannot handle missing data $P(Y = \text{Bedroom} \mid X =$  $)$

Conditional Generative Models

Class **conditional generative models** are also possible:

P(X=  | Y = Bedroom)

It's often useful to condition on rich side information Y

P(X=  | Caption = "A black table with 6 chairs")

A discriminative model is a very simple conditional generative model of Y:

P(Y = Bedroom | X= )

Progress in Generative Models of Images

2014

2015

2016

2017

2018

Ian Goodfellow, 2019

Progress in Generative Models of Images

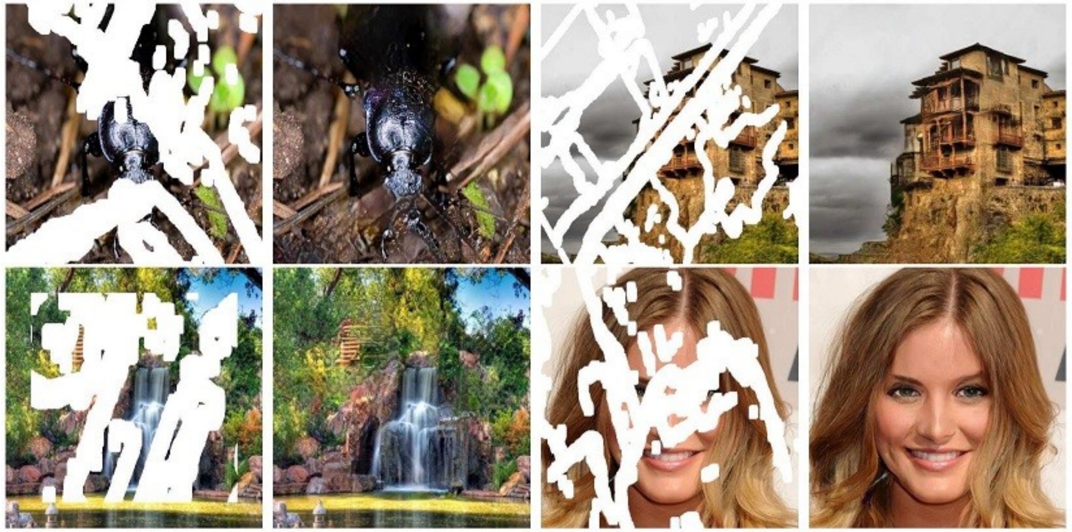
Song et al., Score-Based Generative Modeling through Stochastic Differential Equations, 2021

Progress in Inverse Problems

P(high resolution | low resolution)

Menon et al, 2020

P(full image| mask)

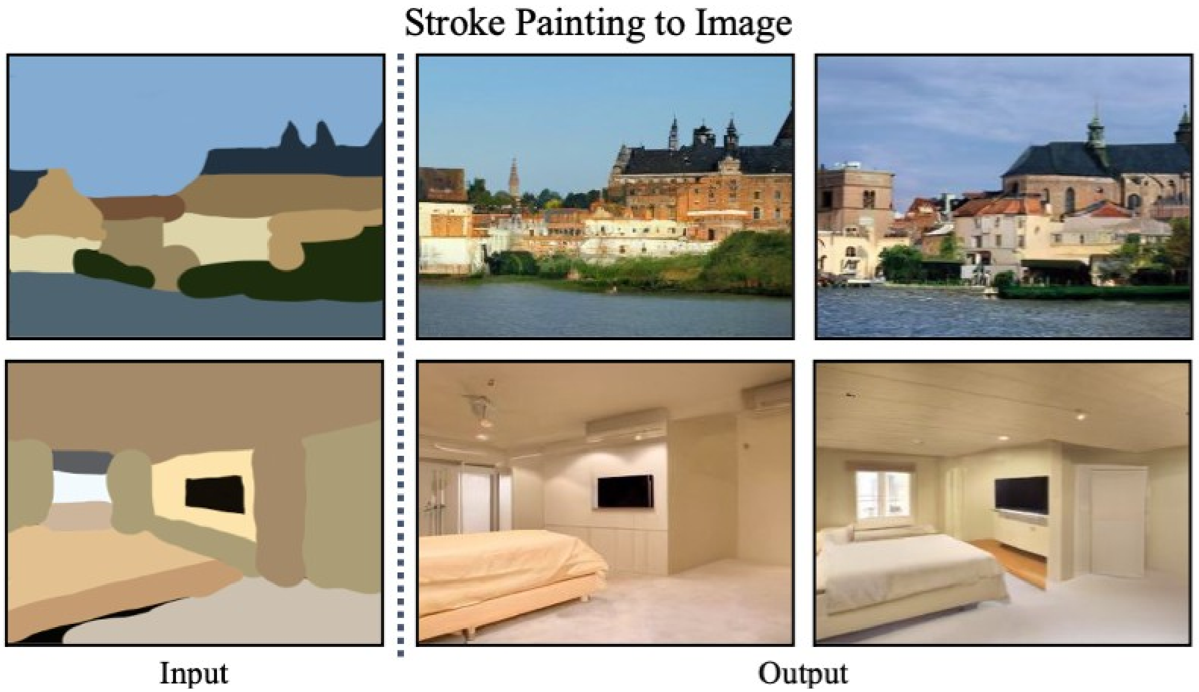


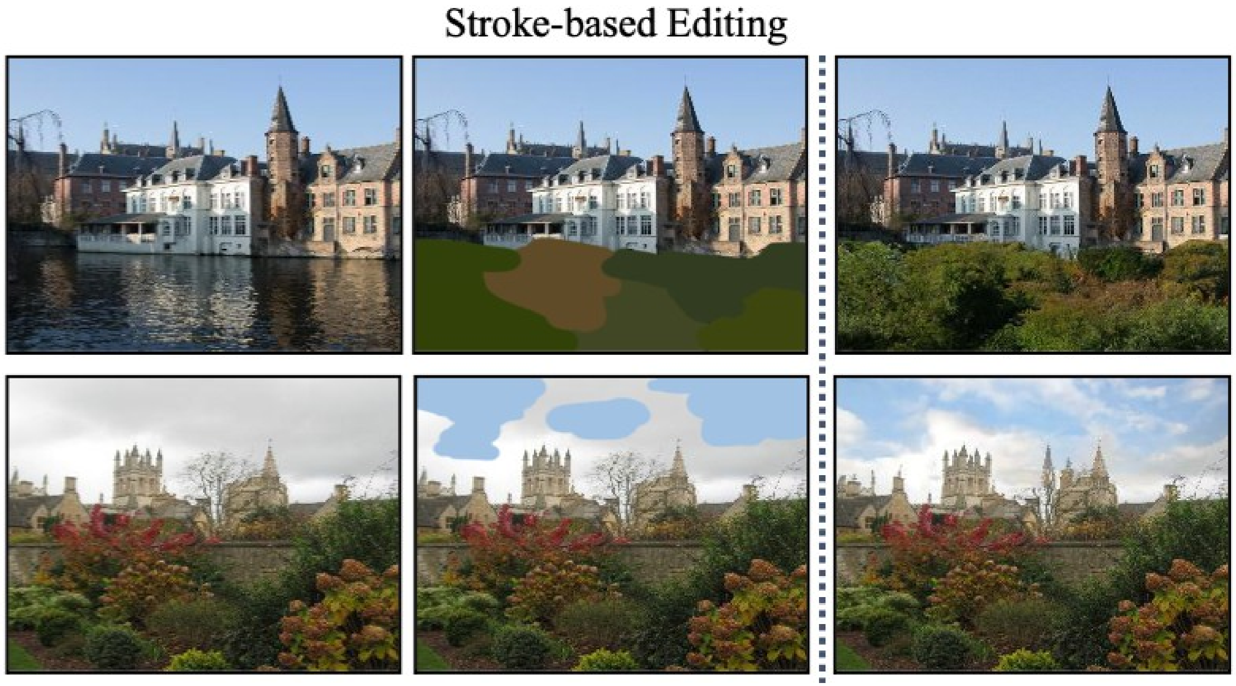
Liu al, 2018

P(color image| greyscale)

Antic, 2020

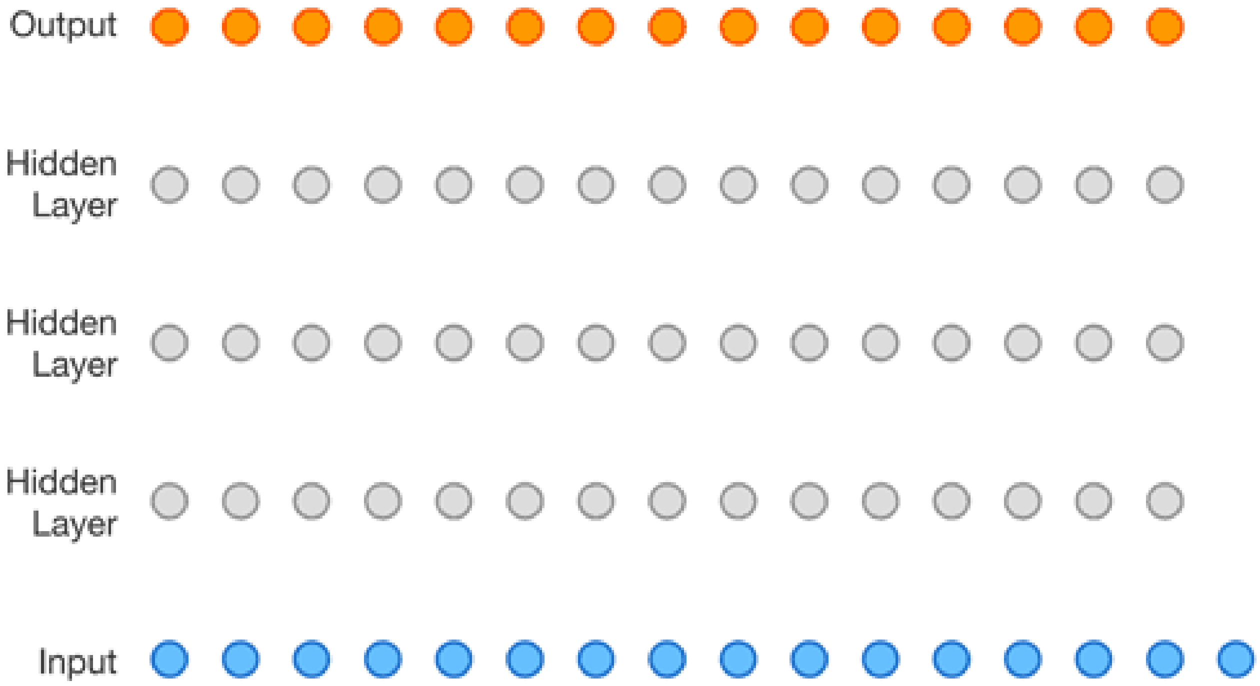
Progress in Inverse Problems





WaveNet

Generative model of speech signals



Text to Speech


Parametric


Concatenative


WaveNet

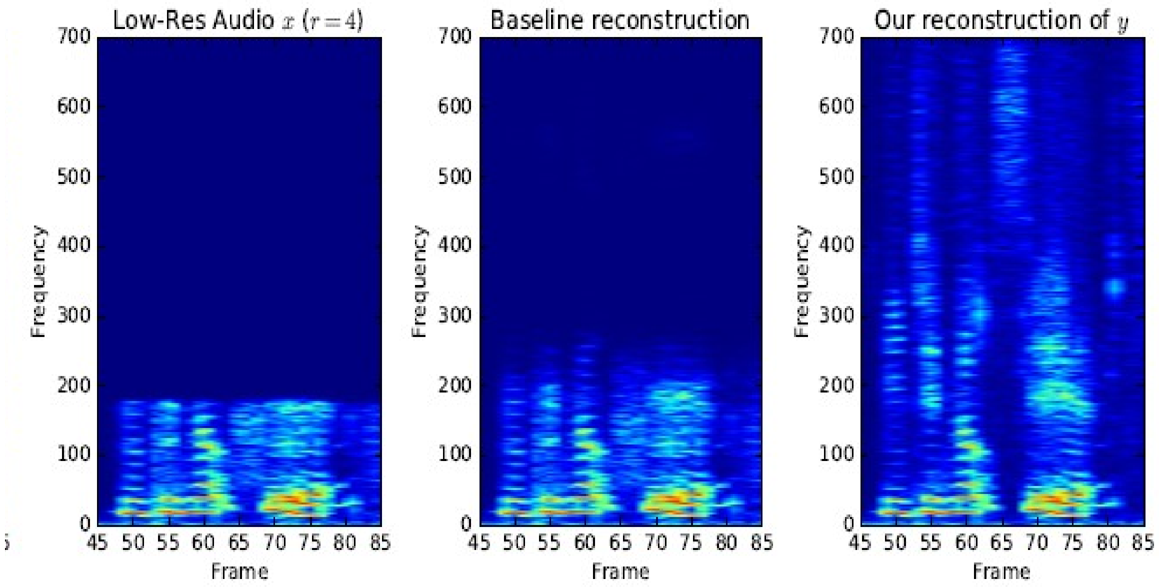

Unconditional


Music

van den Oord et al, 2016c

Audio Super Resolution

Conditional generative model P(high-res signal | low-res audio signal)




Low res signal


High res audio signal

Kuleshov et al., 2017

Machine Translation

Conditional generative model $P(\text{ English text} | \text{ Chinese text})$

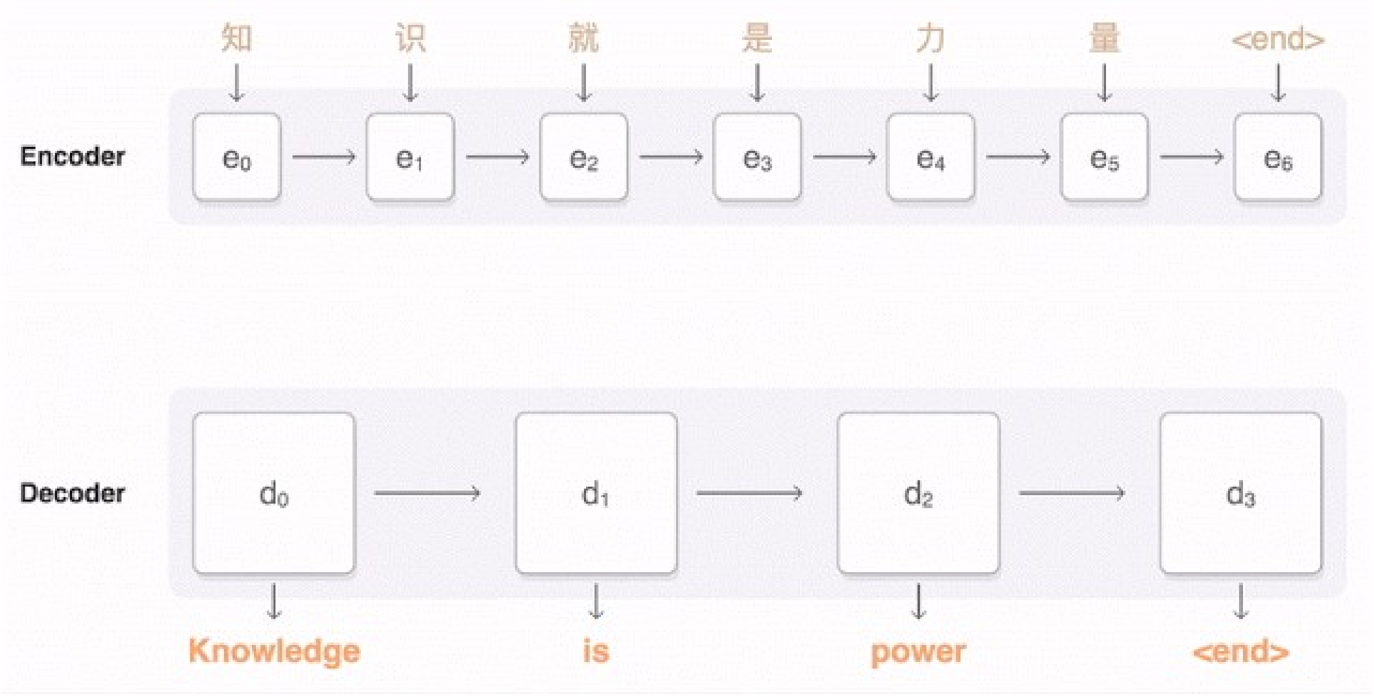


Figure from Google AI research blog.

Language Generation

Custom prompt

To get an A+ in deep generative models, students have to

Completion

To get an A+ in deep generative models, students have to be willing to work with problems that are a whole lot more interesting than, say, the ones that most students work on in class. If you're a great student, the question above can be avoided and you'll be able to do great work, but if you're not, you will need to go beyond the basics before getting good.

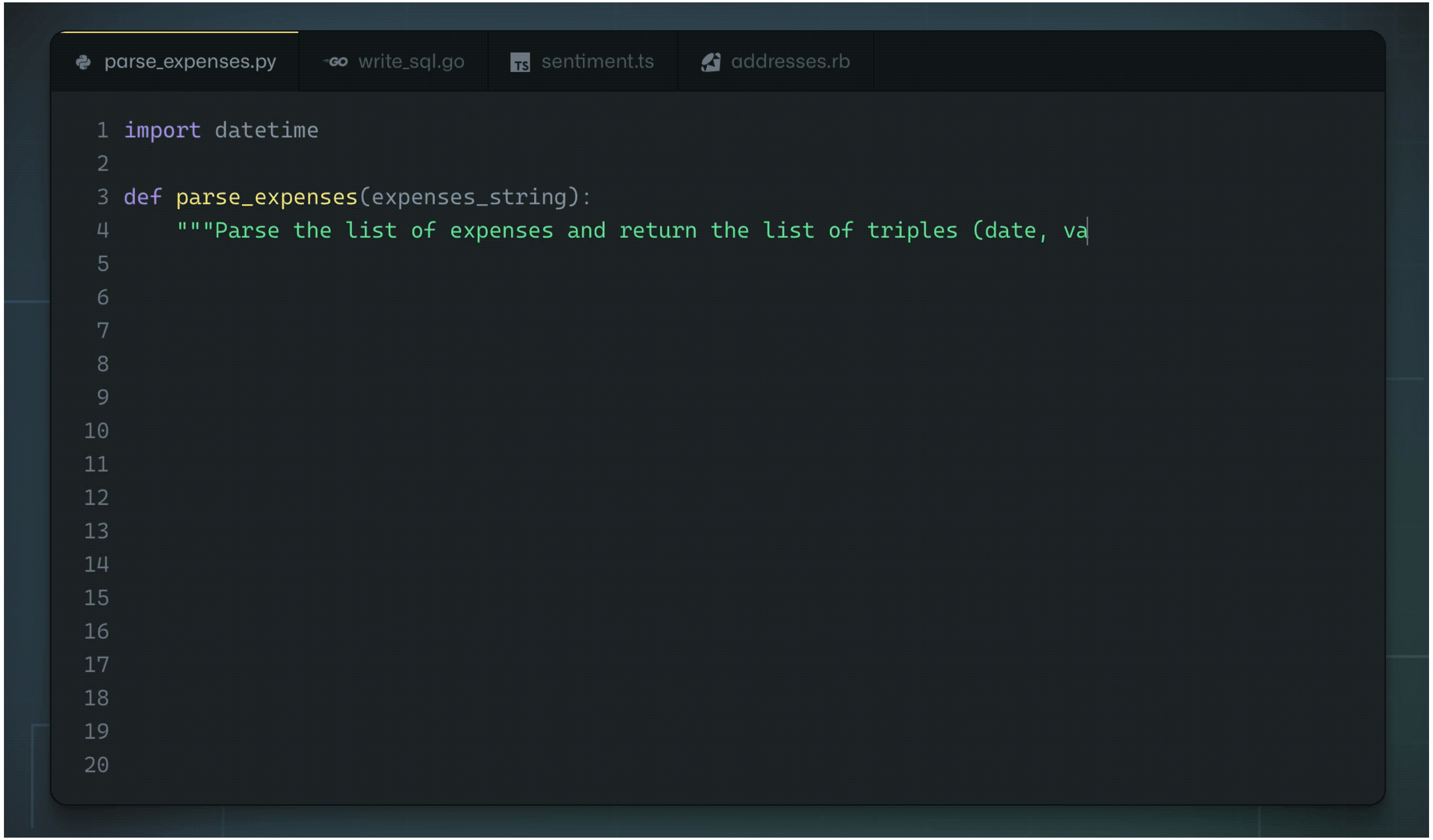
Now to be clear, this advice is not just for the deep-learning crowd; it is good advice for any student who is taking his or her first course in machine learning.

The key point is that if you have a deep, deep brain of a computer scientist, that's just as important to you.

$P(\text{next word} \mid \text{previous words})$

Radford et al., 2019
Demo from talktotransformer.com

Code Generation



OpenAI Codex

Images and Text

TEXT PROMPT an armchair in the shape of an avocado. . . .

AI-GENERATED IMAGES

Edit prompt or view more images↓

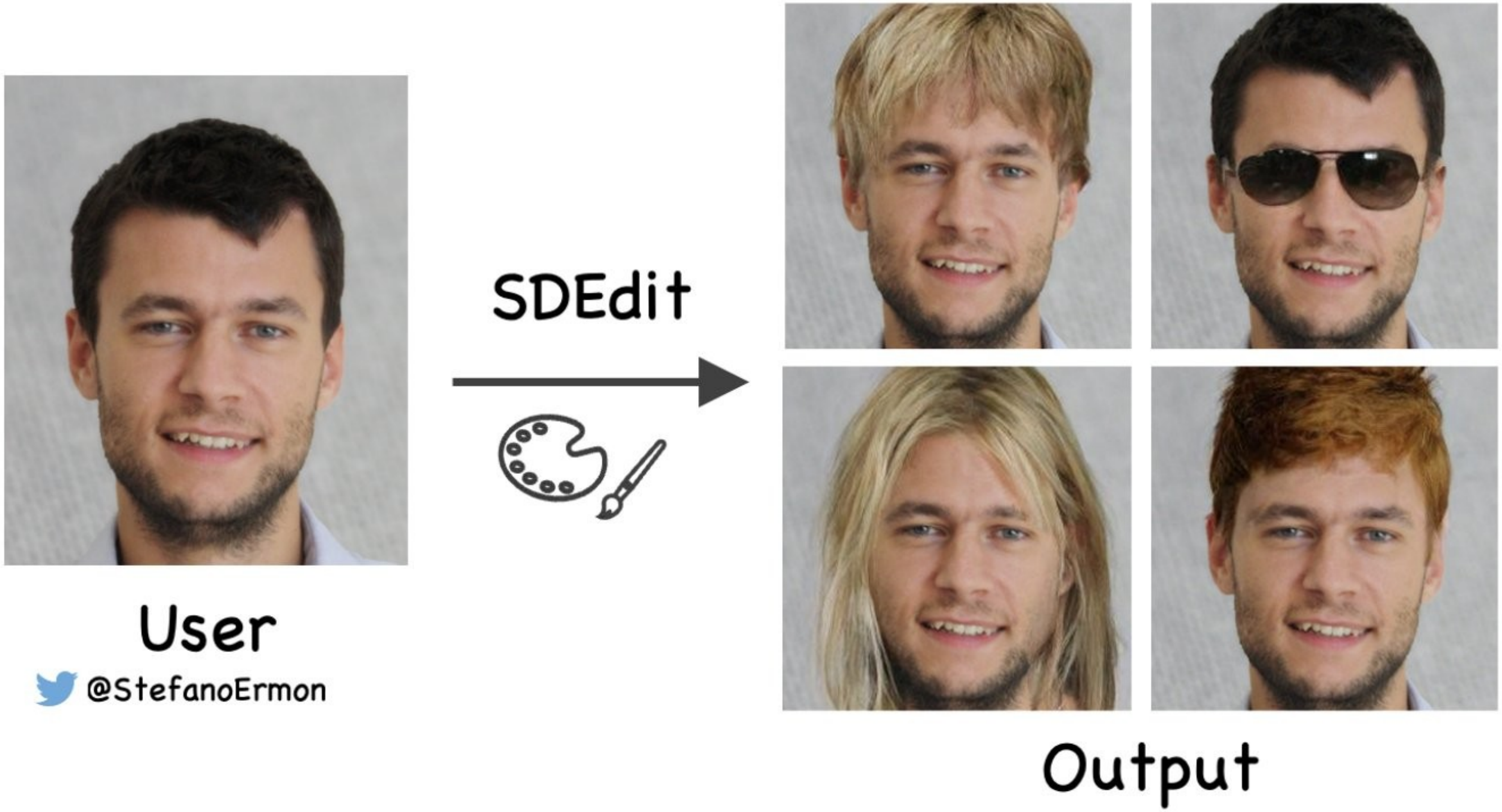
$$P(\text{image} \mid \text{caption})$$

TEXT PROMPT a store front that has the word ‘openai’ written on it. . . .

AI-GENERATED IMAGES

DeepFakes

Which image is real?



DeepFakes

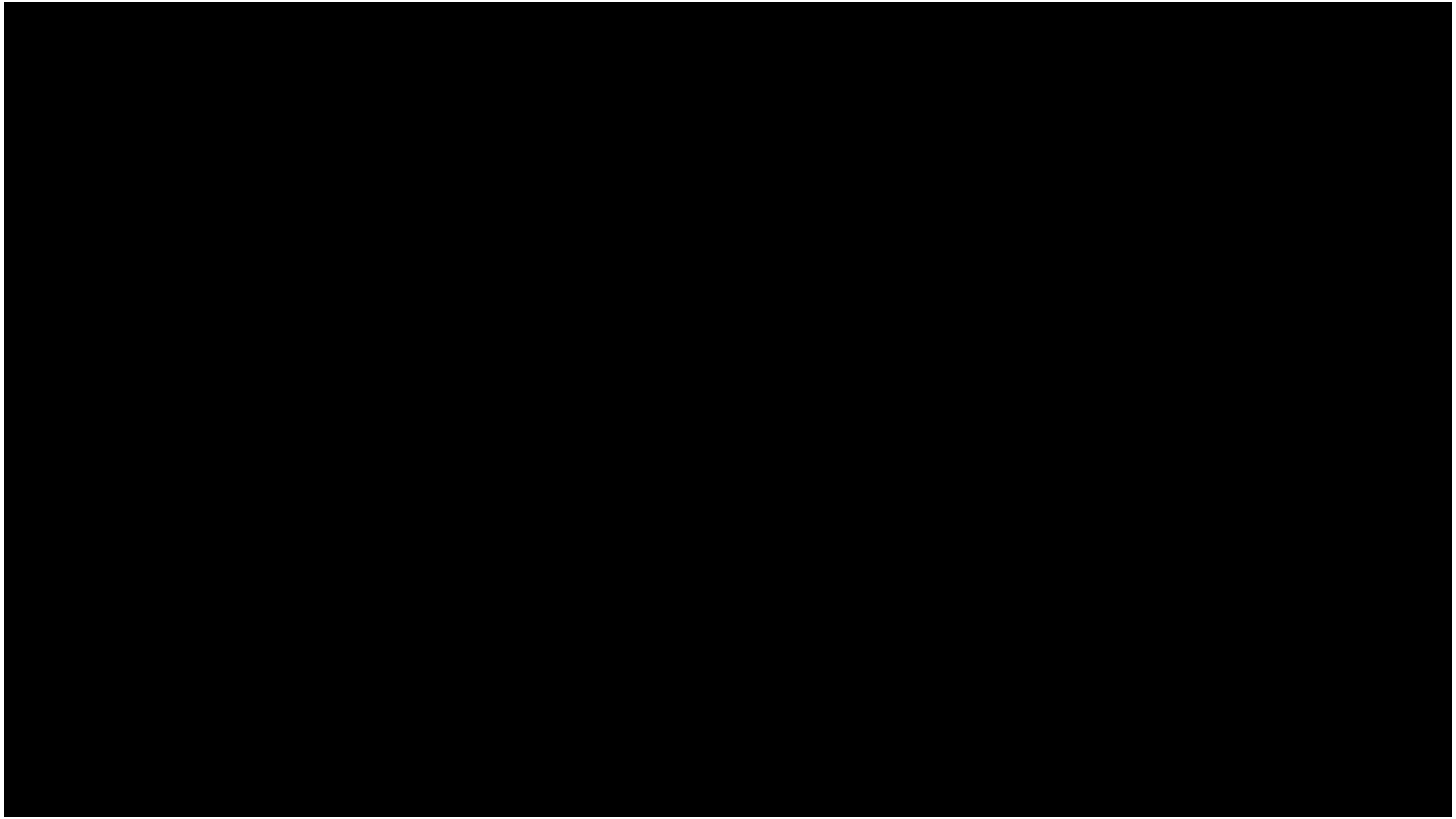


Image Translation

Conditional generative model $P(\text{ zebra images} | \text{ horse images})$

Zhu et al., 2017

Imitation Learning

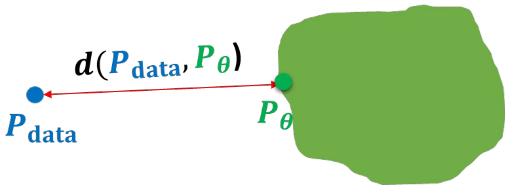
Conditional generative model P(actions | past observations)

Li et al., 2017

Roadmap and Key Challenges

- **Representation**: how do we model the joint distribution of many random variables?
  - Need compact representation
- **Learning**: what is the right way to compare probability distributions?

$\mathbf{x}_i \sim P_{\text{data}}$
$i = 1, 2, \ldots, n$



$\theta \in M$

Model family

- **Inference**

Syllabus

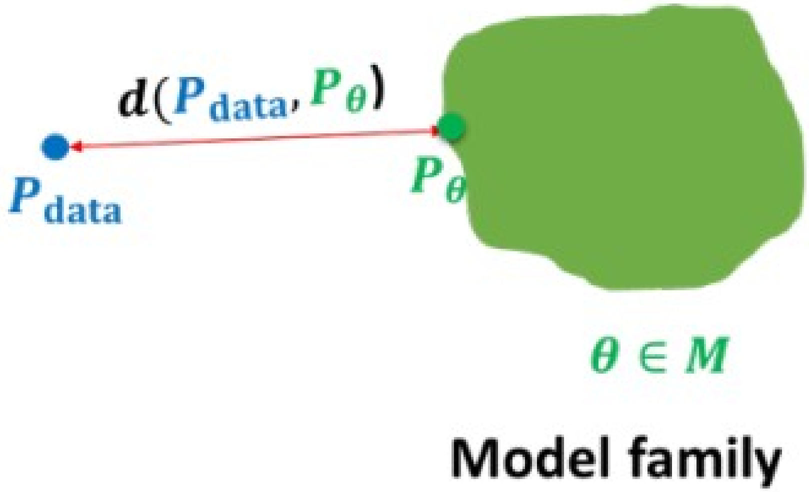
- Probabilistic Graphical Models
  - Bayesian Networks
  - Markov Random Fields
- Sampling Methods
- Causality and Causal models
- Deep Generative Models
  - Variational Autoencoders
  - Flow-based Models
  - GAN
  - Diffusion (score) based models
- NLP generative models
  - Language Models
  - Prompt engineering
  - Voice generation

Grading

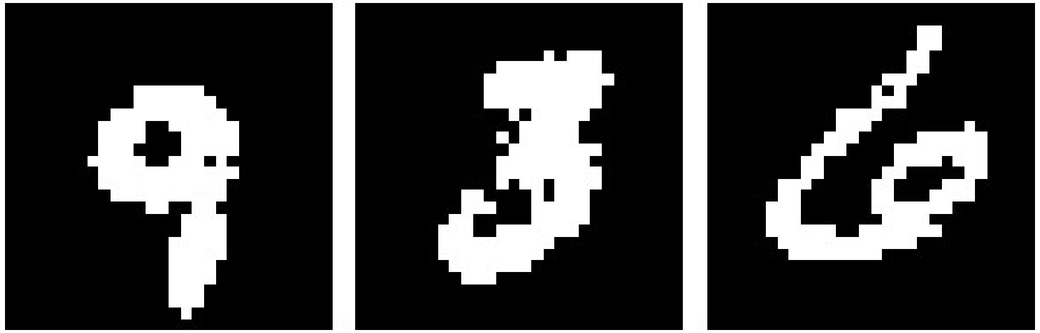
- 4 Homework Assignments:
  - 8 points
  - Grace: 12 days
- Midterm Exam:
  - 20th Aban
  - 6 points
- Final Exam:
  - 4th Bahman
  - 6 points

Learning a generative model

$\mathbf{x}_i \sim P_{\text{data}}$
$i = 1, 2, \ldots, n$



Example of joint distribution



- Suppose $X_1, \ldots, X_n$ are binary (Bernoulli) random variables, i.e., $\text{Val}(X_i) = \{0, 1\} = \{\text{Black}, \text{White}\}$.
- How many possible states?

$$\underbrace{2 \times 2 \times \cdots \times 2}_{n \text{ times}} = 2^n$$

- Sampling from $p(x_1, \ldots, x_n)$ generates an image
- How many parameters to specify the joint distribution $p(x_1, \ldots, x_n)$ over n binary pixels?

$$2^n - 1$$

Independent random variables

- If $X_1, \ldots, X_n$ are independent, then

$$p(x_1, \ldots, x_n) = p(x_1)p(x_2)\cdots p(x_n)$$

- How many possible states? 2^n
- How many parameters to specify the joint distribution $p(x_1, \ldots, x_n)$?
  - How many to specify the marginal distribution $p(x_1)$? 1
- 2^n **entries can be described by just** n **numbers** (if $|Val(X_i)| = 2$)!
- Independence assumption is too strong. Model not likely to be useful
  - For example, each pixel chosen independently when we sample from it.


Chain Rule

- Using Chain Rule

$$p(x_1, \ldots, x_n) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1, x_2) \cdots p(x_n \mid x_1, \cdots, x_{n-1})$$

- How many parameters? $1 + 2 + \cdots + 2^{n-1} = 2^n - 1$
  - $p(x_1)$ requires 1 parameter
  - $p(x_2 \mid x_1 = 0)$ requires 1 parameter, $p(x_2 \mid x_1 = 1)$ requires 1 parameter
    Total 2 parameters.
  - $\cdots$
- $2^n - 1$ is still exponential, chain rule does not buy us anything.
- Now suppose $X_{i+1} \perp X_1, \ldots, X_{i-1} \mid X_i$, then

$$\begin{aligned} p(x_1, \ldots, x_n) &= p(x_1)p(x_2 \mid x_1)p(x_3 \mid \cancel{x_1}, x_2) \cdots p(x_n \mid \cancel{x_1, \cdots,} x_{n-1}) \\ &= p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_2) \cdots p(x_n \mid x_{n-1}) \end{aligned}$$

- How many parameters? $2n - 1$. Exponential reduction!