CS273B Lecture 7: generative models

James Zou

10/16/17

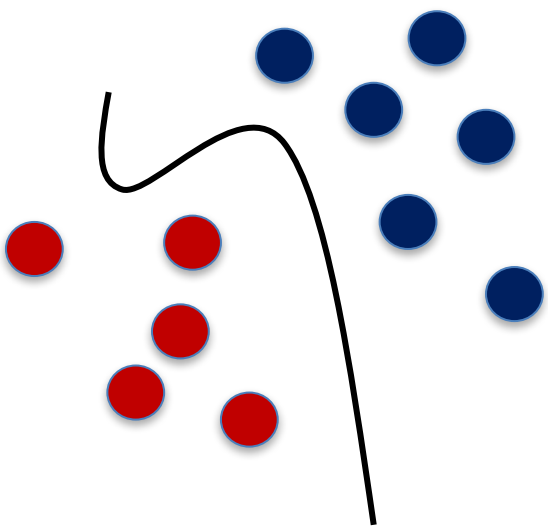
Lecture outline

1. Method 1: RNN.
   Example: molecule design.

2. Method 2: Generative adversarial networks (GAN).
   Example: electronic health records.

Generative vs. discriminative models

Discriminative model

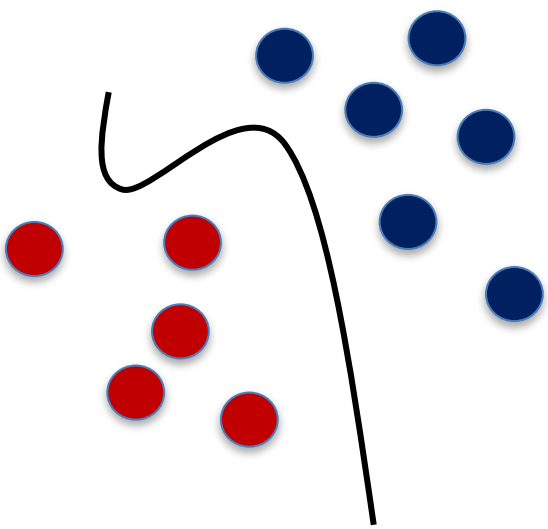


Decision boundary
(SVM, neural net, etc.)

$$P(y|x)$$

Generative vs. discriminative models

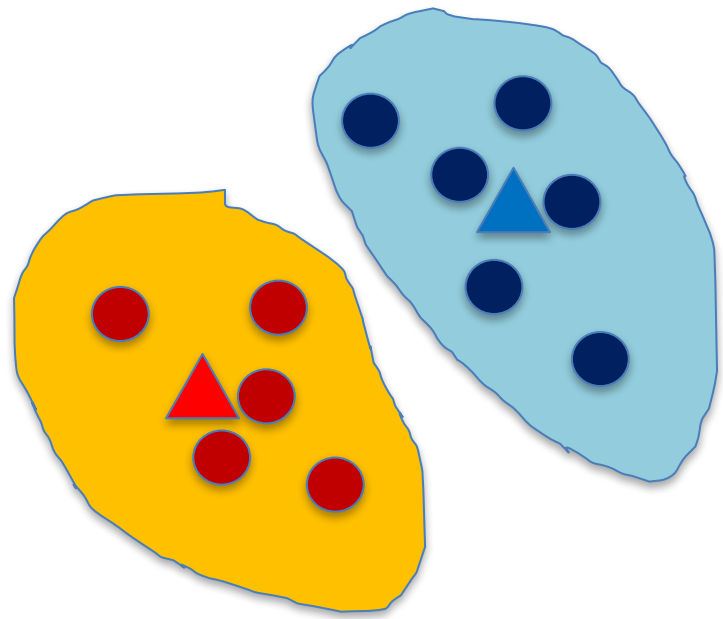
Discriminative model



Decision boundary
(SVM, neural net, etc.)

$P(y|x)$

Generative model



Mixture of Gaussians

$P(x, y)$

Q: What are the pro/cons of each method?

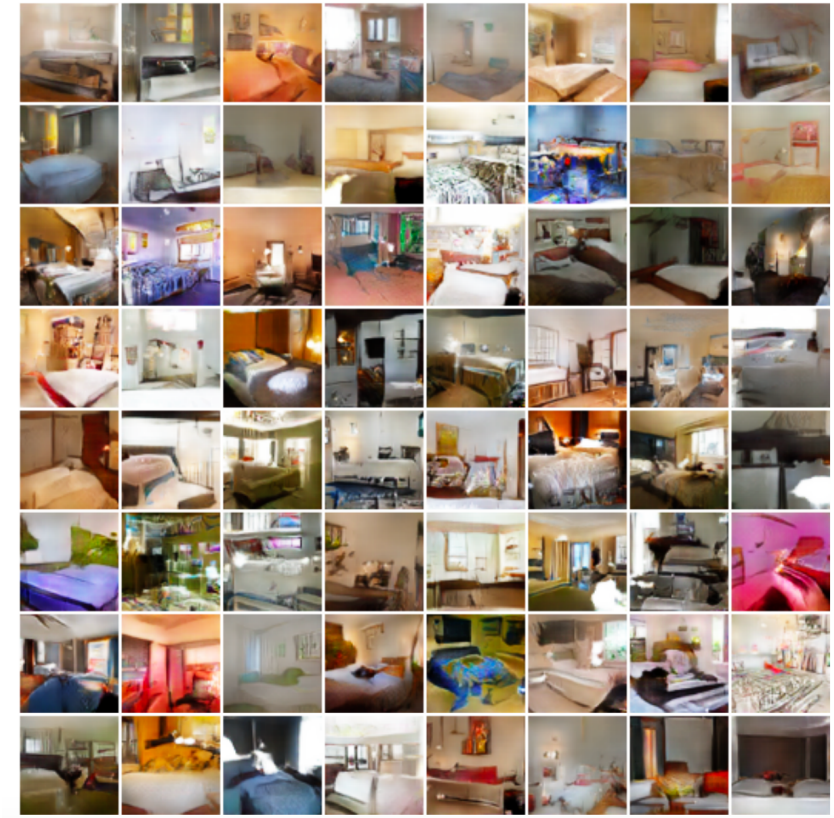
Generative vs. discriminative models

Discriminative model

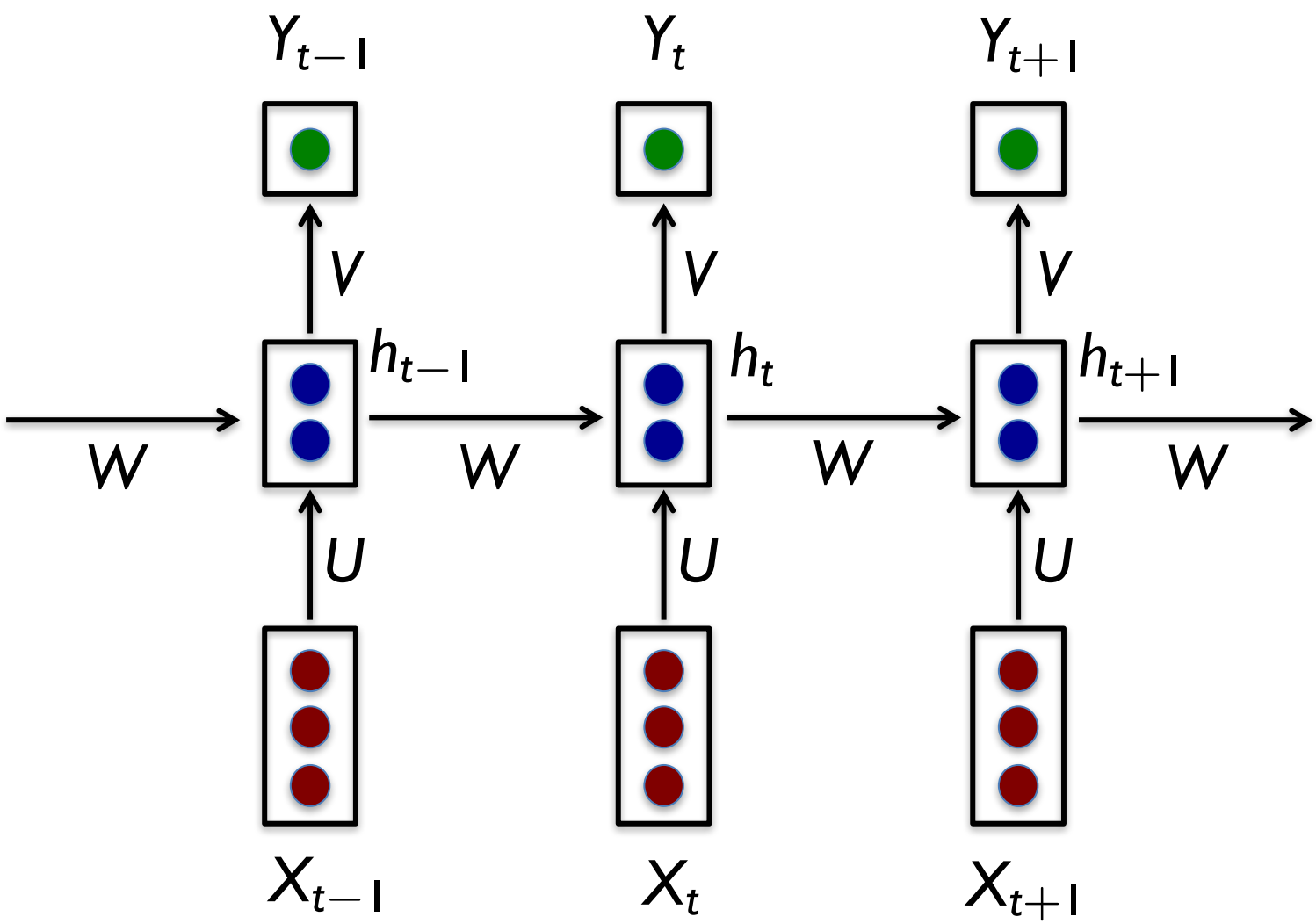
Bedroom or living room?

Generative model

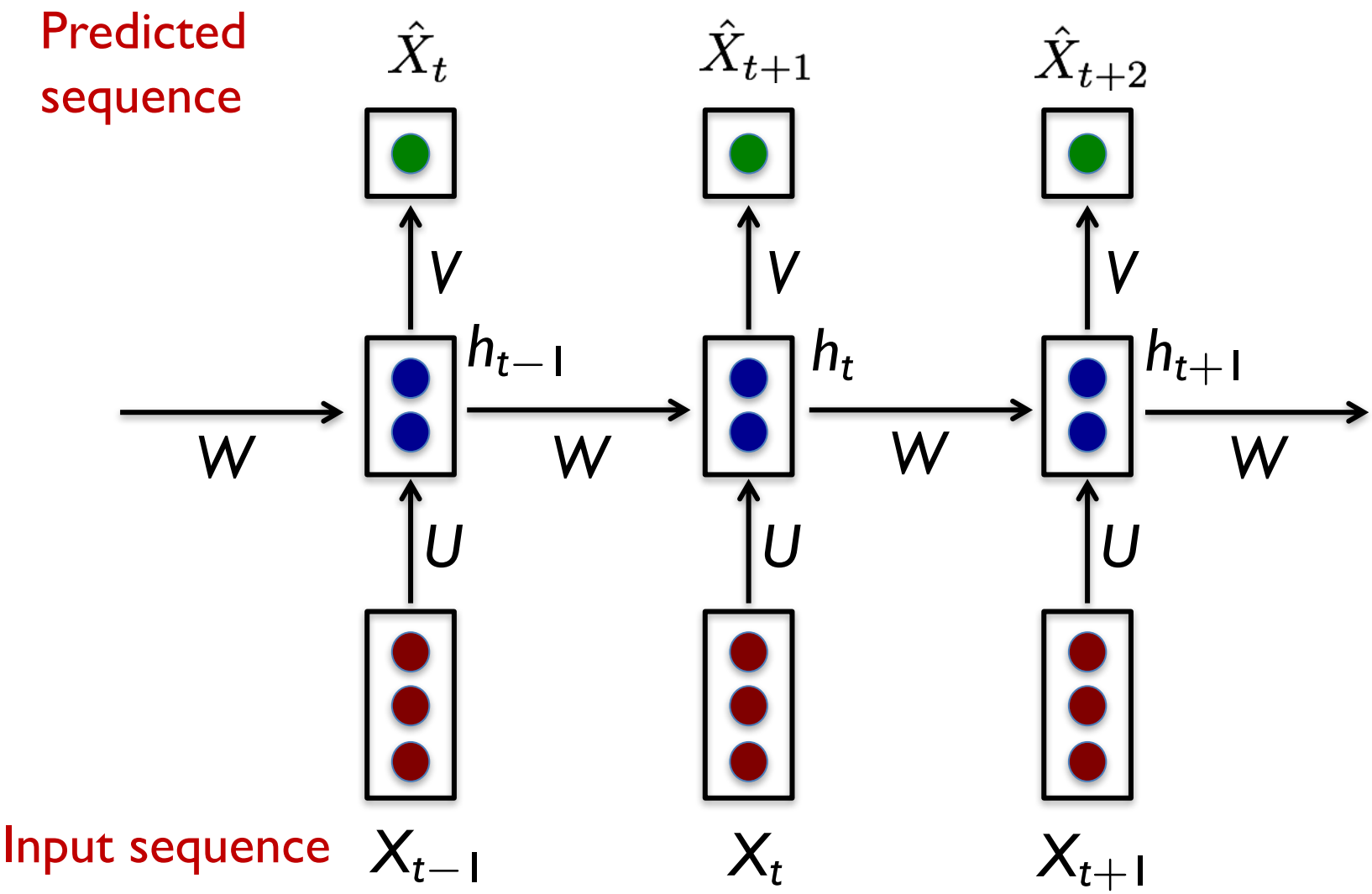
Generate new bedrooms



Recurrent neural network

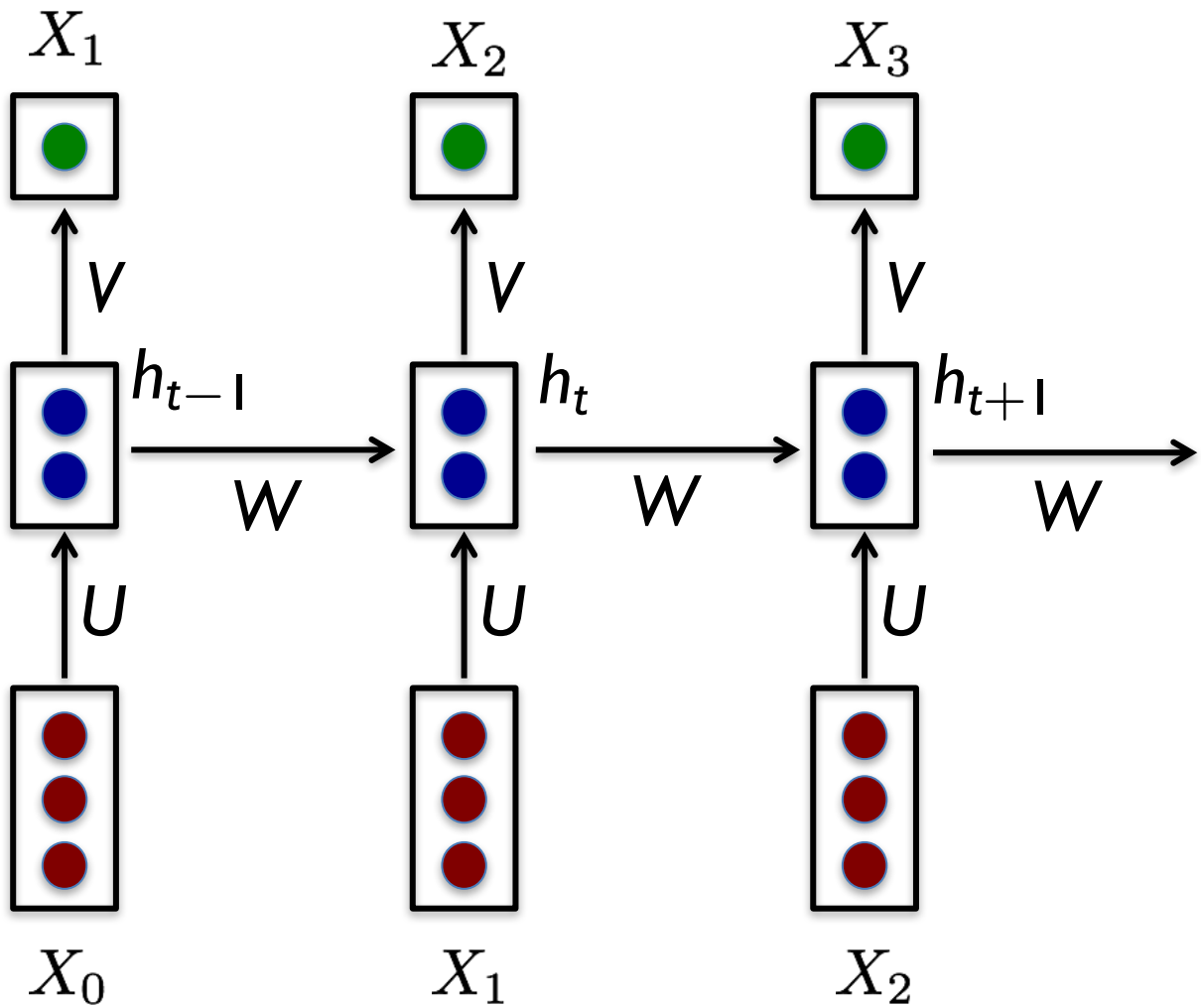


RNN: predict the next symbol



RNN to generate sequence

Generated sequence



Random initialization

RNN to generate molecules

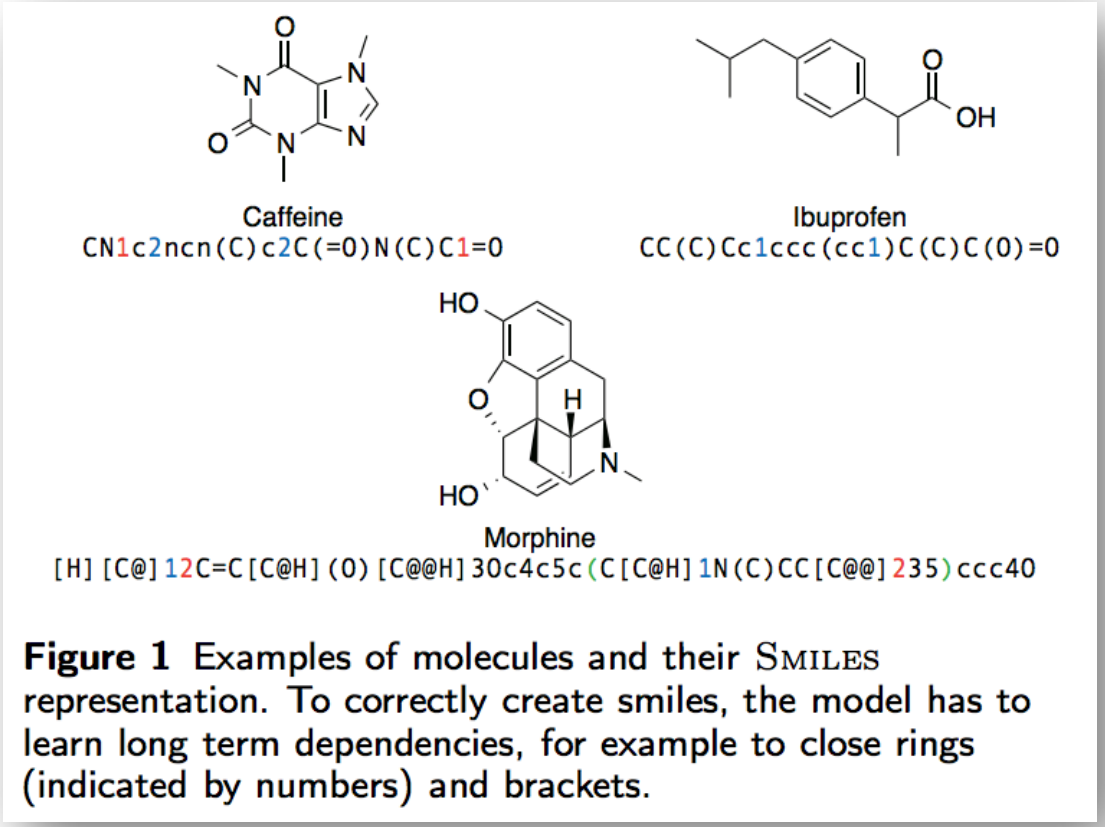
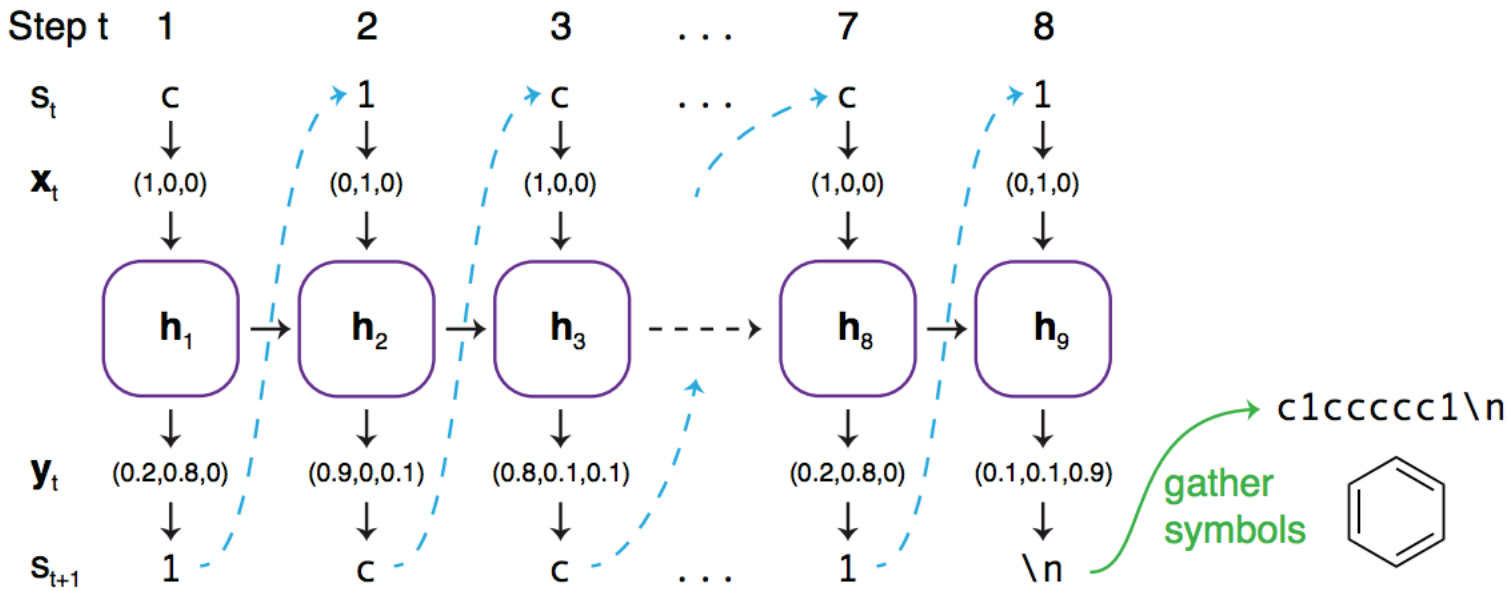


Figure 1 Examples of molecules and their SMILES representation. To correctly create smiles, the model has to learn long term dependencies, for example to close rings (indicated by numbers) and brackets.

Simplified molecular-input line-entry system (SMILES)

Segler et al 2017

RNN to generate molecules



RNN generative process: 3 stacked LSTM

Segler et al 2017

Does the RNN generate valid molecules?

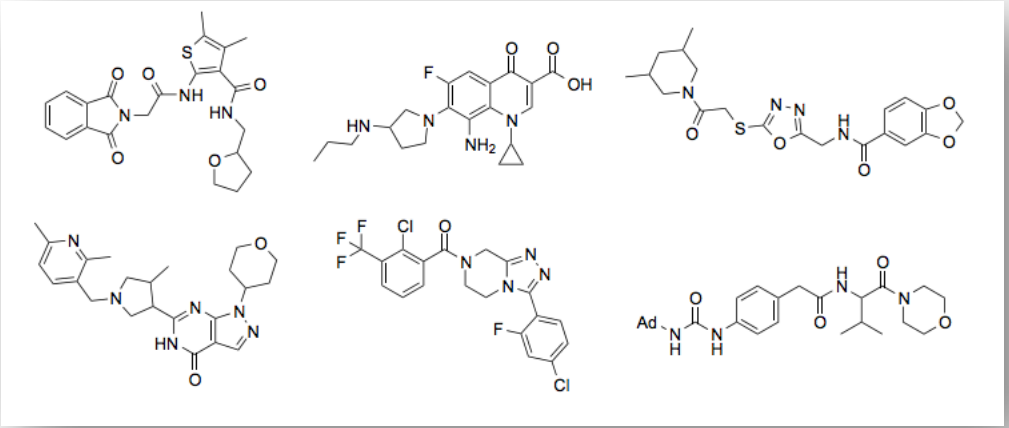
Generate sequences from different training stages of the RNN.

Batch	Generated Example	valid
0	`Oc.BK5i%ur+7oAFc7L3T=F8B5e=n)CS6RCTAR((OVCp1CApb)`	no
1000	`OF=CCC2OCCCC)C2)C1CNC2CCCCCCCCCCCCCCCCCCCCCCC`	no
2000	`O=C(N)C(=O)N(c1occc1OC)c2ccccc2OC`	yes
3000	`O=C1C=2N(c3cc(ccc3OC2CCC1)CCCc4cn(c5c(Cl)cccc54)C)C`	yes

Segler et al 2017

Examples of generated molecules

A sample of generated molecules

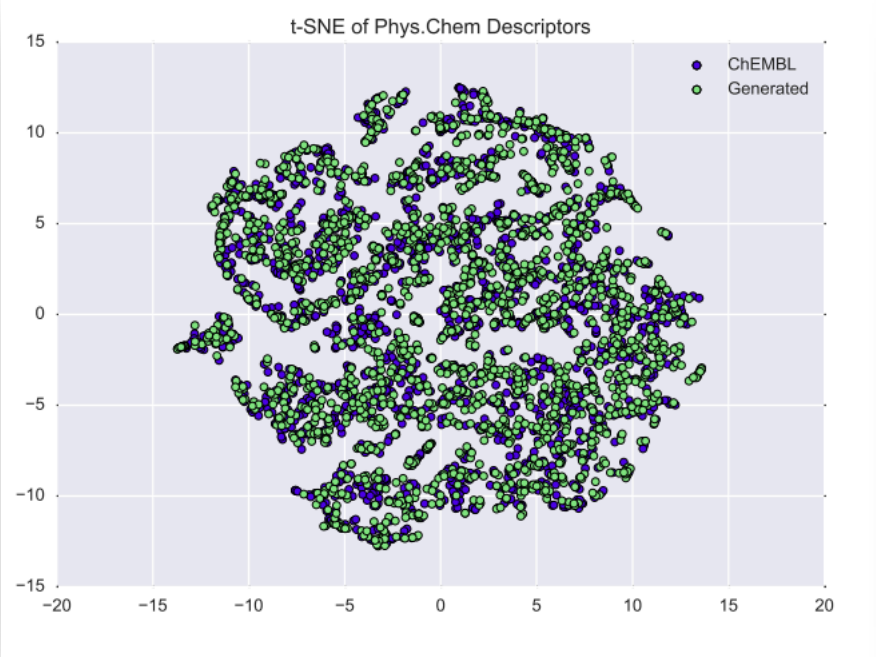


Q: how do we evaluate the quality of this generative model?

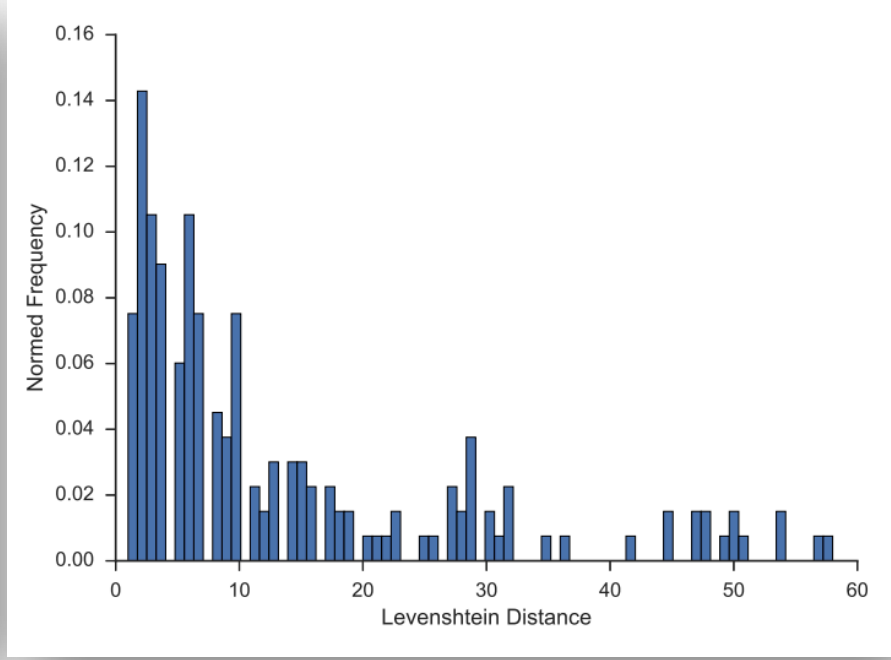
Segler et al 2017

Similarities to existing molecules

t-SNE visualization of new and old molecules

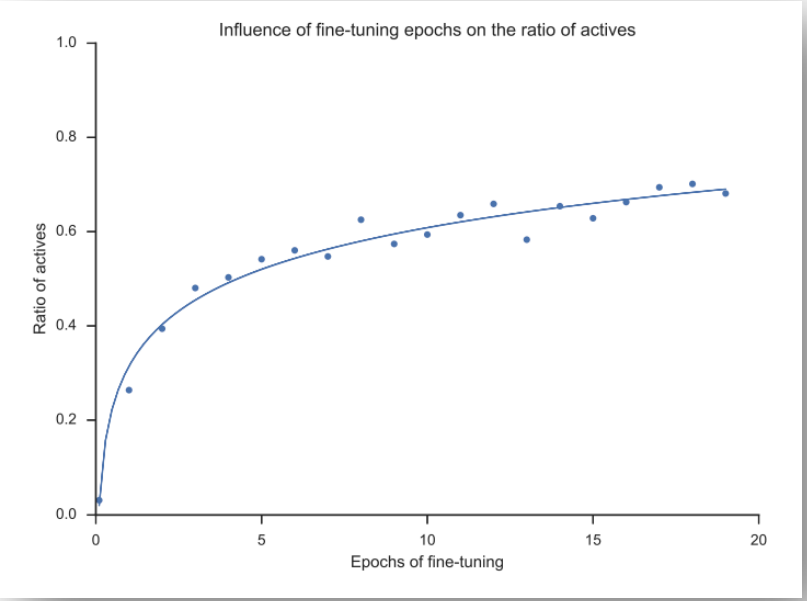


Edit distance between new and old molecules



Segler et al 2017

Fine-tuning to generate desirable properties



1. Train a base RNN on all the molecules.
2. Fine-tune on subset (732) of molecules active for a target receptor.

Segler et al 2017

Fine-tuning to generate desirable properties

What if we don't have examples of molecules with desirable property?

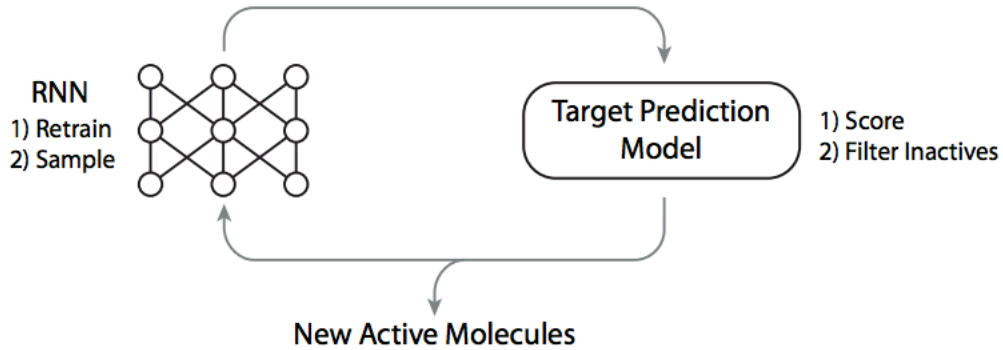
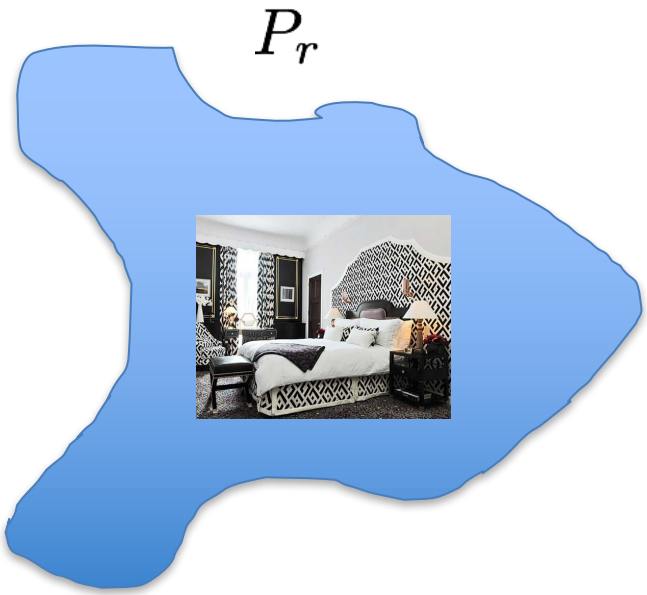


Figure 10 Scheme of our *de novo* design cycle. Molecules are generated by the chemical language model and then scored with the target prediction model (TPM). The inactives are filtered out, and the RNN is retrained. Here, the TPM is a machine learning model, but it could also be a robot conducting synthesis and biological assays, or a docking program.

Segler et al 2017

Generative models as mimicry

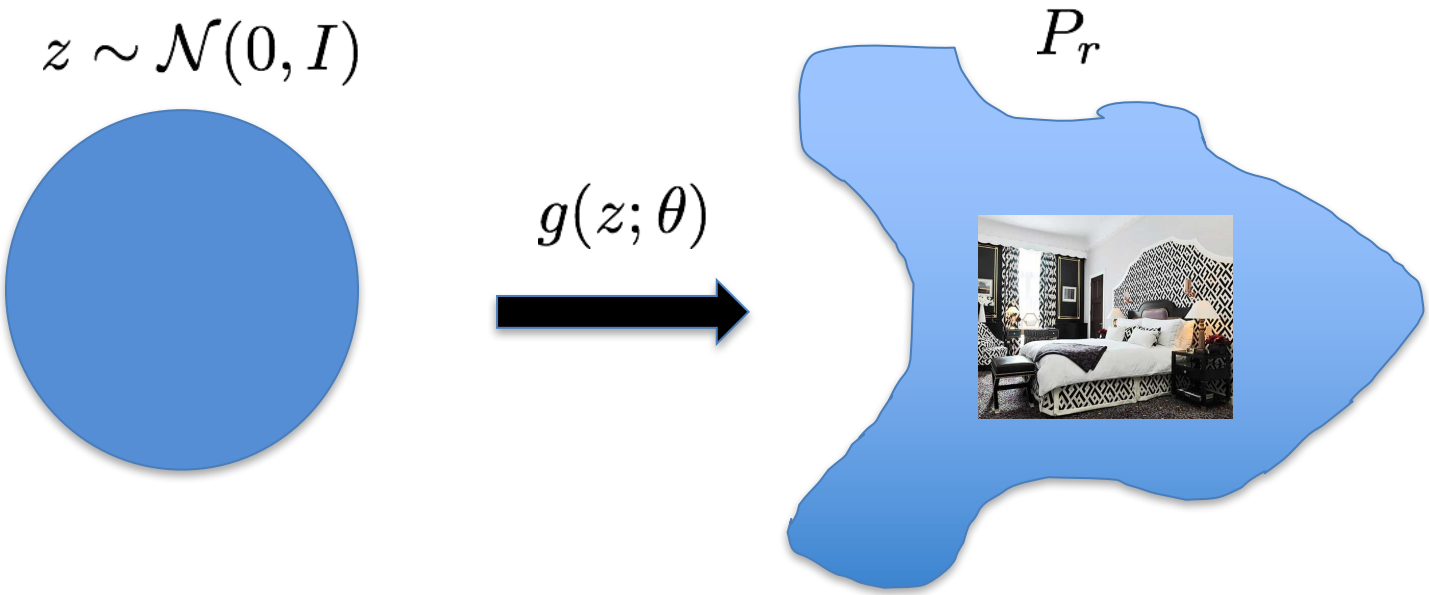
The probability distribution of bedrooms is complex.



Generative models as mimicry

Approximate it with a simple distribution: Gaussian

The probability distribution of bedrooms is complex.



Generative models as mimicry

- Given samples from complex target distribution P_r

- We construct an approximate generative model to P_θ by:
  1. sample $z \sim \mathcal{N}(0, I)$
  2. compute $x = g(z; \theta)$

- x is our approximate sample

- Q: how close is P_θ to P_r ?

Measuring distance between distributions

L_1 distance $\delta(P_r, P_\theta) = \frac{1}{2} \int |P_r(x) - P_\theta(x)| dx$

KL-divergence $KL(P_r || P_\theta) = \mathbb{E}_{P_r}[\log P_r / P_\theta]$.

Earthmover distance $W(P_r, P_\theta)$: the cost of moving P_θ to P_r.

Measuring distance between distributions

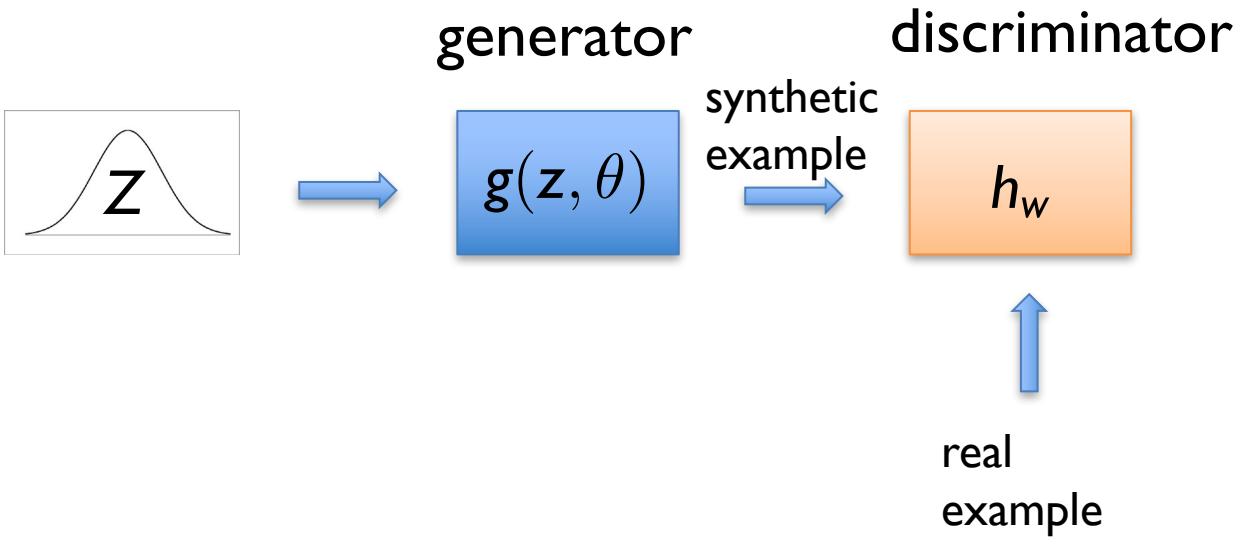
Earthmover (Wasserstein) distance $W(P_r, P_\theta)$

$$W(P_r, P_\theta) \approx \max_w \underbrace{\mathbb{E}_{X \sim P_r}[f_w(X)]}_{\text{discriminator score on real examples}} - \underbrace{\mathbb{E}_{Z \sim N(0,I)}[f_w(g(Z;\theta))]}_{\text{discriminator score on fake examples}}$$

f_w is the discriminator (f_w is a 1-Lipshitz function)

Goal: find θ so that $W(P_r, P_{theta})$ is minimized.

Generative adversarial model



Competition between g_θ and h_w.

f_w aims to maximize $\mathbb{E}_{X \sim P_r}[f_w(X)] - \mathbb{E}_{Z \sim N(0,I)}[f_w(g(Z;\theta))]$

g_θ aims to minimize $\mathbb{E}_{X \sim P_r}[f_w(X)] - \mathbb{E}_{Z \sim N(0,I)}[f_w(g(Z;\theta))]$

Generative adversarial model

WGAN training algorithm

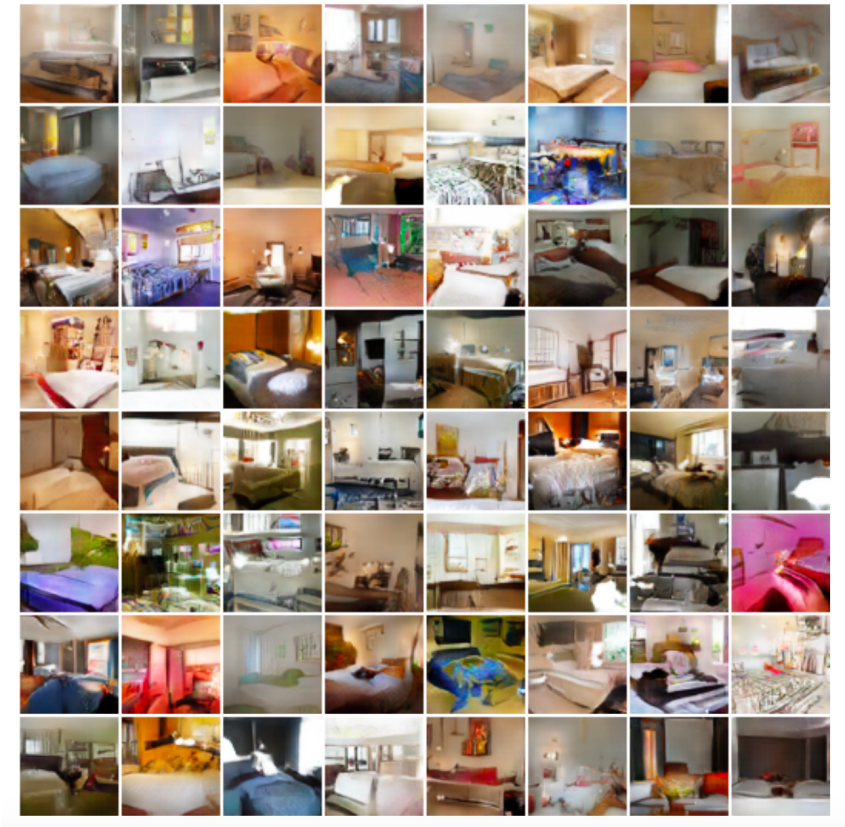
Input: $\{X^{(i)}\} \sim P_r$
Output: θ

Outerloop: optimize θ
    Innerloop: optimize w
        Sample batch $\{X^{(i)}\}$
        Sample $\{Z^{(i)}\} \sim P(Z) = N(0, I)$
        Compute $\nabla_w [\frac{1}{m} \sum f_w(x^{(i)}) - \frac{1}{m} \sum f_w(g_\theta(z^{(i)}))]$
        Update w: $w \leftarrow$ threshold w (to ± 0.01 say)
    Sample $\{Z^{(i)}\} \sim N(0, I)$
    Compute $-\nabla_\theta \frac{1}{m} \sum f_w(g_\theta(z^{(i)}))$
    Update θ

Generative adversarial model

real example

synthetic examples



Generative adversarial model

real example

synthetic examples

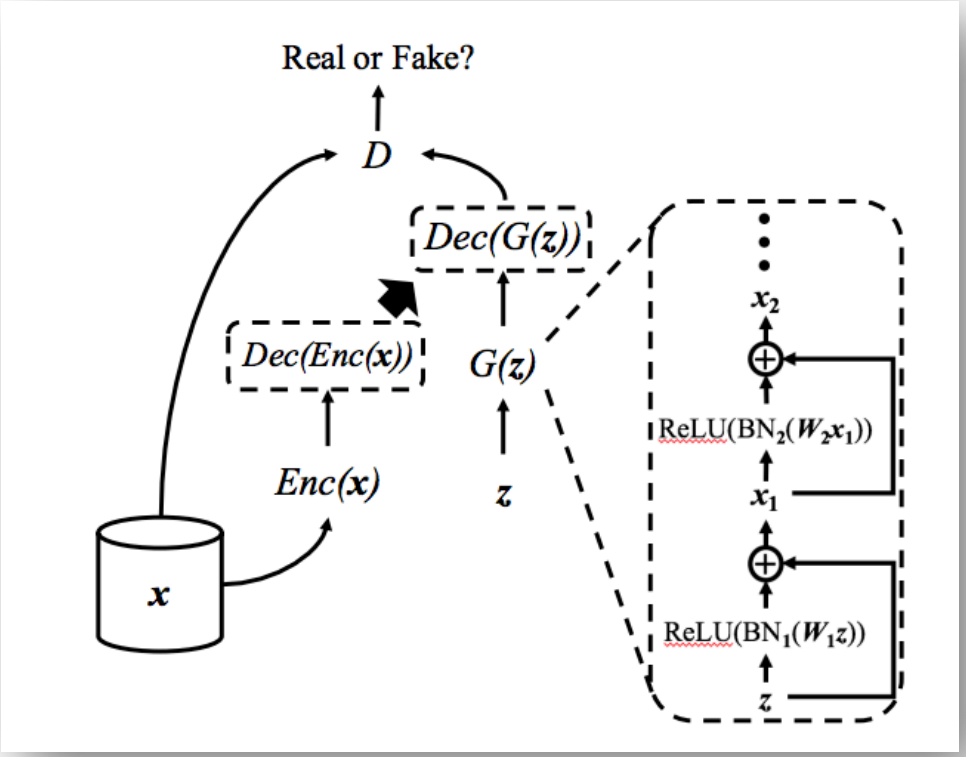
GAN to generate medical records

Dataset	Sutter PAMF	MIMIC-III
# of patients	258,559	46,520
# of unique codes	615	1071
Avg. # of codes per patient	38.37	11.27
Max # of codes for a patient	198	90
Min # of codes for a patient	1	1

For each patient: 615 dimensional vector of binary values and counts.

Choi et al 2017

GAN to generate medical records



Choi et al 2017

GAN to generate medical records

Histogram of counts for the most frequent diseases.

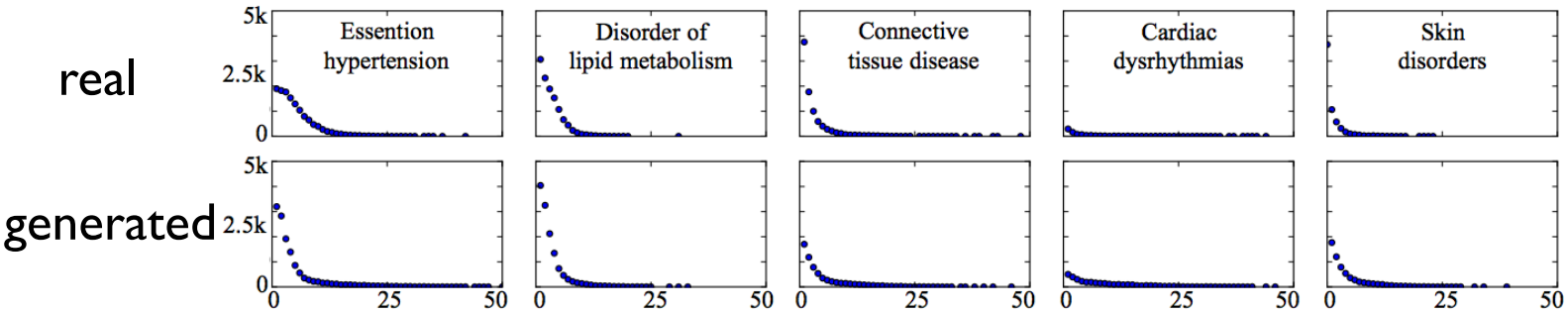
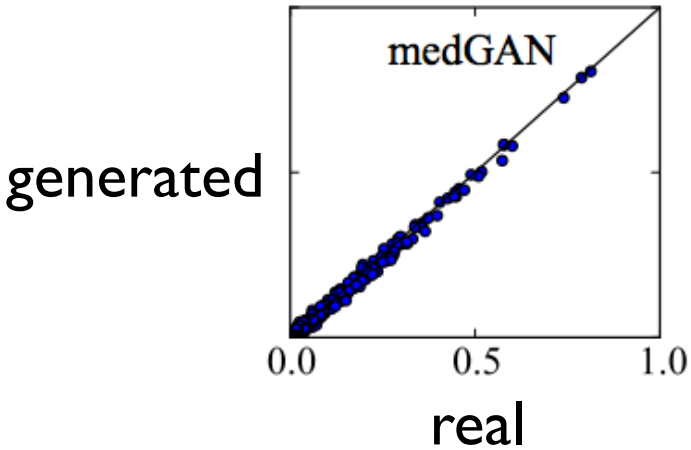


Figure 4: Histogram of counts of five most frequent codes from the Sutter heart failure dataset. The top row was plotted using the real data, the bottom row using medGAN's synthetic data.

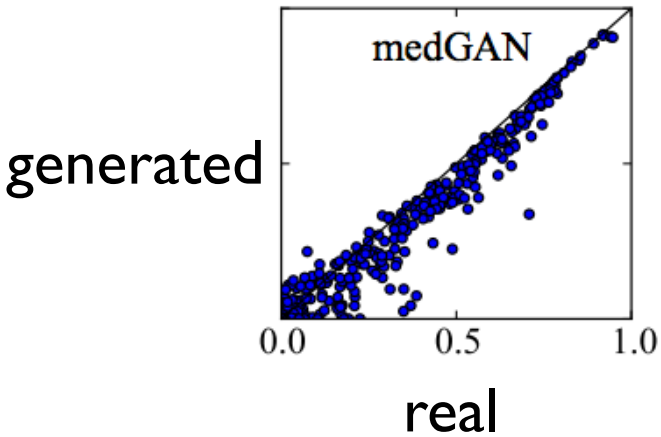
Choi et al 2017

GAN to generate medical records

Avg. count of each code



F_1 score



Choi et al 2017

GAN to generate medical records

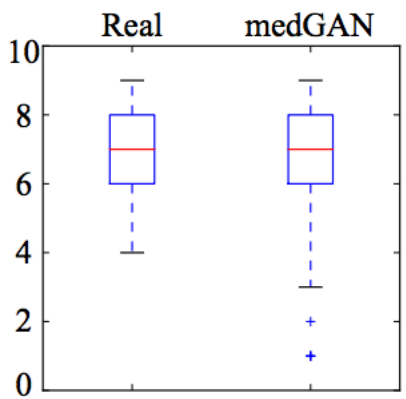


Figure 7: Boxplot of the impression scores from a medical expert. The synthetic data from **medGAN** provides comparable impression scores in terms of realisticness level.

Choi et al 2017