Generative Models
Discriminative vs Generative

Understanding Machine Learning

Archetype: Naïve Bayes

probabilistic model:

$$P(Y|X_1,\cdots,X_p) = \frac{P(Y,X_1,\cdots,X_p)}{P(X_1,\cdots,X_p)} = \frac{P(Y)P(X_1,\cdots,X_p|Y)}{P(X_1,\cdots,X_p)} \propto P(Y)P(X_1,\cdots,X_p|Y)$$

Bayes' rule   constant   to be estimated

approach:

1. estimate $P(Y,\boldsymbol{X})$ → generative model (can be used to generate new samples)
2. calculate $P(Y|\boldsymbol{X})$ from $P(Y,\boldsymbol{X})$ → used for discriminative task (classification)

Independence Assumption

(naïve) assumption: conditional independence of features given target

$$P\big(X_j|Y, X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_p\big) = P\big(X_j|Y\big)$$

$$\implies P\big(Y|X_1, \cdots, X_p\big) = \frac{P(Y)\prod_{j=1}^{p} P\big(X_j|Y\big)}{P\big(X_1, \cdots, X_p\big)}$$
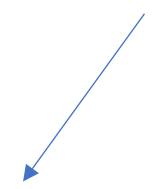
→ independent feature contributions (ignoring feature correlations)
→ robust against curse of dimensionality

Estimation of Feature Contributions

separate estimations of $P(X_j|Y)$ for each feature

requires assumption of distributions (e.g., Gaussian naïve Bayes) or non-parametric methods (kernel density estimation)

Gaussian feature likelihoods: $\quad P(x_{ij}|y) = \dfrac{1}{\sqrt{2\pi\sigma_{y,j}^2}} \exp\left(-\dfrac{(x_{ij}-\mu_{y,j})^2}{2\sigma_{y,j}^2}\right)$

parameter estimation (e.g., mean and variance of Gaussians) can be done with maximum likelihood method (y known in training)

→ no Bayesian methods needed

Maximum a Posteriori Classification

$$\hat{y}_i = \underset{y}{\mathrm{argmax}}\, P(y) \prod_{j=1}^{p} P(x_{ij}|y)$$

despite potentially inaccurate probability estimates (due to naïve independence assumption), good identification of correct class via maximum probability

→ bad for regression tasks (if independence assumption is too naïve, i.e., features are correlated)
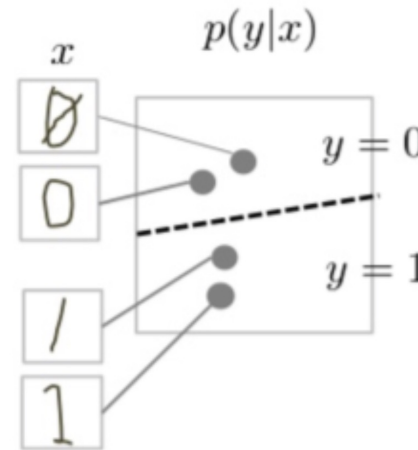
Generative vs Discriminative Models

generative models: predict joint probability $P(Y, \boldsymbol{X})$ (what allows to create new data samples) or directly generates new data samples
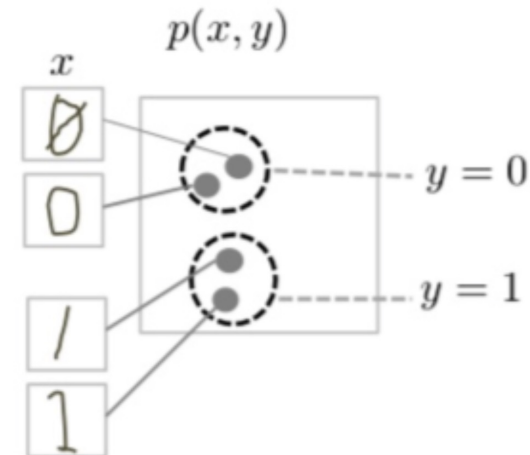
discriminative models: predict conditional probability $P(Y|\boldsymbol{X})$ or directly output (label for classification, real value for regression)

task of generative models more difficult: model full data distribution rather than merely find patterns in inputs to distinguish outputs

discriminative model

x $p(y|x)$

$y = 0$

$y = 1$

generative model

x $p(x, y)$

$y = 0$

$y = 1$

source

Naïve Bayes and Logistic Regression

generative-discriminative pair of classification algorithms

- binary case: logit of naïve Bayes' outputs, $\log\left(\frac{P(y_i=1|\boldsymbol{x}_i)}{P(y_i=0|\boldsymbol{x}_i)}\right)$, corresponds to output of logistic regression's linear predictor
- for discrete inputs or Gaussian naïve Bayes: naïve Bayes can be reparametrized as linear classifier

for discriminative task: identical in asymptotic limit (infinite training samples) if independence assumption holds (otherwise naïve Bayes less accurate)

naïve Bayes has greater bias but lower variance than logistic regression $\rightarrow$ to be preferred for scarce training data (if bias, i.e., independence assumption, correct)
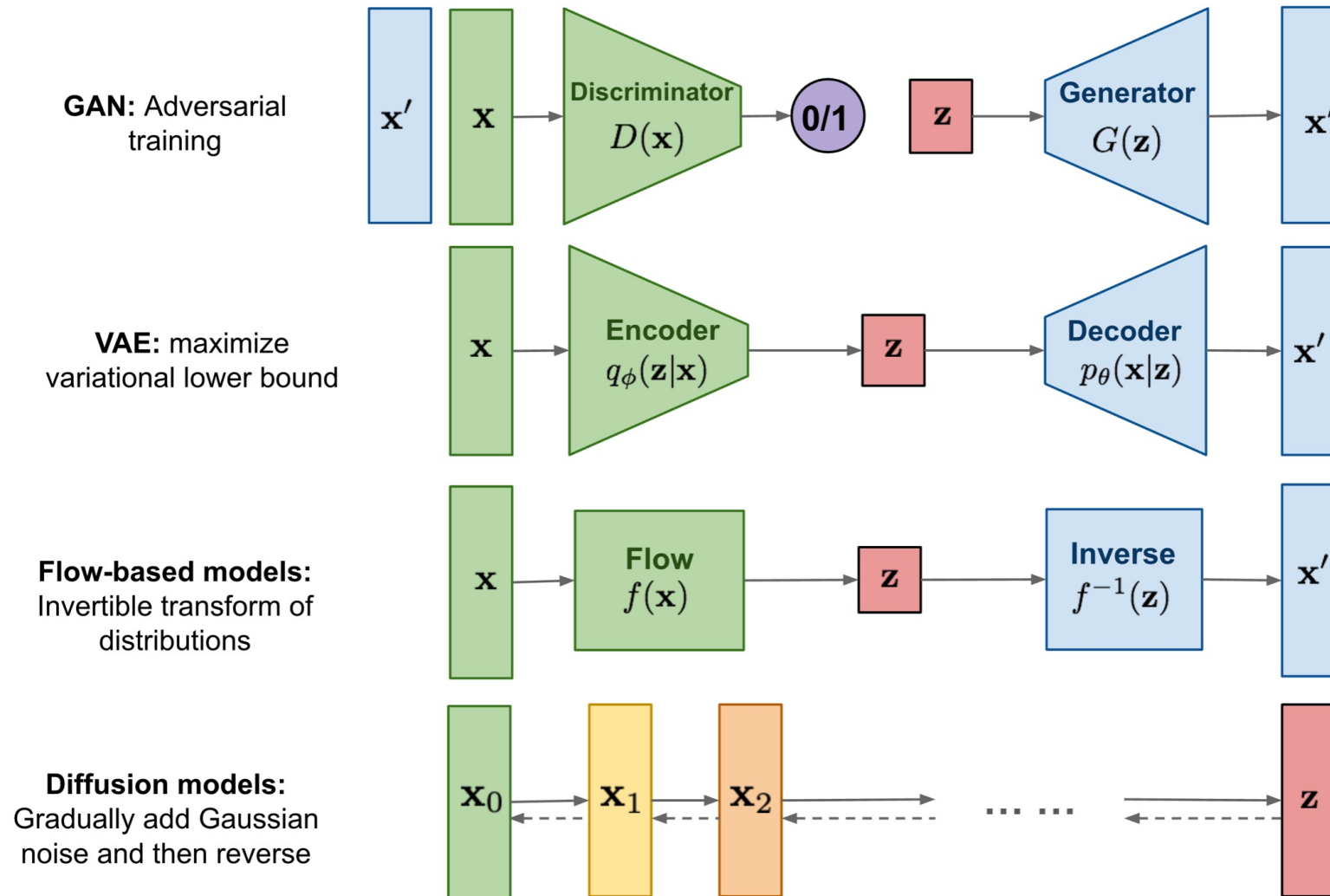
Data Generation

generative models can be used for discriminative tasks (although potentially inferior to direct discriminative methods)

but generative methods do more than discriminative ones: model full data distribution

→ allows generation of new data samples (can be images, text, video, audio, proteins, materials, time series, structured data, …)

large (auto-regressive) language models examples of generative models

Different Types of Generative Models

GAN: Adversarial training

$\mathbf{x}'$ | $\mathbf{x}$ → **Discriminator** $D(\mathbf{x})$ → 0/1 | $\mathbf{z}$ → **Generator** $G(\mathbf{z})$ → $\mathbf{x}'$

VAE: maximize variational lower bound

$\mathbf{x}$ → **Encoder** $q_\phi(\mathbf{z}|\mathbf{x})$ → $\mathbf{z}$ → **Decoder** $p_\theta(\mathbf{x}|\mathbf{z})$ → $\mathbf{x}'$

Flow-based models: Invertible transform of distributions

$\mathbf{x}$ → **Flow** $f(\mathbf{x})$ → $\mathbf{z}$ → **Inverse** $f^{-1}(\mathbf{z})$ → $\mathbf{x}'$

Diffusion models: Gradually add Gaussian noise and then reverse

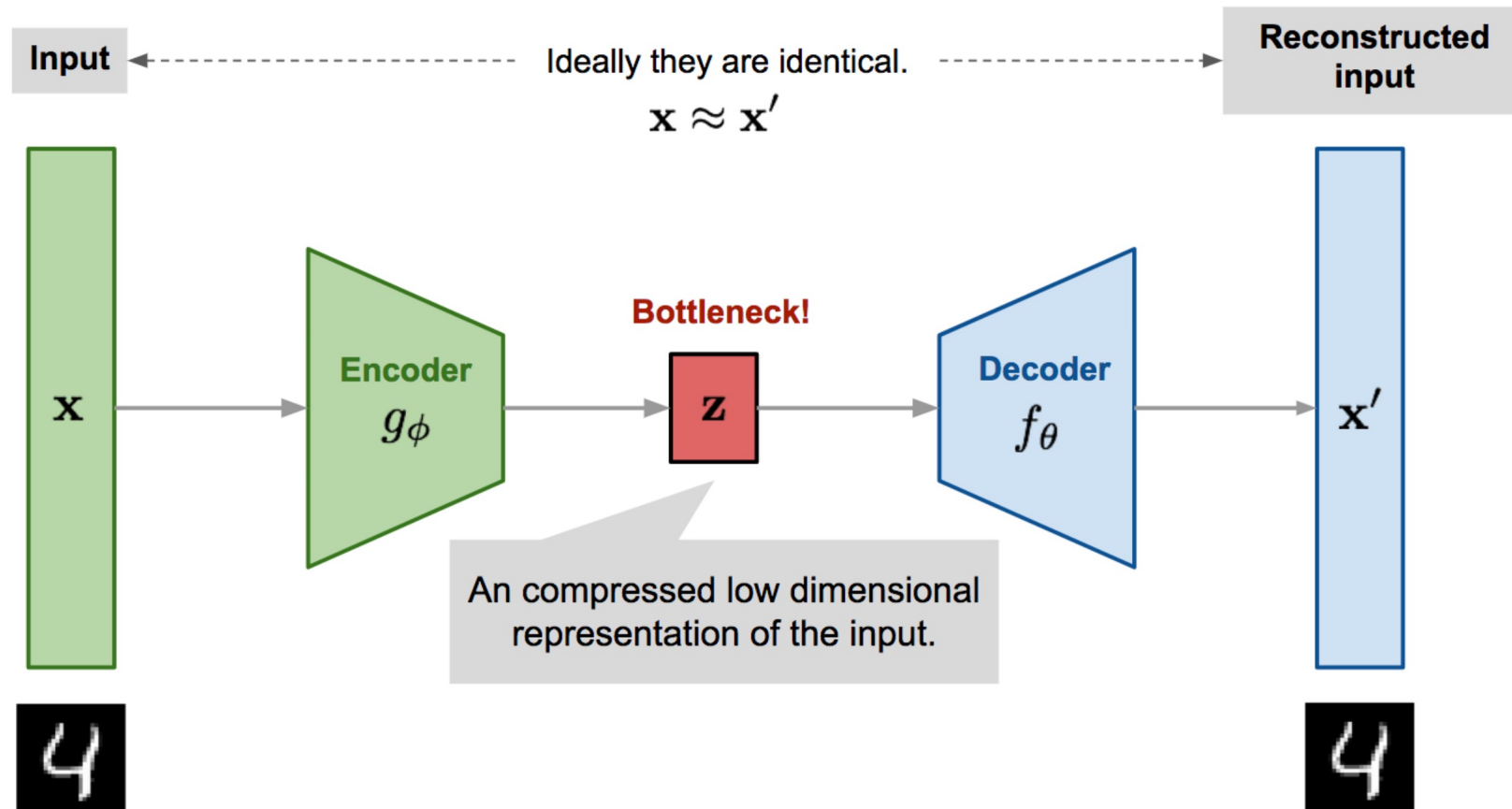$\mathbf{x}_0$ → $\mathbf{x}_1$ → $\mathbf{x}_2$ → ... ... → $\mathbf{z}$

Variational Inference

- Bayesian …

Recap: Autoencoder

(deep) encoder network

(deep) decoder network

learned together by minimizing differences between original input and reconstructed input (expressed as losses)
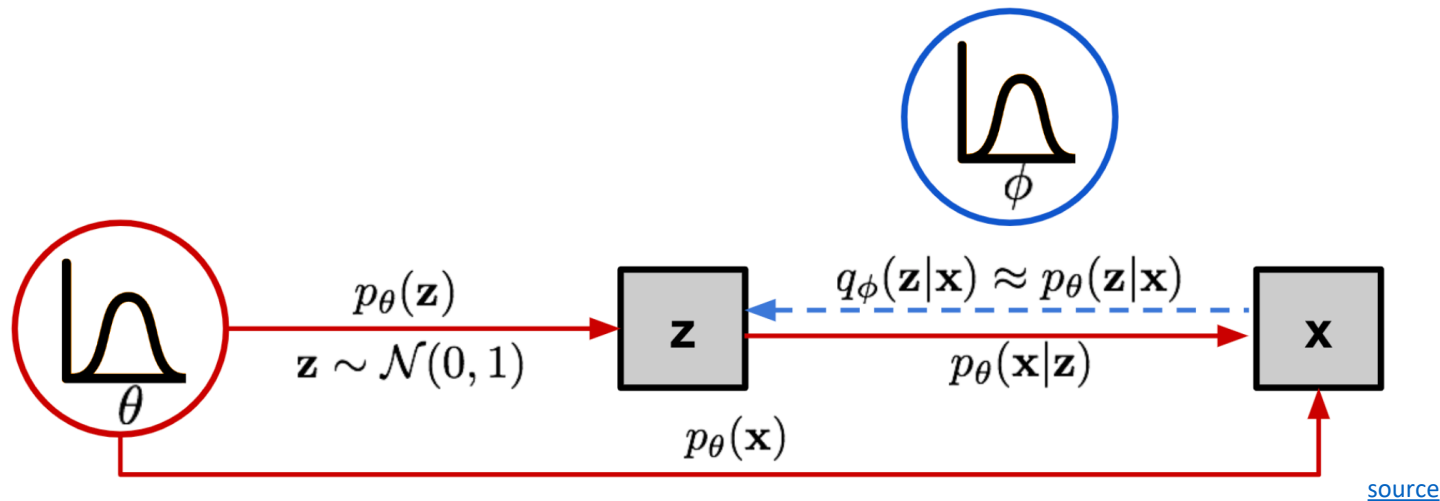
compressed intermediate representation: dimensionality reduction

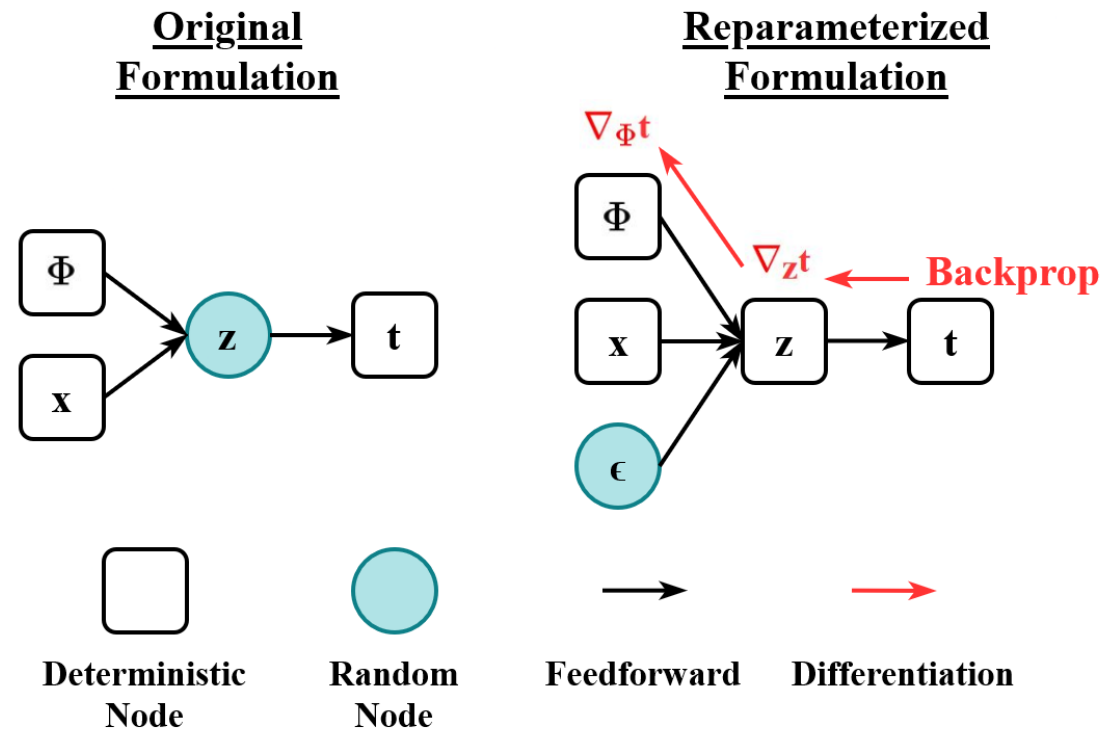Input ← - - - - - - - - - - - - Ideally they are identical. - - - - - - - → **Reconstructed input**

$$x \approx x'$$

Bottleneck!

x → **Encoder** g_ϕ → z → **Decoder** f_θ → x'

An compressed low dimensional representation of the input.

Variational Autoencoder (VAE)

- …
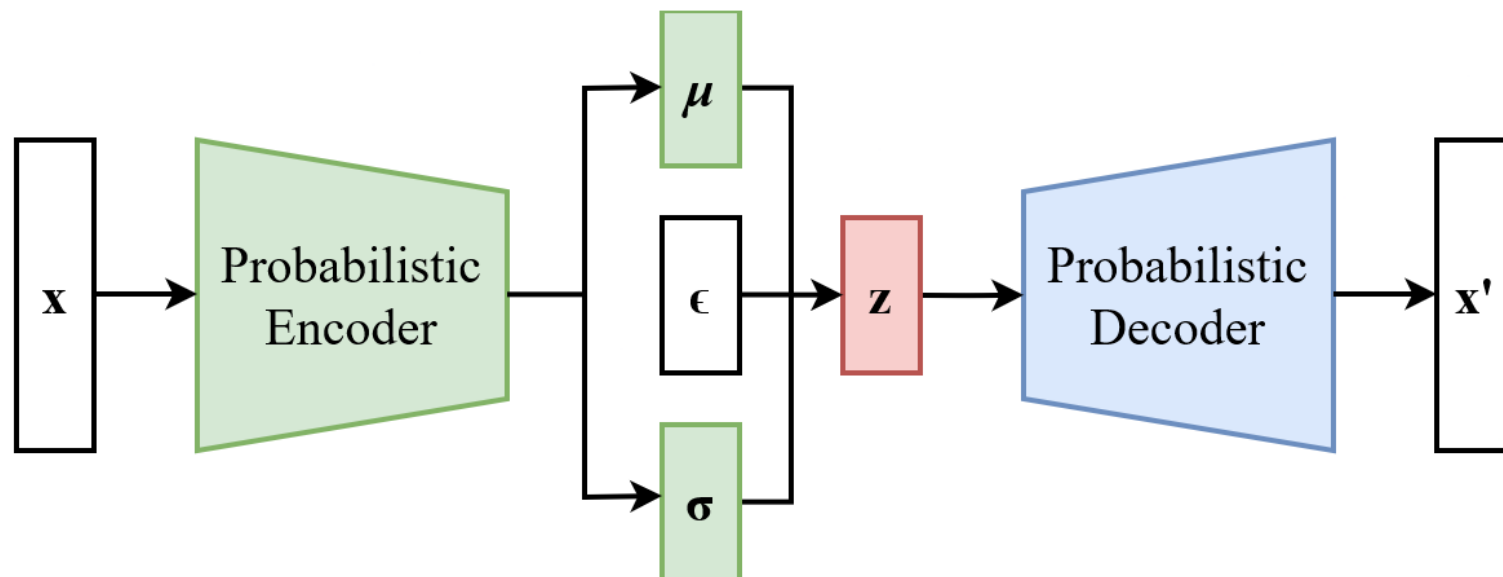- VAE relies on a surrogate loss

ELBO

- reparameterization trick

from wikipedia

15

- …

from wikipedia

Generative Adversarial Networks (GAN)

- ...

- GAN models are known for potentially unstable training and less diversity in generation due to their adversarial training nature.

Flow-Based Methods

- …

- Flow models have to use specialized architectures to construct reversible transform.

Energy-Based Methods

- …

Diffusion Models

- …

- Diffusion models are inspired by non-equilibrium thermodynamics. They define a Markov chain of diffusion steps to slowly add random noise to data and then learn to reverse the diffusion process to construct desired data samples from the noise. Unlike VAE or flow models, diffusion models are learned with a fixed procedure and the latent variable has high dimensionality (same as the original data).

Image Generation

[DALL-E 2](#)

...

[Stable Diffusion](#)

[DreamStudio](#)

A dream of a distant galaxy, by Caspar David Friedrich, matte painting trending on artstation HQ

Literature

papers:

- variational autoencoder
- [GAN](GAN)
- normalizing flows
- latent diffusion

Movie-like Intelligence

emergent capabilities of complex systems almost impossible to foresee

mini examples in contemporary ML:

- large language models
- multi-agent reinforcement learning

one idea: reward is enough

philosophical: emotions or consciousness might also occur as emergent capabilities