Probabilistic Machine Learning

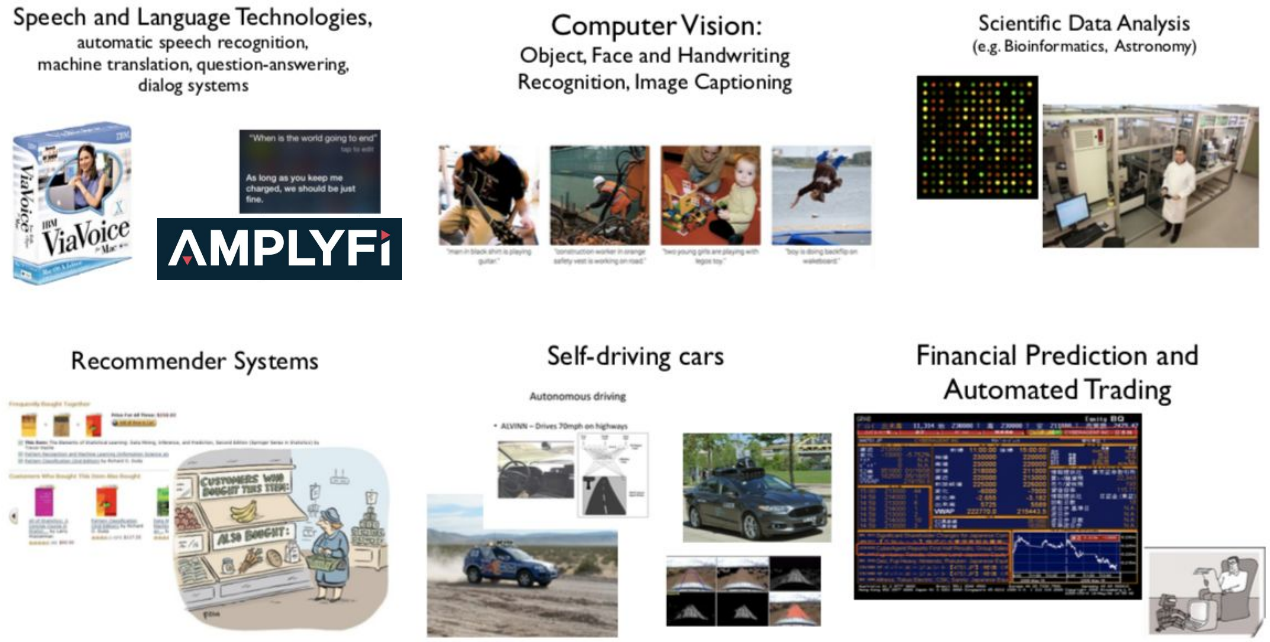
or...the importance of modeling unknowns

Yunus Saatchi, October 2018

yunus.saatchi@amplyfi.com

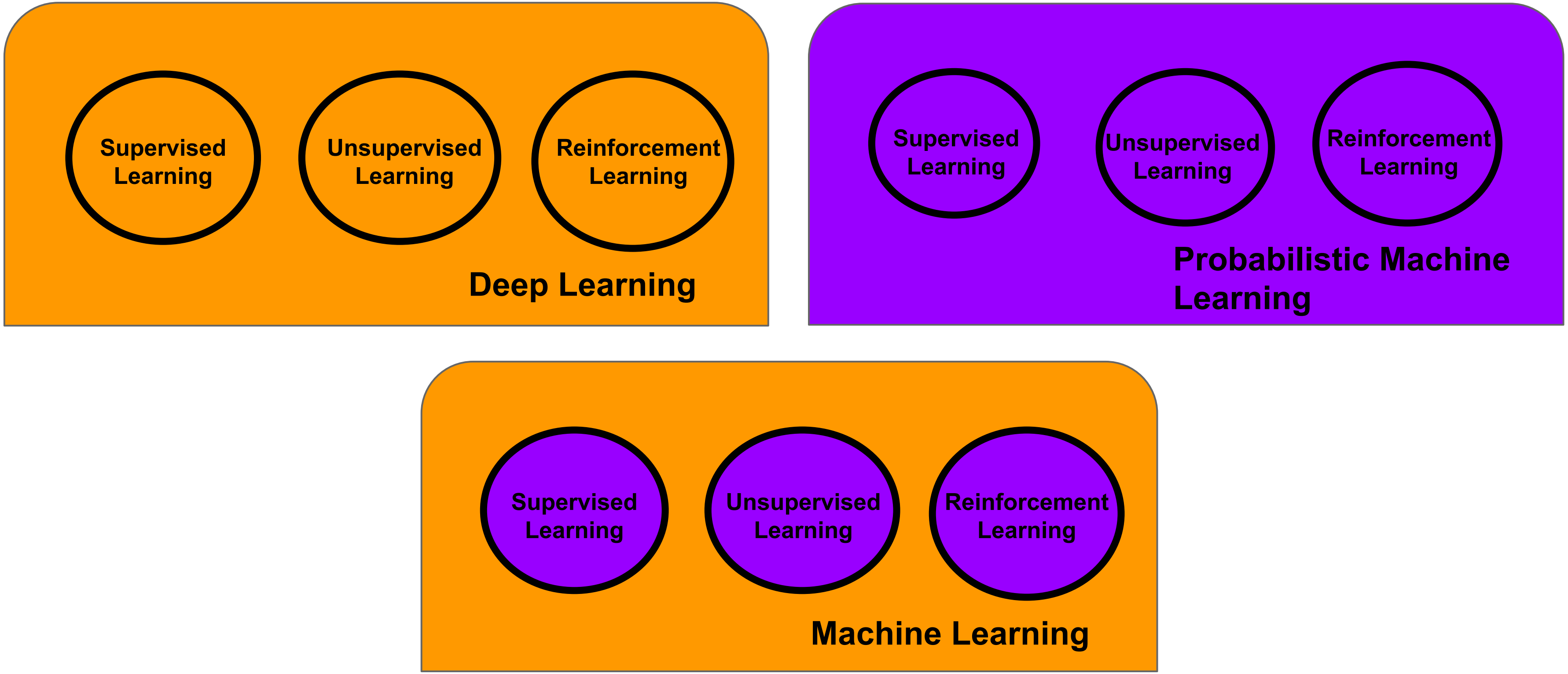
Machine Learning is HOT

Many application domains in almost all verticals of tech



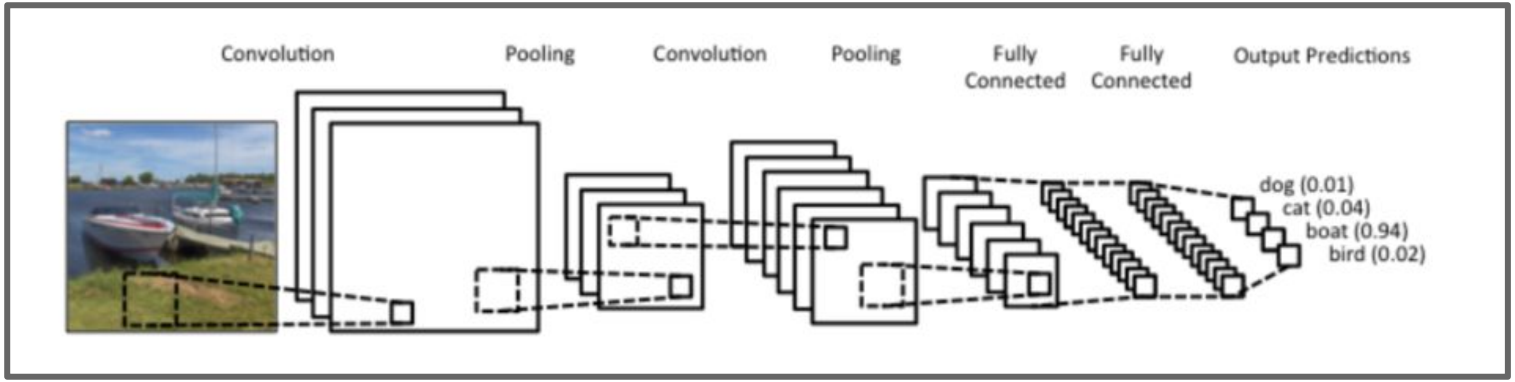
Machine Learning Concept Map

As of 2018...



Deep (Supervised) Learning

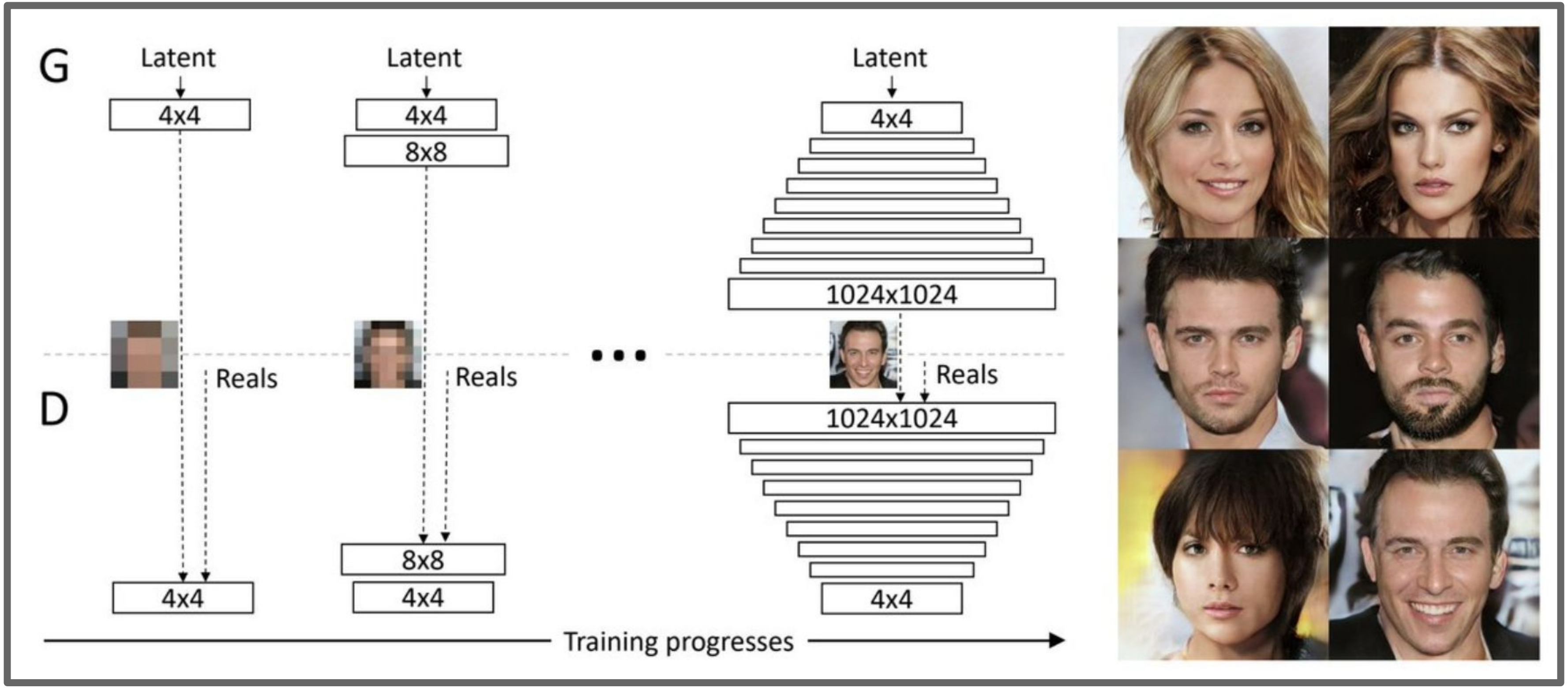
The darling child of Machine Learning



- Actually invented in the 80s and 90s!
- Become popular after a landmark paper by Krizhevsky et al. demonstrating state-of-art computer vision results. Key idea: Train it on a lot of data!
- Works very well with web-scale data.
- And a lot of computation.
- And some very good autodiff libraries: e.g. PyTorch, TensorFlow and MXNet.

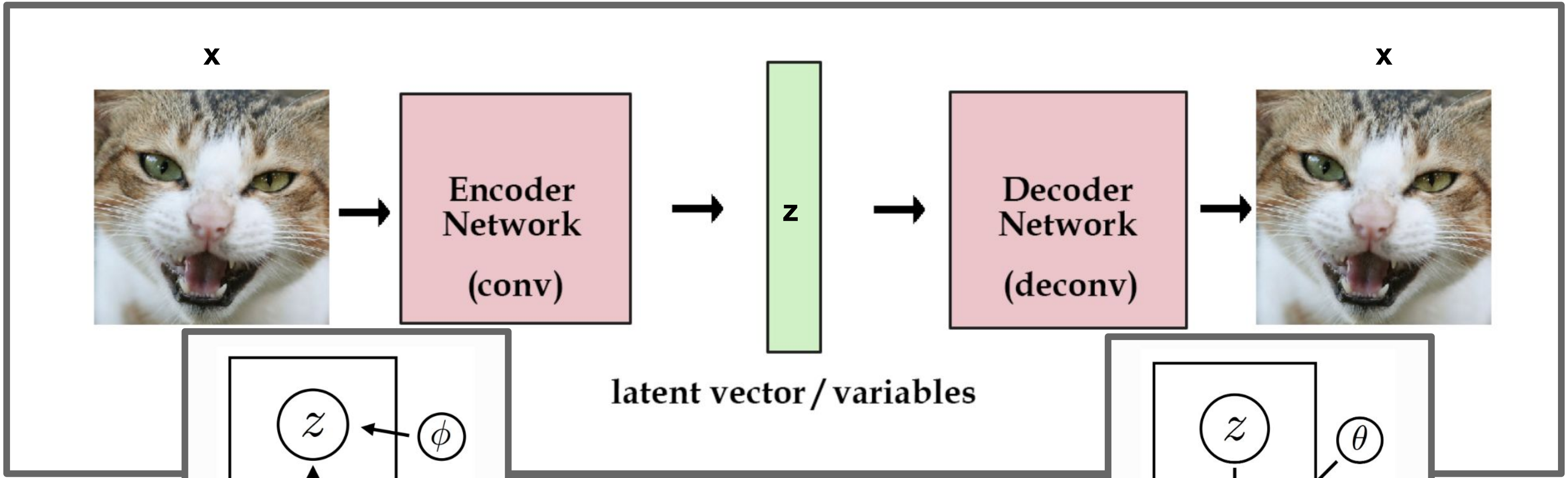
Deep Unsupervised/Semi-supervised Learning

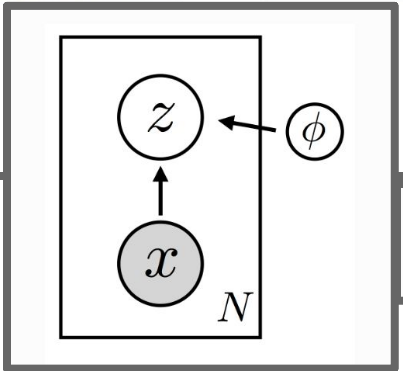
Generative Adversarial Networks



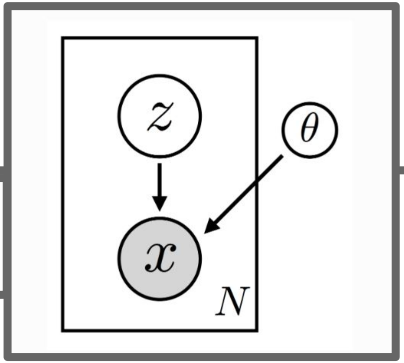
Deep Unsupervised/Semi-supervised Learning

Variational Autoencoders



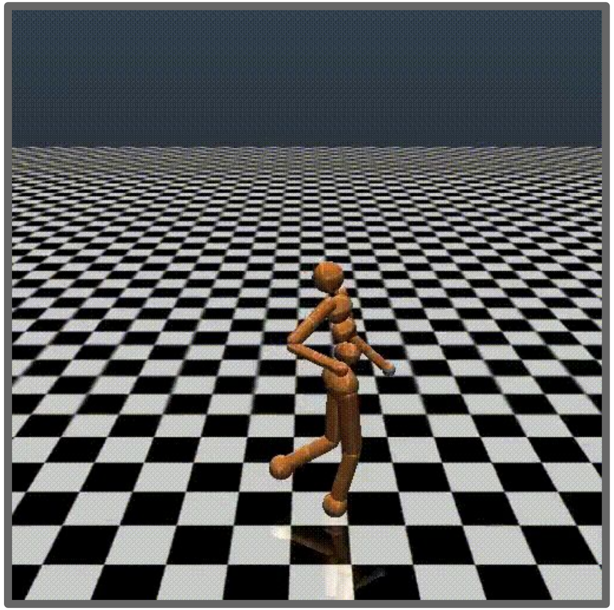


$$\text{ELBO} \equiv \mathbb{E}_{q_\phi(\mathbf{z})}\left[\log p_\theta(\mathbf{x}, \mathbf{z}) - \log q_\phi(\mathbf{z})\right]$$



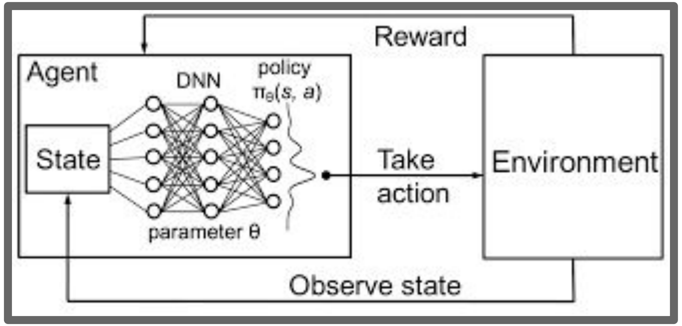
Deep Reinforcement Learning

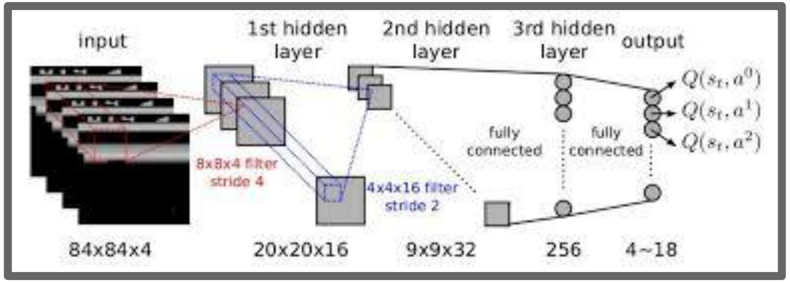
And the art of machine learning hype using games



Deep Reinforcement Learning

And the art of machine learning hype using games





$$Q^\pi(s_t, a_t) = \underline{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots | s_t, a_t]$$

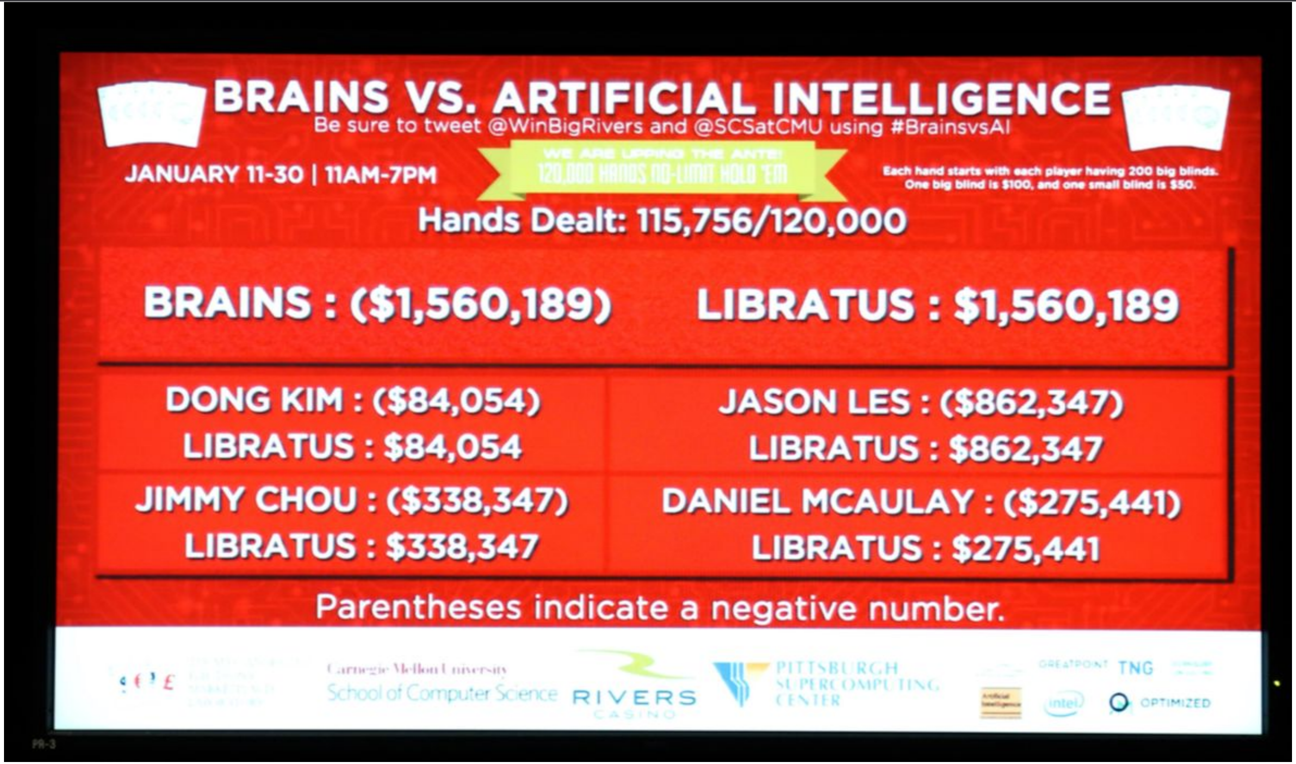
Q value for that state given that action

Expected discounted cumulative reward ...

given that state and that action

Deep Reinforcement Learning

But not all games...



Probabilistic Machine Learning

In one slide

Everything follows from two simple rules:

Sum rule: $P(x) = \sum_y P(x, y)$

Product rule: $P(x, y) = P(x)P(y|x)$

Learning:

$$P(\theta|\mathcal{D}, m) = \frac{P(\mathcal{D}|\theta, m)P(\theta|m)}{P(\mathcal{D}|m)}$$

$P(\mathcal{D}|\theta, m)$ likelihood of parameters θ in model m
$P(\theta|m)$ prior probability of θ
$P(\theta|\mathcal{D}, m)$ posterior of θ given data $\mathcal{D}$

Prediction:

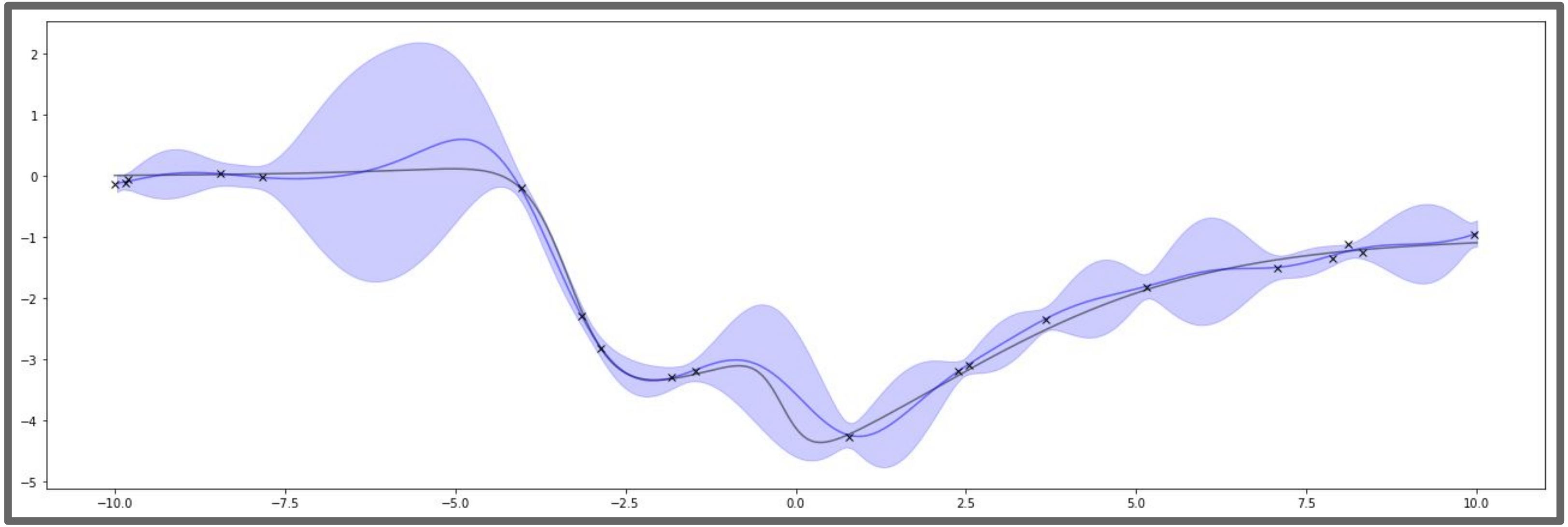
$$P(x|\mathcal{D}, m) = \int P(x|\theta, \mathcal{D}, m)P(\theta|\mathcal{D}, m)d\theta$$

Model Comparison:

$$P(m|\mathcal{D}) = \frac{P(\mathcal{D}|m)P(m)}{P(\mathcal{D})}$$

Probabilistic Supervised Learning

Example: Gaussian Processes



$$p(y|x, \mathcal{D})$$

Probabilistic Unsupervised Learning

Probabilistic Graphical Models & Probabilistic Programming

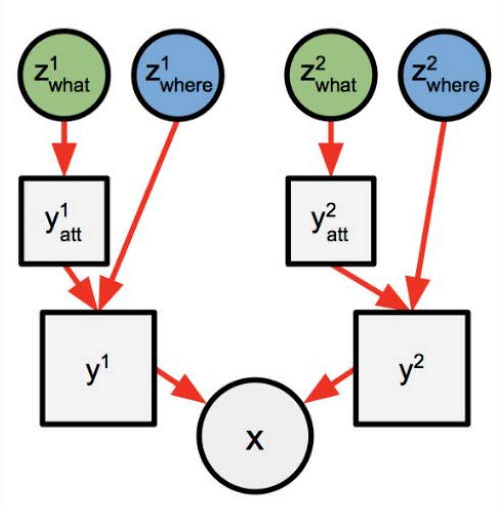
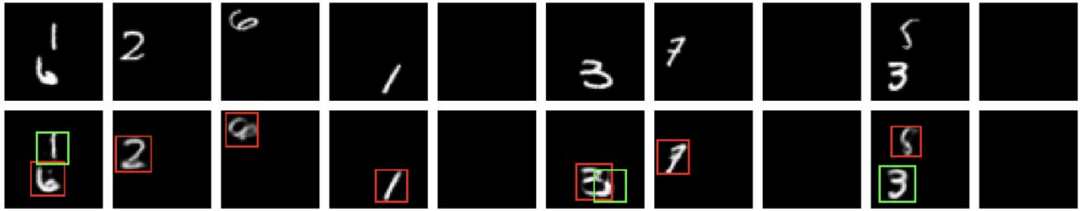


Figure 1: Two steps of the generative process.

```
data = mnist.view(-1, 50 * 50)

svi = SVI(model,
          guide,
          optim.Adam({'lr': 1e-4}),
          loss=TraceGraph_ELBO())

for i in range(5):
    loss = svi.step(data)
    print('i={}, elbo={:.2f}'.format(i, loss / data.size(0)))
```



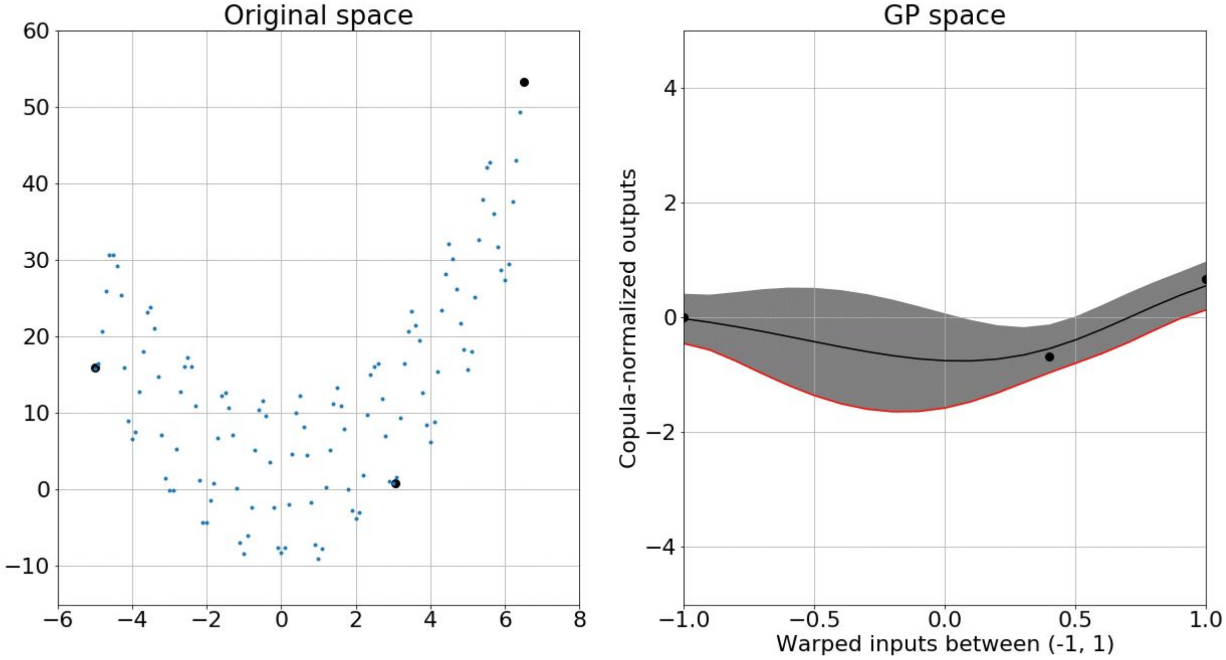
$$p(x|\mathcal{G})$$

Taken from: http://pyro.ai/examples/air.html

Probabilistic Reinforcement Learning

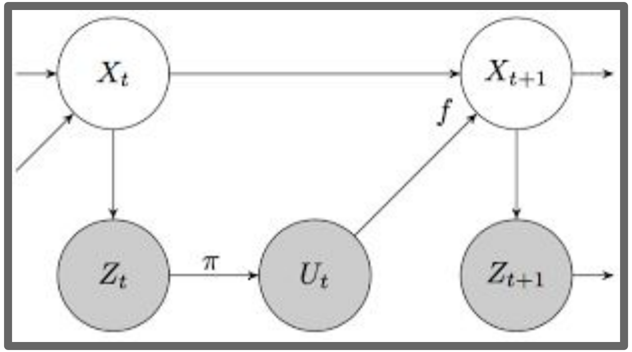
Gaussian Processes & Bayesian Optimization

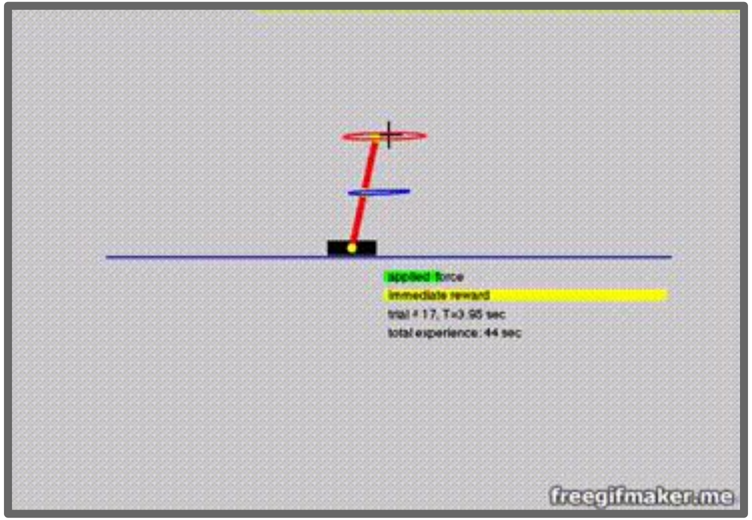
$$p(\mathrm{argmax}_x f(x)|\mathcal{D})$$



Probabilistic Reinforcement Learning

Model-based Reinforcement Learning





$$\mathrm{argmax}_\theta \mathbb{E}\left(\left(\sum_{t=0}^{T} \gamma^t r_t\right) \middle| \pi(a|s,\theta)\right)$$

using $$p^\pi\left(\sum_t^T \gamma^t r_t | s_t, a_t\right)$$

The Future is Deep + Probabilistic ML

Use deep learning as a powerful tool for inference in complicated probabilistic models of the real world

Come help build the Future at Amplyfi!

AMPLYFi

are on the hunt for exceptional talent…

Current vacancies:

- DevOps Engineer
- Software Architects
- Senior Software Engineers
- Strategy Analysts

Interested to hear more?

Contact Chloe Murray, Talent Manager
chloe.murray@amplyfi.com
www.amplyfi.com