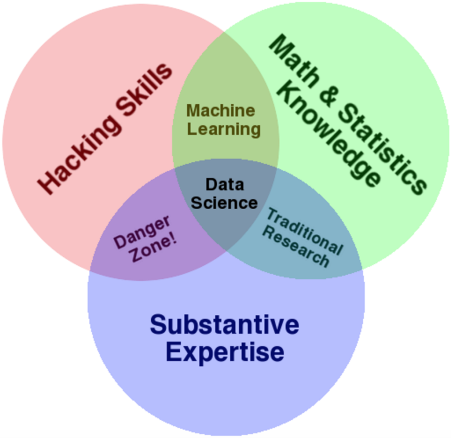
Machine Learning Trends – Generative Adversarial Networks

Fernando Perez-Cruz

Swiss Data Science Center

Bern Winter School on Machine Learning

Conway's Data Science Venn Diagram



Outline

Unsupervised Learning

From David MacKay's 2004 book:

Postscript on Supervised Neural Networks

One of my students, Robert, asked:

> Maybe I'm missing something fundamental, but supervised neural networks seem equivalent to fitting a pre-defined function to some given data, then extrapolating – what's the difference?

I agree with Robert. The supervised neural networks we have studied so far are simply parameterized nonlinear functions which can be fitted to data. Hopefully you will agree with another comment that Robert made:

> Unsupervised networks seem much more interesting than their supervised counterparts. I'm amazed that it works!

Unsupervised Learning

AI: Nothing but Tinned Human Thought?

In spite of all the feverish talk about it these days, true artificial intelligence is still remaining a fancy. We know very well what to expect from an autonomously behaving animal or of a human being, yet it is only all too plain that when dealing with digital devices we get nothing but long past human thought trying to foresee the current situation.

We will never have fully autonomous vehicles and robots without breaking through this glass ceiling, and even now we live with severe social and economical restrictions.

Confess, even if you talk to a digital agent in plain language you feel very well that all the reactions you'll ever get have long been prepared, that you are talking to a menu, and as soon as you deviate from standard situations you'd rather contact a real person.

Before looking beyond the glass ceiling, let me argue my case. Two mutually jealous fields have been attempting over the past 5 or 6 decades to emulate human intelligence: one based on the algorithmic approach (AI in the narrow sense), one based on artificial neural networks (ANNs).

Unsupervised Learning

Yann LeCun
November 9, 2016 ·

Christof von der Malsburg, the renowned theoretical neuroscientist, is arguing that today's popular AI methods (read: supervised learning) rely too much on human intervention to be a path to "real" AI.

I totally agree. I have made a similar point in all my talks of the last two years.

The path to AI goes through unsupervised learning.

I first met Christof when I was a young grad student around 1984. I explained to him the idea of backprop, and he said "this seems like an interesting idea. I shall need to familiarize myself with it".

Inthe mid 1980s he was an early proponent of the idea of "fast weights", ie synaptic weights that change on a fast time scale (commensurate with the time it takes to perceive). These are ideas that are no becoming popular in the deep learning community with attention mechanisms, gating, and multiplicative interactions.

AI: Tinned Human Thought? | Platonite

AI: Tinned Human Thought? by Christoph Malsburg | Nov 5, 2016 | Thoughts | AI: Nothing but Tinned Human Thought? In spite of all the feverish talk about it these days, true artificial intelligence is still remaining a fancy. We know very well what to expect from an autonomously behaving animal or of...

PLATONITE.COM

275 10 Comments 44 Shares

Machine Learning

Supervised Learning

- Well-know problem with available data and prior knowledge.
- Model: $p(y|\mathbf{x})$ or $\widehat{y} = f(\mathbf{x})$.
- Tools to solve them: ensamble methods or neural nets.
- Main Issue: Computation.

Unsupervised Learning

- Data unstructured or unknown.
- Model: $p(y, \mathbf{x})$ or $p(\mathbf{x})$.
- Tools to solve them: density estimation or latent variable models.
- Main issue: Scalability and automatization.

Machine Learning

Supervised Learning

- Well-know problem with available data and prior knowledge.
- Model: $p(y|\mathbf{x})$ or $\widehat{y} = f(\mathbf{x})$.
- Tools to solve them: ensamble methods or neural nets.
- Main Issue: Computation.

Unsupervised Learning

- Data unstructured or unknown.
- Model: $p(y, \mathbf{x})$ or $p(\mathbf{x})$.
- Tools to solve them: density estimation or latent variable models.
- Main issue: Scalability and automatization.

What about reinforcement learning?

Density Estimation

- Standard problem in statistics.
- learn the distribution from the data.
- Classically: parametric family of densities:

$$p_\theta, \qquad \theta \in \Theta$$

- Maximum likelihood estimation:

$$\theta^* = \arg\max_\theta E_{p(\mathbf{x})}[\log(\mathbf{x})]$$

Prescribed models

- Prescribed models definition:

  ensure that p_θ defines a proper density.
- Ability to evaluate density p_θ at sample points $\mathbf{x}$:
  - trivial for exponential families or mixtures.
  - impractical for complex model.
- What are the other strategies for more complex models?

Latent variable models

DeFinetti's Theorem:

$$p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \int \prod p(\mathbf{x}_i|\theta)p(\theta)d\theta$$

for exchangeable observations.

Dimensionality reduction:

- Principal Component Analysis/Factor Analysis.
- Nonnegative Matrix Factorization.
- LLE/Isomap/GPLVM.
- Restricted Boltzmann Machine.
- Dirichlet Processes (aka Chinese Restaurant Process).
- Beta Processes (aka Indian Buffet Process).

Interpretability

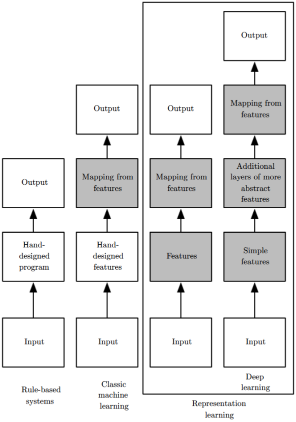
F. Dohsi-Velez et al. (NIPS 2015)

Objectives such as data exploration present unique challenges and opportunities for problems in unsupervised learning. While in more typical scenarios, the discovered latent structures are simply required for some downstream task – such as features for a supervised prediction problem – in data exploration, the model must provide information to a domain expert in a form that they can readily interpret. It is not sufficient to simply list what observations are part of which cluster; one must also be able to explain why the data partition in that particular way. These explanations must necessarily be succinct, as people are limited in the number of cognitive entities that they can process at one time.

Outline

Generative Adversarial Networks

- Rely on representation learning, as in supervised learning.
- Substitute shallow likelihood for a neural network.
- Goal is not to model p_θ, but generate from it
- In a way similar to inverting the cumulative function: $x = F^{-1}(u), \quad u \sim \mathcal{U}[0,1]$



From Optimal Discrimination to Generation

- Proposed by Goodfellow and co-workers in 2014.
- Transform density estimation into a classification problem:
  - Given data $\mathcal{D} = \mathbf{x}_1, \ldots, \mathbf{x}_n$ from p.
  - Given n data points generated from a model p_θ.
  - Find the optimal classifier:

$$q_\theta = p/(p + p_\theta)$$

- Train the model (i.e. the generator) by minimizing the logistic likelihood:

$$\theta = \arg\min \ell^*(\theta) = \mathbf{E}_{\widetilde{p}_\theta}[y \ln q_\theta(\mathbf{x}) + (1-y)\ln(1 - q_\theta(\mathbf{x}))]$$

where $\widetilde{p}_\theta(\mathbf{x}, y = 1) = p(\mathbf{x})/2$ and $\widetilde{p}_\theta(\mathbf{x}, y = 0) = p_\theta(\mathbf{x})/2$

From Real Discrimination to Generation

- Optimal classifier is generally inaccessible.
- Instead we define a classification model:

$$q_\phi : \mathbf{x} \to [0;1], \qquad \phi \in \Phi$$

- Define objective via a bound:

$$\ell^*(\theta) \geq \sup_\phi \ell(\theta, \phi)$$

$$\ell(\theta, \phi) := \mathbf{E}_{\widetilde{p}_\theta} [y \ln q_\phi(\mathbf{x}) + (1 - y) \ln(1 - q_\phi(\mathbf{x}))]$$

  - find the best classifier within restricted family.
  - typically q_ϕ is a deep neural network.
  - training objective for generator is implicit.

Optimizing GANs

- Saddle-point problem:

$$\theta^* = \arg\min_{\theta}\{\sup_{\phi} \ell(\theta, \phi)\}$$

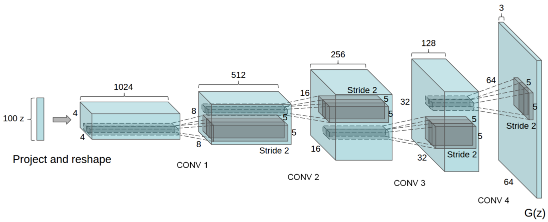
  - explicitly performing the inner sup is not practical.
  - varios methods for optimization / solving games.
- Stochastic gradient descent:

$$\theta^{t+1} = \theta^t - \eta \nabla_\theta \ell(\theta^t, \phi^t)$$
$$\phi^{t+1} = \phi^t + \eta \nabla_\phi \ell(\theta^{t+1}, \phi^t)$$

  It may diverge.
- On going research: Tens of papers in the last couple of years. Jury still out.

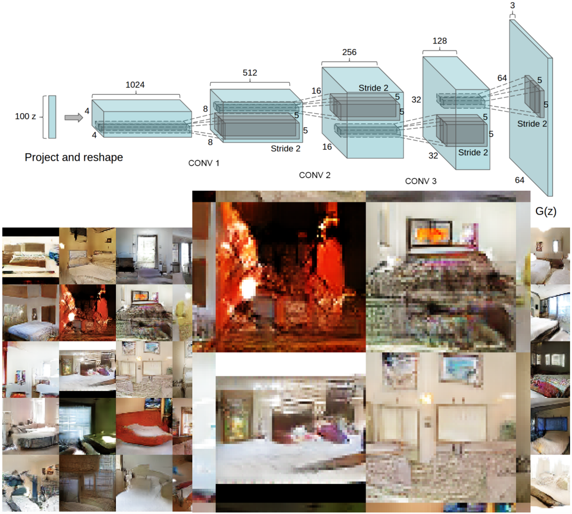
Example: Image Generation

From Radford, Metz and Chintala 2015.



Example: Image Generation

From Radford, Metz and Chintala 2015.



Evaluating GANs

- Convergence to $p(\mathbf{x})$:
  - it is analyzed in Goodfellow et al. 2014.
  - moment matching (Liu, Bousquet and Chaudhuri 2017).
  - AdaGAN (Tolstikhin et al. 2017)
- How to measure quality of implicit models? (fundamental question)
  - out-of-sample evaluations is not available for implicit models.
  - visual inspection (inception score).
- Trade-offs:
  1. noisy samples (e.g. blurry images), but adequate representation of the variability.
  2. faithful (as in good looking) samples, but lack of representation ("mode dropping").

  which one is better?
- Evaluation that is based on each application.