What We Did Not Cover

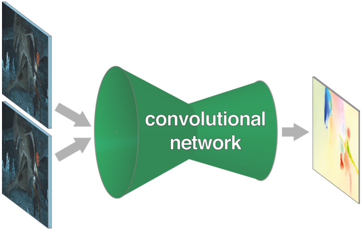
COMPSCI 371D — Machine Learning

What We Did Not Cover

1. Much More Detail
2. Statistical Machine Learning
3. Other Supervised Techniques
4. Reducing the Burden of Labeling
5. Unsupervised Methods
6. Addressing Multiple Learning Tasks Together
7. Prediction over Time

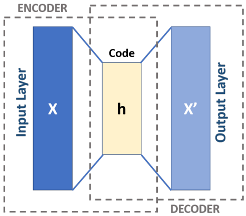
Much More Detail

- Computationally efficient training algorithms: Optimization techniques
- Deep learning architectures for special problems: Image motion analysis, video analysis, ...



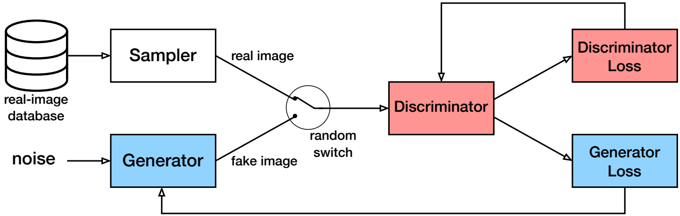
Beyond Discriminative Neural Networks

- Abstraction for its own sake: Auto-encoders
- A game-theoretical technique to draw from a distribution: Generative Adversarial Networks
- See also recent diffusion methods



Which image is fake?

Generative Adversarial Networks



- Discriminator guesses if input is real or fake
- Discriminator loss penalizes wrong predictions
- Generator loss penalizes correct predictions
- After training keep only the generator

Statistical Machine Learning

- How to measure the size of $\mathcal{H}$: Vapnik-Chervonenkis dimension, Rademacher complexity
- How large must T be to get an h that is within ϵ of a performance target with probability greater than $1 - \delta$: Probably Approximately Correct (PAC) learning
- $\mathcal{H}$ is *learnable* if there exists a size of T that is large enough for this goal to be achieved
- Which $\mathcal{H}$s are learnable?
- How large must S be to get a performance measure accurate within ϵ: Concentration bounds, statistical estimation theory, PAC-like techniques

Other Supervised Techniques

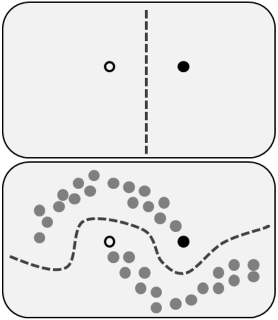
- Boosting: How to use many bad predictors to make one good one

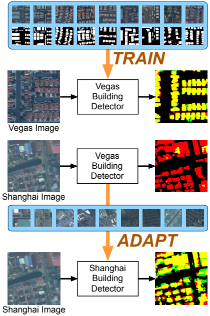
  Similar in principle to ensemble predictors, different assumptions and techniques
- Learning to rank

  Example: Learning a better Google

Reducing the Burden of Labeling

- Semi-supervised methods: Build models of the data $\mathbf{x}$ to leverage sparse labels y
- Domain adaptation: Train a classifier on source-domain labeled data $(\mathbf{x}, y)$ and target-domain unlableled data $\mathbf{x}$ so that it works well in the target domain

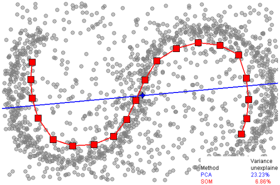




https://en.wikipedia.org

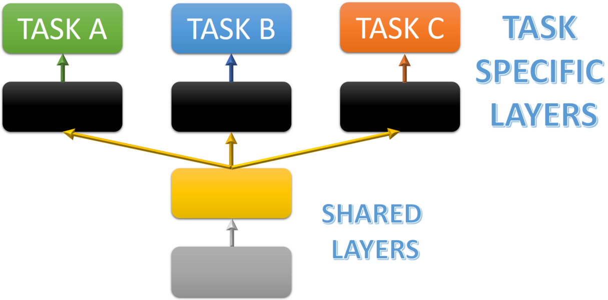
Unsupervised Methods

- Dimensionality reduction:
  Compressing $X \subseteq \mathbb{R}^d$ to $X' \subseteq \mathbb{R}^{d'}$ with $d' \ll d$
  - Principal or Independent Component Analysis (PCA, ICA)
  - Manifold learning, GANs
- Clustering:
  - K-means
  - Expectation-Maximization
  - Agglomerative methods
  - Splitting methods



Addressing Multiple Learning Tasks Together

- Multi-task learning: How to learn representations that are common to different but related prediction tasks



Prediction over Time

- State-space methods
  - Time series analysis
  - Stochastic state estimation
  - System identification
- Recurrent neural networks
- Reinforcement learning: Actions over time
  Learning policies underlying observed sequences