CSCE 625: Introduction to Machine Learning

- Yoonsuck Choe

- October 28, 2013

What Is Machine Learning?

- A subfield of AI that is rapidly growing in importance.

- Performance of a system increased based on learning experience.

- Learning from data.

Why Machine Learning?

- Abundance of data: the data deluge.
    - Scientific instruments.
    - Data acquisition devices.
    - Internet and the web.
    - All sectors of human society producing and digitizing data.

- Not enough human expertise or man power to make sense of such huge amounts of data.

Machine Learning in the News

IBM's Watson

- IBM's Watson beats human champions: Jeopardy (game show)

- Google detects cats from YouTube videos.

- Google Glass app recognizes people it sees.

- Legal, medical, financial applications.
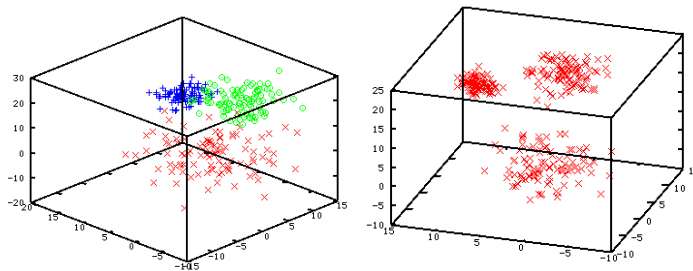
What Does It Take to do ML?

A lot of math:

- Linear algebra

- Calculus

- Probability and statistics

- Differential geometry

- Numerical methods

Types of Machine Learning

- Supervised learning
  - Input-Target pairs
  - $\{\langle \vec{x}_i, \vec{t}_i \rangle | i = 1, 2, .., n\}$

- Unsupervised learning
  - A bunch of inputs (unlabeled)
  - $\{\vec{x}_i | i = 1, 2, ..., n\}$

- Reinforcement learning
  - $\text{state}_1 \xrightarrow{\text{action}_1} \text{state}_2 \xrightarrow{\text{action}_2} \text{state}_3, ...$, reward
  - $s_{t+1} = \delta(s_t, a_t),\ r_{t+1} = \rho(s_t, a_t)$

Example Data

- Left: supervised

- Right: unsupervised

- Typically very high dimensional (10,000, 1 million [or more]).

High-dimensional Data

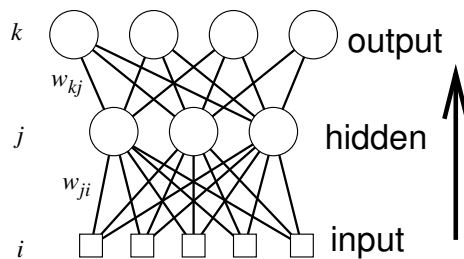$$\longrightarrow \underbrace{[240, 240, ..., 232, ...]}_{\text{2,500-D vector}}$$
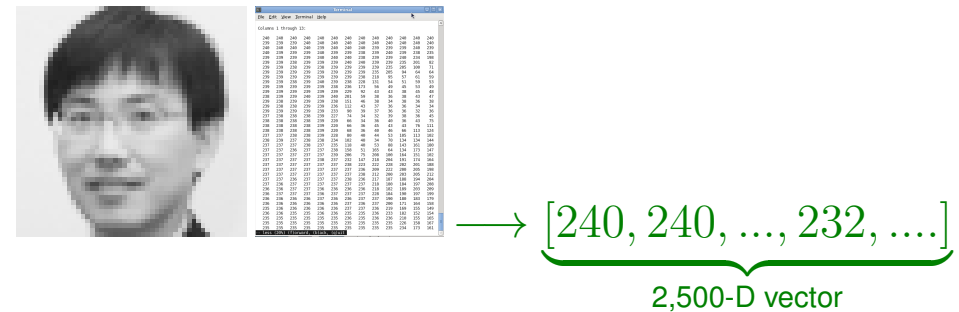
- Images: these are 2D images, but ...

- These are 50×50 = 2,500-dimensional vectors.
 - Each such image is a single point in 2,500-dimensional space.

Supervised Learning

Supervised Learning

- Regression: approximating $y = f(x)$

- Classification: face recognition, hand-written character recognition, credit risk assessment, etc.

- Techniques:

  - Neural networks

  - Decision tree learning

  - Support vector machines

  - Radial basis functions

  - Naive Bayes learning

  - k-nearest neighbor

Neural Networks


- Input, hidden, and output units.

- Connection weights are adjusted based on $\langle \vec{x}_t, \vec{t}_t \rangle$ and error in the output.

Decision Tree Learning

Ex Num	Outlook	Temp.	Humidity	Wind	Water	Forecast	Enjoy Sport?
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
...	...	...	...	...	...	...	...
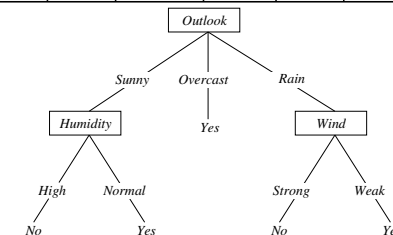
- Building a tree from scratch, one attribute at a time.

- Maximized information gain (checking which attribute reduces uncertainty the most?).

Support Vector Machine

Support Vectors
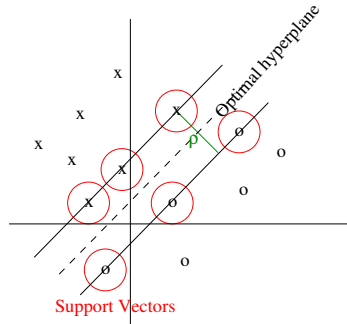
- Similar to a one-layer neural network.

- Learning rule is different.

- Nice optimality properties.

13

Supervised Learning Issues

- How well will it do on training inputs?

- How well will it do on novel inputs?

  – Generalization.

- How many samples needed for sufficient

  performance and generalization?

  – Sample complexity

  – Curse of dimensionality

  – Computational learning theory

- Catastrophic forgetting (online learning hard).

14

Addendum: Curse of Dimensionality
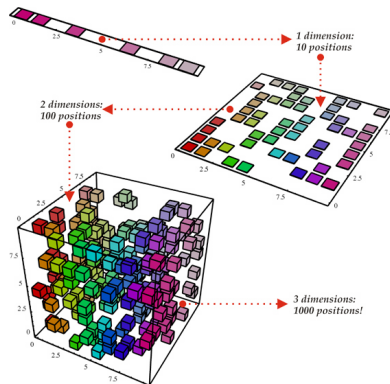
From: Yoshua Bengio's page

- Exponentially many points needed to achieve same
  density of training samples.
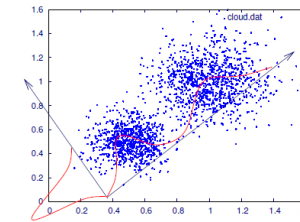
15

Unsupervised Learning

16

Unsupervised Learning

- Clustering, feature extraction, blind source separation, dimensionality reduction, etc.

- Techniques:

  - Principal Component Analysis (PCA)

  - Self-Organizing Maps (SOM)

  - Independent Component Analysis (ICA)

  - Multi-Dimensional Scaling (MDS)

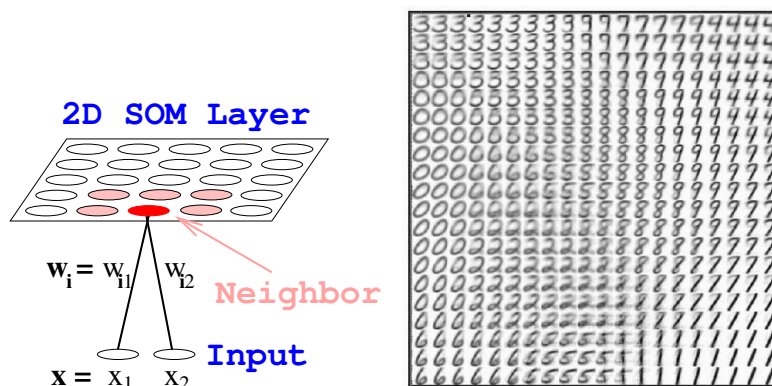 - ISOMAP, Locally Linear Embedding (LLE)

Principal Component Analysis

- Finding orthogonal axes that result in maximum variance when projected.

- Large portion of information resides in the first few principal components.
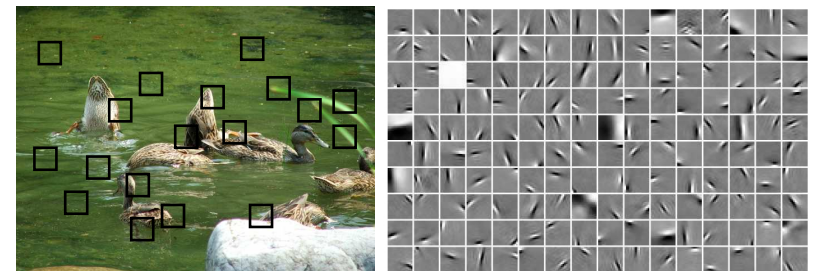
- Dimensionality reduction.

Self-Organizing Maps

2D SOM Layer

$\mathbf{w_i} = w_{i1} \quad w_{i2}$

Neighbor

$\mathbf{x} = x_1 \quad x_2$ **Input**

- Units occupy a regular grid (1D, 2D, 3D), with reference vector.

- Inputs matched to units with most similar reference vectors.

- Reference vectors adjusted based on match and neighbor on grid.

- Nearby units represent similar inputs.

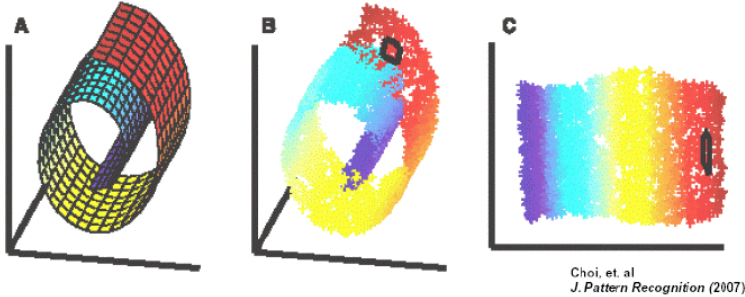# Independent Component Analysis

Hoyer and Hyvärinen (2000)

- Find additive sources (right) based on their mixtures (e.g., image patches to the left).

- Sources assumed to be statistically independent from each other and non-Gaussian.

- Feature extraction, blind source separation.

Manifold Learning: ISOMAP, etc.

Choi, et. al
J. Pattern Recognition (2007)

- Low-dimensional manifold embedded in high-dimensional space.

- Recover the manifold. Geodesic distance a central concept.

- Dimensionality reduction, visualization, etc.

Unsupervised Learning Issues

- Discovering structure.

- Discovering features.

- Removing redundancy.

- How many clusters?

- What distance measures to use?

Reinforcement Learning

Reinforcement Learning

- Very different from supervised and unsupervised learning.

- Multi agent control, robot control, game playing, scheduling, etc.

- Techniques:

  - Value function-based: Q-learning, Temporal difference (TD) learning

  - Direct policy search: Neuroevolution, genetic algorithms.

Learning the Meaning of Neural Spikes

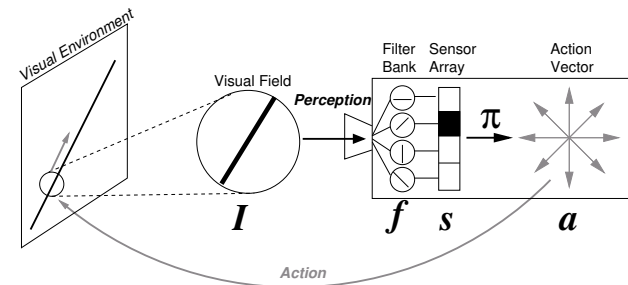- What do these blinking lights mean? (Choe et al. 2007).

What If They Are Brain Responses to **Something**

They Are Visual Cortical Responses to **Oriented Lines**

This is a problem of *grounding*.

Use Reinforcement Learning

- Direct access to **encoded internal state** (sensory array) only.

- Action is enabled, which can **move the gaze**.

- How does this solve the grounding problem?

Action for Unchanging Internal State

- Diagonal motion causes the *internal state* to **remain unchanging** over time.

- Property of such a movement **exactly reflects** the property of the input I: Semantics figured out through action.
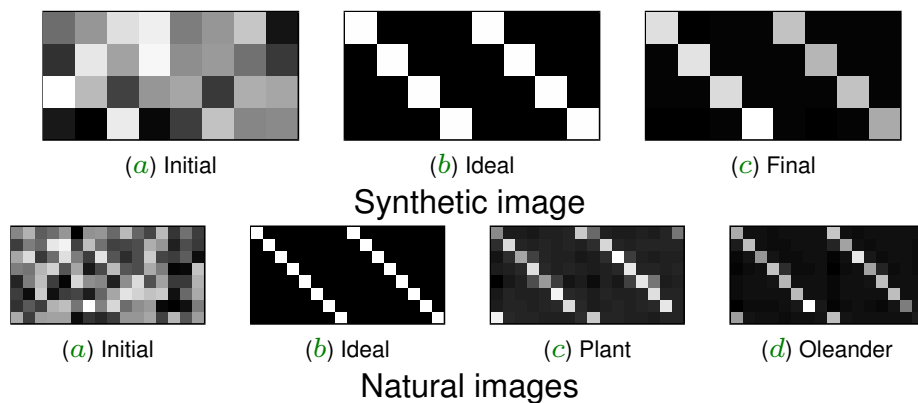
Reinforcement Learing

A: direction of motion

S: sensory state (orientation)	→	↗	↑	↘	←	↙	↓	↘
⊖	0.5	0	0	0	0.5	0	0	0
⊘	0	0.5	0	0	0	0.5	0	0
⊖	0	0	R(s ,a)	0	0	0	0.5	0
⊘	0	0	0	0.5	0	0	0	0.5

- Learn state-to-action mapping to maximize invariance in internal state.

Results: Learned $R(s, a)$

(a) Initial (b) Ideal (c) Final

Synthetic image

(a) Initial (b) Ideal (c) Plant (d) Oleander

Natural images

- Learned $R(s, a)$ close to ideal.

Results: Gaze Trajectory

(a) Input (b) Initial (c) Final
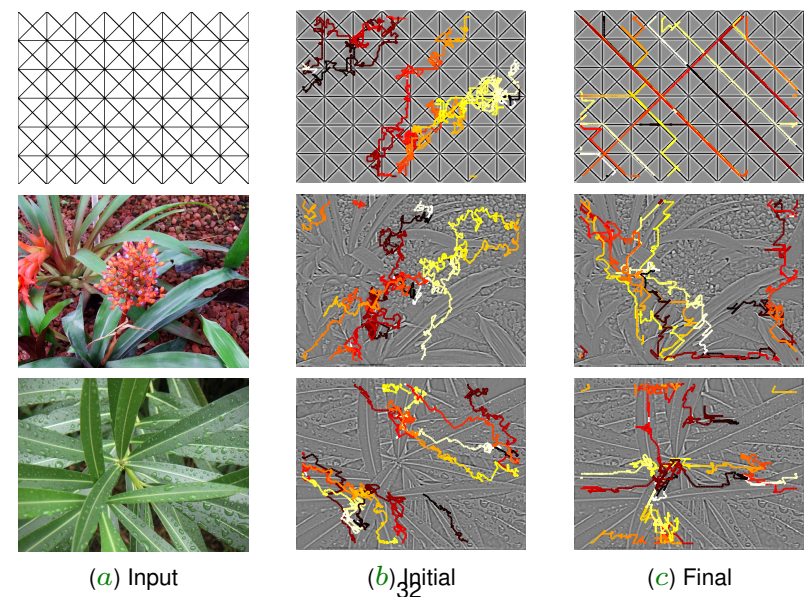
Brief Summary

- Decoding of encoded representation can be done **without** external reference.

- Action and changes in the internal representation induced by action is the key.

- Reinforcement learning plays a key role.

Reinforcement Learning Issues

- Discrete states and actions is a norm.

- Scalability an issue.

- Certain assumptions: state-action pair visited infinitely often.

- Online learning, safety, transfer, etc.

Wrap Up

Summary

- Machine learning is a rapidly developing field with great promise:
  - Big data
  - New theoretical insights (e.g., deep learning)
- Three types of ML:
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning
- Need to look beyond ML:
  - ML good at solving problems, but not posing problems (Choe and Mann 2012).