Artificial Intelligence at Chicago

Partha Niyogi

The University of Chicago
Statistics — data analysis and probabilistic models
Computer Science — design and analysis of algorithms
Mathematics — analytic tools

Artificial Intelligence:
Pattern recognition in speech, vision, text, bioinformatics, finance etc.

Cognitive and Neural Sciences:
Human learning of language, concepts, movement, vision, etc.
Canonical Problems

- Classification and Regression
- Clustering
- Dimensionality Reduction and Data Representation
- Density Estimation
A Pattern Recognition Example

\( P \) on \( X \times Y \)

\[ X = \mathbb{R}^N; \ Y = \{0, 1\}, \mathbb{R} \]

\((x_i, y_i)\) labeled examples

find \( f : X \rightarrow Y \) \quad \text{Ill Posed}
Regularization Principle

\[ f = \arg \min_{f \in H_K} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \gamma \|f\|_K^2 \]

- Splines
- Ridge Regression
- SVM

- \( K : X \times X \to \mathbb{R} \) is a p.d. kernel
  - e.g. \( e^{-\frac{\|x-y\|^2}{\sigma^2}} \), \((1 + x \cdot y)^d\), etc.

- \( H_K \) is a corresponding RKHS
  - e.g., certain Sobolev spaces, polynomial families, etc.
Simplicity is Relative
Intuitions

- $\text{supp } P_X$ has manifold structure
- *geodesic* distance v/s *ambient* distance
- geometric structure of data should be incorporated
- $f$ versus $f_M$
Manifold Regularization

$$\min_{f \in H_K} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \gamma_A \|f\|^2_K + \gamma_I \|f\|^2_I$$

$$\|f\|^2_I = \begin{cases} 
\text{Laplacian} & \int \langle \text{grad}_M f, \text{grad}_M f \rangle = \int f \Delta_M f \\
\text{Iterated Laplacian} & \int f \Delta_M^i f \\
\text{Heat kernel} & \text{e}^{-\Delta_M t} \\
\text{Differential Operator} & \int f(Df) 
\end{cases}$$

Representer Theorem: $$f = \sum_{i=1}^{n} \alpha_i K(x, x_i) + \int_M \alpha(y) K(x, y)$$

Belkin, Niyogi, Sindhwani (2004)
Approximating $\|f\|_I^2$

$\mathcal{M}$ is unknown but $x_1 \ldots x_M \in \mathcal{M}$

\[
\|f\|_I^2 = \int_{\mathcal{M}} \langle \nabla f, \nabla f \rangle \approx \sum_{i \sim j} W_{ij} (f(x_i) - f(x_j))^2
\]
\[ M \approx G = (V, E) \]

\[ e_{ij} \in E \text{ if } \|x_i - x_j\| < \epsilon \]

\[ W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}} \]

\[ \Delta_M \approx L = D - W \]

\[ \int \langle \text{grad} f, \text{grad} f \rangle \approx \sum_{i,j} W_{ij} (f(x_i) - f(x_j))^2 \]

\[ \int f(\Delta f) \approx f^T L f \]
Manifold Regularization

\[
\frac{1}{n} \sum_{i=1}^{n} V(f(x_i), y_i) + \gamma_A \|f\|_K^2 + \gamma_I \sum_{i \sim j} W_{ij} (f(x_i) - f(x_j))^2
\]

Representer Theorem: \( f_{opt} = \sum_{i=1}^{n+m} \alpha_i K(x, x_i) \)

- \( V(f(x), y) = (f(x) - y)^2 \): Least squares
- \( V(f(x), y) = (1 - yf(x))_+ \): Hinge loss (Support Vector Machines)
Ambient and Intrinsic Regularization

\( \gamma_A = 0.03125 \) \( \gamma_I = 0 \)

\( \gamma_A = 0.03125 \) \( \gamma_I = 0.01 \)

\( \gamma_A = 0.03125 \) \( \gamma_I = 1 \)
Experimental Results: USPS
Convergence Theorem

with prob. $> 1 - \delta$

$$|E_{\gamma_A, \gamma_I, n} - E_{opt}| \leq C + \gamma_A \|f_{opt}\|_K^2 + \gamma_I \int_{\mathcal{M}} f_{opt}(\Delta^I f_{opt})$$

$$C = \frac{4}{\beta^{3/2}} \sqrt{\frac{1}{n} \log \left( \frac{2}{\delta} \right)}$$

$$\beta^2 = \frac{\gamma_A}{\kappa^2} + \frac{\gamma_I}{\mu^2}$$

$$\kappa^2 = \sup_{x \in X} K(x, x) \quad \mu^2 = \sup_{x \in \mathcal{M}} \sum_i \left( \frac{1}{\lambda_i} \right)^l \phi_i^2(x)$$
An Acoustic Example

\[ u(t) \quad \rightarrow \quad s(t) \]

\[ l \]
An Acoustic Example

\[ u(t) \leftrightarrow l \leftrightarrow s(t) \]

One Dimensional Air Flow

(i) \[ \frac{\partial V}{\partial x} = -\frac{A}{\rho c^2} \frac{\partial P}{\partial t} \]

(ii) \[ \frac{\partial P}{\partial x} = -\frac{\rho}{A} \frac{\partial V}{\partial t} \]

\[ V(x,t) = \text{volume velocity} \]

\[ P(x,t) = \text{pressure} \]
\[ u(t) = \sum_{n=1}^{\infty} \alpha_n \sin(n\omega_0 t) \in l_2 \]

\[ s(t) = \sum_{n=1}^{\infty} \beta_n \sin(n\omega_0 t) \in l_2 \]
Vocal Tract modeled as a sequence of tubes. (e.g. Stevens, 1998)

Jansen and Niyogi (in prep.)
Machine vision: inferring joint angles.
Corazza, Andriacchi, Stanford Biomotion Lab, 05, Partiview, Surendran
1 (a) He ran from there with his money.

1 (b) He his money with there from ran. (⋆)
1 (a) He ran from there with his money.

1 (b) He his money with there from ran. (*)

Linguistic Experience $\leftrightarrow$ Linguistic Knowledge
$\mathcal{G}$ \hspace{5pt} $g_t \in \mathcal{G}$ target grammar

$S_n$ \hspace{5pt} $+$ve examples

$\mathcal{A}$ \hspace{5pt} Learner (Child)

$\mathcal{A}(S_n) = h_n$
\[ G \quad g_t \in G \text{ target grammar} \]

\[ S_n \quad +ve \text{ examples} \]

\[ A \quad \text{Learner (Child)} \]

\[ A(S_n) = h_n \]

Learnability \[ h_n \rightarrow g_t \]

Gold (1967); Valiant (1984)
The Evolution of English

Her ... Aelfred cyning ... gefeaht wid ealne, here, and hine

*Here Alfred king fought against whole army and it*

geflymde and him aefter rad od pet geweorc, and paer saet

*put to flight and it after rode to the fortress and there camped*

XIII niht, and pa sealde se here him gislas and myccl
das, pet he of his rice woldon, and him eac geheton

*fourteen nights and then gave the army him hostages and great oaths that they from his kingdom would [go] and him also promised*

pet heora cyng fulwithe onfon wolde, and hi paet gelaston

*that their king baptism receive would and they that did*
pa Darius geseah paet he oferwunnen beon wolde
then Darius saw that [he conquered be would]

& him aefterfylgende waes
and [him following was]

Nu ic wille eac paes maran Alexandres gemunende beon
now I will also [the great Alexander considering be]

(Orosius 128.5)

(Orosius 236.29)

(Orosius 110.10)
How does the mind/brain work?

How can we replicate intelligent phenomena in machines?
Faculty

Language
  - John Goldsmith
  - Gina Levow
  - Partha Niyogi
  - Terry Regier (Psychology)
  - Howard Nusbaum (Psychology)
  - Yasemin Altun (Toyota)

Vision
  - Yali Amit
  - Pedro Felzenszwalb
  - Christin Sminchesescu (Toyota)
Partha Niyogi
Yali Amit
Marc Coram (Statistics)
David McAllester (Toyota)
Stephen Smale (Toyota)
Adam Kalai (Toyota)
Friends

- Statistics and Mathematics
- Linguistics and Psychology
- Computational Neuroscience
- Toyota Technological Institute