# Kernel methods on finite groups

Risi Imre Kondor

Center for Automated Learning and Discovery
School of Computer Science
Carnegie Mellon University

# The learning task

Data: $(x_i, y_i)$ pairs $x_i \in \Omega$

Task: predict $y$ from $x$

$K : \Omega \times \Omega \mapsto \mathbb{R}$ expresses our prior beliefs about similarity

eg. $K \sim e^{-(x_1 - x_2)^2/2\sigma^2}$

# The kernel trick

$\phi : \Omega \mapsto \mathcal{H}$

$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$$

*“ K induces an isometric embedding of $\Omega$ in the Hilbert space $\mathcal{H}$”*

# Positive de£niteness

$K$ symmetric and

$$\int_{\Omega}\int_{\Omega} f(x_1)\, f(x_2)\, K(x_1, x_2)\, dx_1\, dx_2 \geq 0 \qquad \forall f \in L_2(\Omega) \quad (Mercer)$$

$$\sum_{x_1}\sum_{x_2}\; c_{x_1}\, c_{x_2}\, K(x_1, x_2) \geq 0 \qquad \forall c_1, c_2, \ldots \in \mathbb{R}$$

Can we do this on discrete $\Omega$?

Can we use the kernel trick to unfold the internal structure of the object $\Omega$?

# Finite Groups

Finite set $G$ with operation $G \times G \mapsto G$

- $x_1 x_2 \in G$ for any $x_1, x_2 \in G$ (closure)
- $x_1(x_2 x_3) = (x_1 x_2) x_3$ (associativity)
- $xe = ex = e$ for any $x \in G$ (identity)
- $x^{-1} x = x x^{-1} = e$ (inverses)

# Symmetric groups $S_n$

$x_1 = (12)(3)(4)(5) \qquad x_1(\text{ABCDE}) = \text{BACDE}$

$x_2 = (1)(324)(5) \qquad x_2(\text{ABCDE}) = \text{ACDBE}$

$x_3 = x_1 x_2 \qquad x_3(\text{ABCDE}) = x_1(x_2(\text{ABCDE}))$

rankings, orderings, allocation, etc.
natural sense of distance

# Stationary kernels

$$K(x_1, x_2) = f(x_2\, x_1^{-1})$$

$$\sim$$

$$K(x_1, x_2) = f(x_2 - x_1)$$

eg. $K \sim e^{-(x_1 - x_2)^2/2\sigma^2}$

$f$ positive definite function on $G$

# Bochner’s Theorem

$f$ is positive definite and symmetrical on $\mathbb{R}^m$ iff $\widehat{f}(\boldsymbol{\omega}) > 0$

$$\widehat{f}(\boldsymbol{\omega}) = \frac{1}{\sqrt{2\pi}^{\,m}} \int e^{-i\,\boldsymbol{\omega}\cdot\boldsymbol{x}} f(\boldsymbol{x})\, d\boldsymbol{x}$$

is there an analogue for finite groups?

# Representation theory

$\rho : G \to \mathbb{C}^{\,d\times d}$

$$\rho(x_1 x_2) = \rho(x_1)\rho(x_2)$$

Equivalence:

$$\rho_1(x) = t^{-1}\rho_2(x)\, t \qquad \forall\, x \in G$$

Reducibility:

$$t^{-1}\rho(x)\, t = \left(\begin{array}{c|c} \rho_1(x) & 0 \\ \hline 0 & \rho_2(x) \end{array}\right) \qquad \forall\, x \in G$$

# Irreducible representations of $S_5$

$$\rho_{\mathrm{trivial}}(x) \equiv (1) \qquad \rightarrow \quad \rho^{(5)}$$

$$\rho_{\mathrm{sign}}(x) \equiv (\mathrm{sgn}(x)) \qquad \rightarrow \quad \rho^{(1,1,1,1,1)}$$

$$\rho_{\mathrm{def.}}(x) \in \mathbb{C}^{5\times 5} \qquad \boldsymbol{e}_{x(i)} = \rho_{\mathrm{def.}}(x)\,\boldsymbol{e}_i$$

$\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4, \boldsymbol{e}_5$ basis

# Irreducible representations of $S_5$

$$t^{-1}\,\rho_{\text{def.}}(x)\,t = \left(\begin{array}{c|c} \rho_{\text{tr.}}(x) & \\ \hline & \\ & \rho^{(4,1)}(x) \\ & \end{array}\right)$$

$$\begin{aligned}\rho_{\text{reg.}} = \rho^{(5)} \oplus 4\rho^{(4,1)} \oplus 5\rho^{(3,2)} \oplus 6\rho^{(3,1,1)} \oplus \\ 5\rho^{(2,2,1)} \oplus 4\rho^{(2,1,1,1)} \oplus \rho^{(1,1,1,1,1)}\end{aligned}$$

# Fourier transforms on Groups

$$\widehat{f}(\rho) = \sum_{x \in G} \rho(x)\, f(x) \qquad\qquad \rho \in \mathcal{R}$$

$$\mathcal{F} :\ \mathbb{C}G \mapsto \bigoplus_{\rho \in \mathcal{R}} \mathbb{C}^{d_\rho \times d_\rho}$$

inversion:

$$f(x) = \frac{1}{|\,G\,|} \sum_{\rho \in \mathcal{R}} d_\rho \, \mathrm{trace}[\widehat{f}(\rho)\rho(x^{-1})]$$

# Main Theorem

The function $f$ is positive definite on $G$ if and only if the matrices $\widehat{f}(\rho)$ are all positive definite.

# Conjugacy classes

conjugacy classes: $x_1 \cong_{\text{Conj.}} x_2$ iff $x_1 = t^{-1}\, x_2\, t$ for some $t$

eg. on $S_n$

$$(\cdot)(\cdot)(\cdot)(\cdot)\ldots$$
$$(\cdot\ \cdot)(\cdot)(\cdot)(\cdot)\ldots$$
$$(\cdot\ \ \cdot\ \ \cdot)(\cdot)(\cdot)\ldots$$
$$(\cdot\ \cdot)(\cdot\ \cdot)(\cdot)(\cdot)\ldots$$
$$\vdots$$

# Corollary to main theorem

The function $f$ is positive de£nite on $G$ and constant on conjugacy classes if and only if

$$f(x) = \sum_{\rho \in \mathcal{R}} c_\rho \, \chi_\rho \qquad\qquad c_\rho > 0 \; .$$

characters:

$$\chi_\rho(x) = \mathrm{trace}[\rho(x)]$$

— break —

## $S_4$

# Diffusion kernel

$$\boldsymbol{K} \sim e^{\beta \boldsymbol{H}} = \lim_{n \to \infty} \left(1 + \frac{\beta \boldsymbol{H}}{n}\right)^n$$

$$H_{ij} = \begin{cases} 1 & (i,j) \in E \\ -d_i & i = j \\ 0 & \text{otherwise} \end{cases}$$

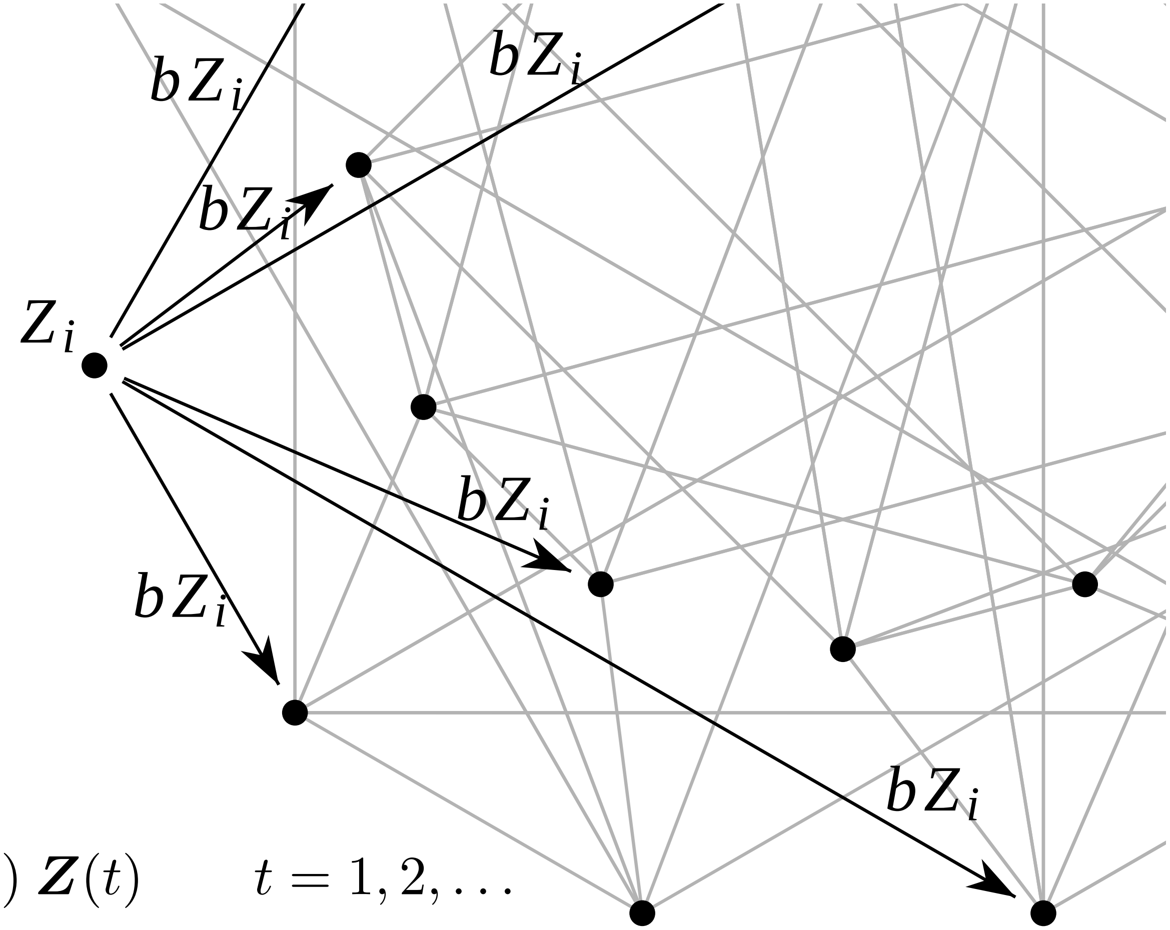
(Laplacian)

# Diffusion on graphs

E$[Z_i(0)] = 0$

Var$[Z_i(0)] = \sigma^2$

$Z_i(0)$ indep.



redistribution:

$$\boldsymbol{Z}(t+1) = (1 + b\boldsymbol{H})\ \boldsymbol{Z}(t) \qquad t = 1, 2, \ldots$$

# Diffusion on graphs (cont.)

evolution operator $\boldsymbol{T}(t) = (1 + b\boldsymbol{H})^t$

$$\boldsymbol{Z}(t) = \boldsymbol{T}(t)\boldsymbol{Z}(0)$$

covariance

$$\mathrm{Cov}_{ij}(t) = \overline{Z_i(t)\, Z_j(t)} = \overline{\Big(\sum_{i'} T_{ii'}\, Z_{i'}(0)\Big)\Big(\sum_{j'} T_{jj'}\, Z_{j'}(0)\Big)}$$

$$\mathbf{Cov}(t) = \sigma^2\, \boldsymbol{T}(t)^T\, \boldsymbol{T}(t) = \sigma^2\, \boldsymbol{T}(t)^2$$

# The continuous limit

$$\boldsymbol{T}(t) = \lim_{n\to\infty}\left(1 + \frac{b\boldsymbol{H}}{n}\right)^{nt} = e^{bt\boldsymbol{H}} \qquad \frac{\partial}{\partial t}\,\boldsymbol{T}(t) = b\boldsymbol{H}\,\boldsymbol{T}(t)$$

$$\mathbf{Cov}(t) = \sigma^2\, e^{\,2bt\boldsymbol{H}}$$

on square grid in $\mathbb{R}^n$, $\lim \boldsymbol{H} = \nabla^2$,

$$\frac{\partial}{\partial t}\mathbf{Cov} = 2b\nabla^2\mathbf{Cov} \qquad \to \qquad \mathrm{Cov}(x_1, x_2) \sim e^{-(x_1-x_2)^2/(2\beta t)}$$

# Diffusion kernels on groups

$$\boldsymbol{K} \sim e^{\,\beta \boldsymbol{H}} = \lim_{n\to\infty}\left(1+\frac{\beta\boldsymbol{H}}{n}\right)^{n} = \lim_{n\to\infty}\left(1+\frac{\beta\boldsymbol{H}}{2n}\right)^{2n}$$

$$H_{x_1 x_2} = g(x_1 x_2^{-1}) \qquad\qquad g(x) = g(x^{-1})$$

- positive definite
- stationary
- continuous bandwidth parameter $\beta$ (infinitely divisible)

# Diffusion kernels in general

$$\boldsymbol{K} \sim e^{\beta H} = \lim_{n \to \infty} \left( 1 + \frac{\beta \boldsymbol{H}}{n} \right)^n$$

*any* graph, *any* positive de£nite matrix

*“The Lie group of positive de£nite kernels over a set $\Omega$ is generated by the set of real symmetric bilinear forms over $\Omega$.”*

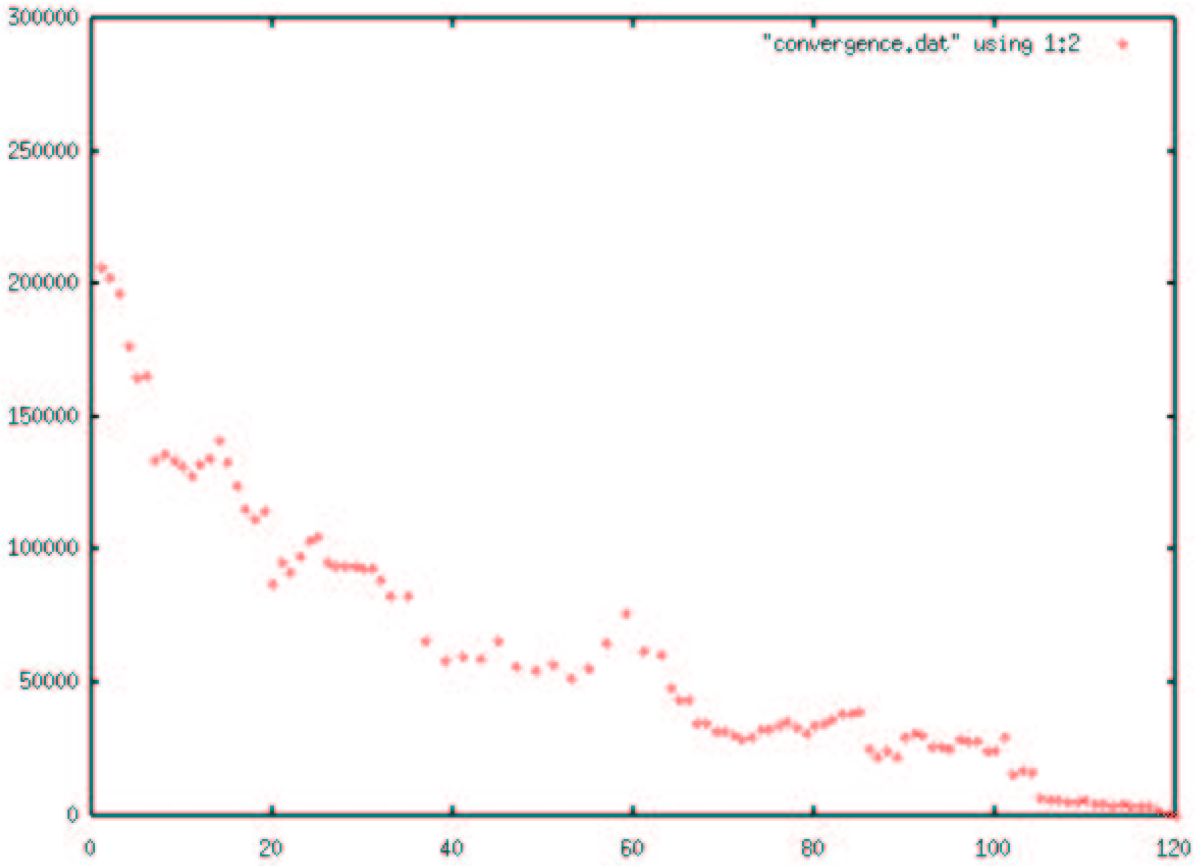
# Experiment

American Psychological Association (1980) (Diaconis 1988)

5 candidates ($5! = 120$)
5738 complete votes

| Ranking | votes |
|---|---|
| 12345 | 30 |
| 12354 | 28 |
| 12435 | 27 |
| 12453 | 29 |
| 12534 | 35 |
| 12543 | 34 |
| 13245 | 102 |
| 13254 | 95 |
| 13425 | 35 |
| 13452 | 37 |
| $\vdots$ | $\vdots$ |

# Experiment


"convergence.dat" using 1:2
300000
250000
200000
150000
100000
50000
0
0
20
40
60
80
100
120

# Summary

Liberate Support Vector Machines, Gaussian Processes, etc. from $\mathbb{R}^m$!

- Real data often lives on discrete objects **with intrinsic structure**
- Many statistical algorithms involve searching, integrating, etc. over a large number of related objects / models

# Thanks

John Lafferty

School of Computer Science,
Carnegie Mellon University

- **J.-P. Serre.** Linear Representations of Finite Groups. Springer-Verlag, 1977.
- **B. E. Sagan.** The Symmetric Group. Springer Verlag, 2001.
- **A. Terras.** Fourier Analysis on Finte Groups and Applications. Cambridge, 1999.
- **P. Diaconis.** Group Representations in Probability and Statistics. Inst. of Math. Stat, 1988.
- **C. Watkins.** Dynamic Alignment Kernels. Tech. Report. Royal Holloway, 1999.
- **D. Haussler.** Convolution Kernels on Discrete Structures. (unpublished) 1999.