CPSC 536F

PSD (positive semidefinite) matrices

and some foundations of ML

(machine learning)

Topics:

(1) Symmetric and positive semidefinite matrices: theory and basic examples

(2) More examples and applications

(3) Kernel functions (usually symmetric and positive definite).

Applications Include:

- SVD (singular value decomposition)

(1) Gives "best" low rank approximation of any matrix

(2) Also writes any matrix in a "convenient form"

- PCA (principal component analysis) is arguably the same thing

---

- Regression and ridge regression

- Pseudoinverse of a matrix (via SVD)

- Graph Laplacians, Expanders, etc.
  1. What do Laplacian eigenvalues mean?

- Graph Edge Laplacians and Reversible Markov chains
- Graphs: div, curl, grad
- Continuous Laplacians and kernels

- Kernel Methods
  1. the "kernel trick"
  2. classical kernels
  3. RKHS (Reproducing Kernels of a Hilbert space)
  4. Applications of kernels

Admin Stuff:

- There's no single textbook: we're trying to give the minimum amount of linear algebra needed to understand the applications of
- I'll provide:
    - My notes for classes
    - References, mostly textbooks or survey articles (free to all UBC community)

- Grades!

≥ 95 I'll take you as a student. I'll write a nice letter of recom.

≥ 90 ... very strong

≥ 85 you're prob not specializing in Comp Sci Theory w/o more background courses

≥ 80 satisfied the expectations of a grad student

Grading based on

- Homework Problems
- Perhaps a project, if you like

Intro to kernel functions:

Origins:

(1) "Harmonics" of ancient Greeks

(2) Heat, Wave, Laplace equation;

Green's function, potential theory

1700's, 1800's

(3) Hilbert "Kernels" ~ 1902

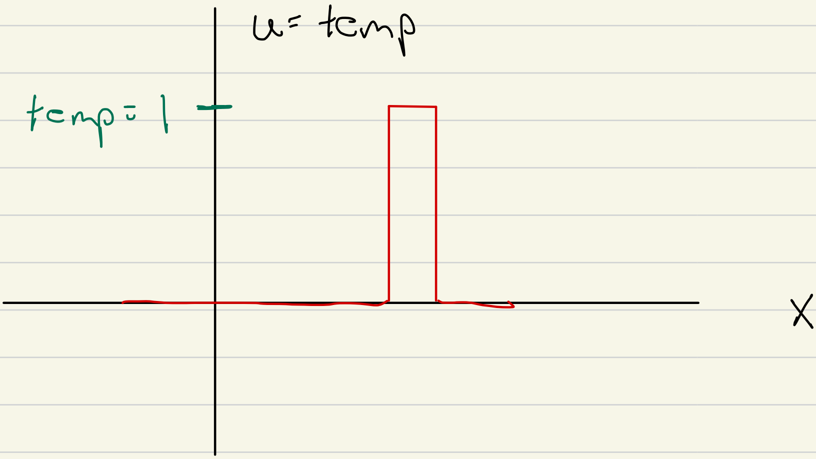
(4) A lot more work in early 1900's

(5) Aronszajn, 1943, 1950

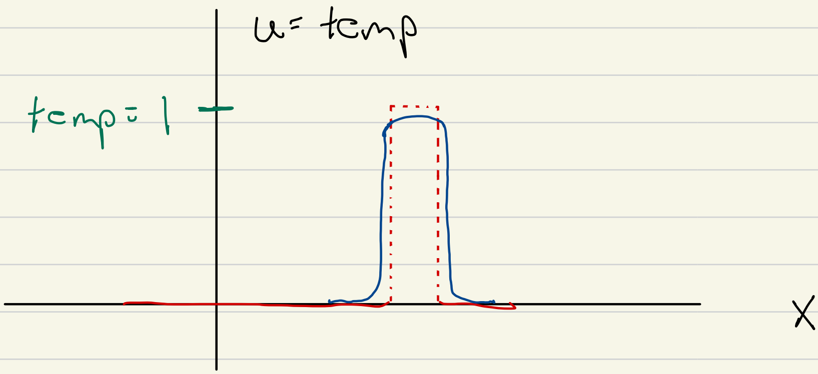
organized the theory

Heat Equation on an infinite rod!

$t = 0$



$t = 0.01$



$$k(x,y;t) = \frac{1}{\sqrt{4\pi t}} e^{-|x-y|^2/4t}$$

Intuition via heat equation properties:

- maximum principle
- analyticity for $t>0$
- $-\Delta u = u_t$ here $\Delta = \left(\frac{\partial}{\partial x}\right)^2$
- $u(x,t) = e^{t\Delta}\left[u(x,0)\right]$

[Δ is negative definite]

- $\int_{-\infty}^{\infty} u(x,t)\,dx =$ total heat at time t

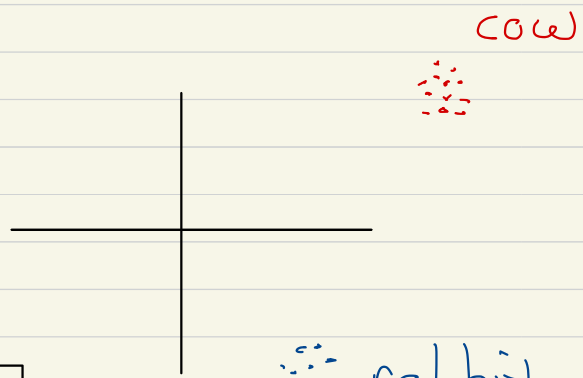
is constant

The "kernel trick"

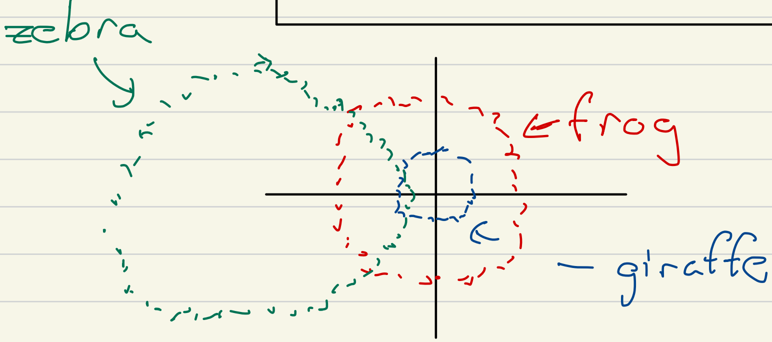
Say you have training data in $\mathbb{R}^N$ (N large)

$$f : \mathbb{R}^N \to \mathbb{R}^2$$

ideal!



bad!



Bad data comes in circles,

annuli, that sometime intersect.

① Idea: $\Phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^6$ via

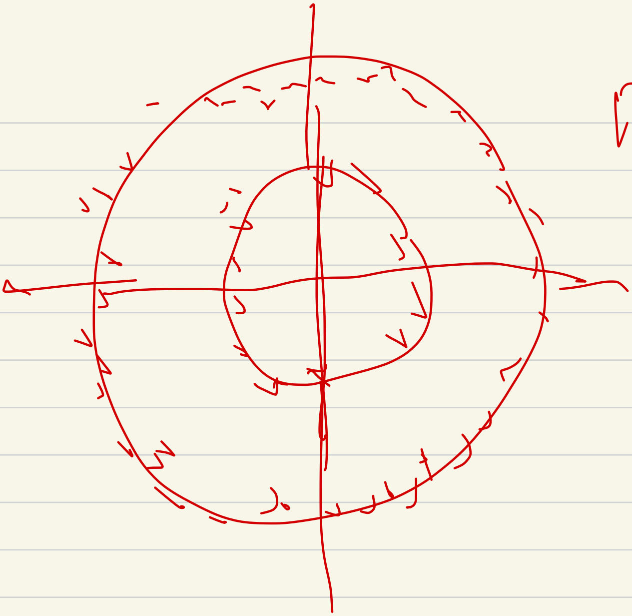
$$\Phi(x,y) = (1, x, y, x^2, xy, y^2)$$

② Trick:

$$\Phi(x_1,y_1) \cdot \Phi(x_2,y_2)$$

$$= (x_1x_2 + y_1y_2 + 1)^2$$

$$= \left((x_1,y_1) \cdot (x_2,y_2) + 1 \right)^2$$



$\mathbb{R}^2$

$$(x,y) \longmapsto (1, x, y, x^2, xy, y^2)$$

Avg of on a circle

of radius r

Details:

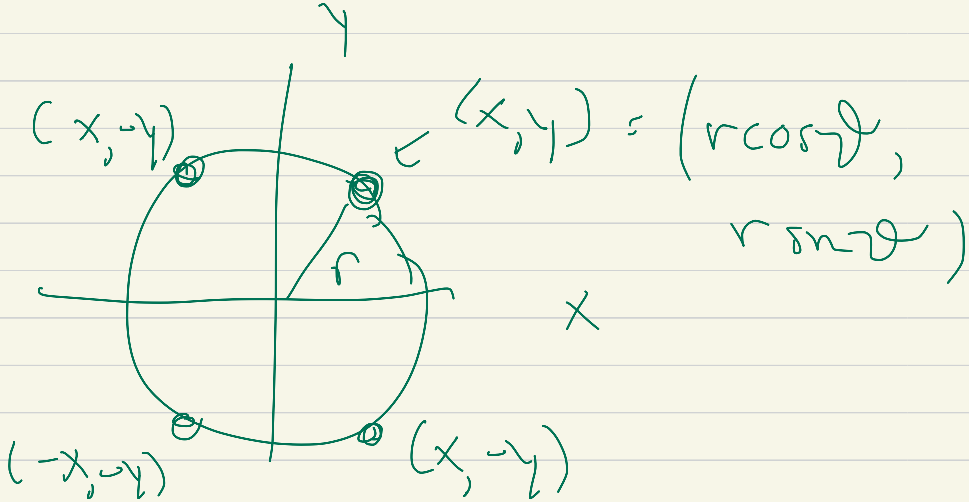
Circle centre (x_0, y_0), radius r!

$$\left.\begin{aligned} x &= r\cos\vartheta + x_0 \\ y &= r\sin\vartheta + y_0 \end{aligned}\right\} \vartheta \in \mathbb{R}$$

Example: Say $x_0 = y_0 = 0$

What is

$$\underset{0 \le \vartheta < 2\pi}{\text{Average}} \quad \Phi\big(x(\vartheta), y(\vartheta)\big)?$$



$$(1, x, y, x^2, xy, y^2)$$

Average over circle $d\theta$

$$\left(1, 0, 0, \frac{r^2}{2}, 0, \frac{r^2}{2}\right)$$

$$x^2 + y^2 = r^2$$