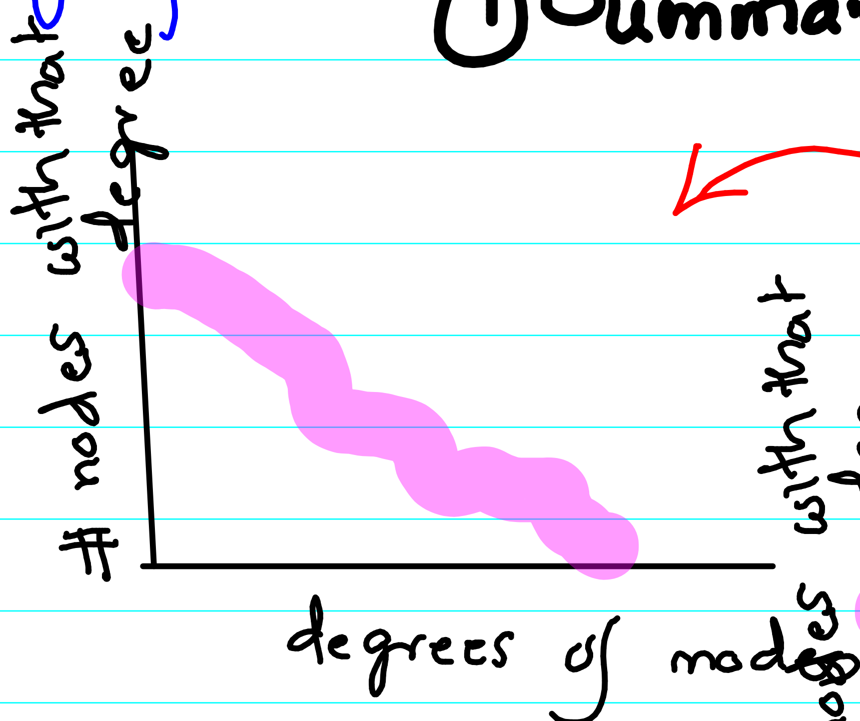
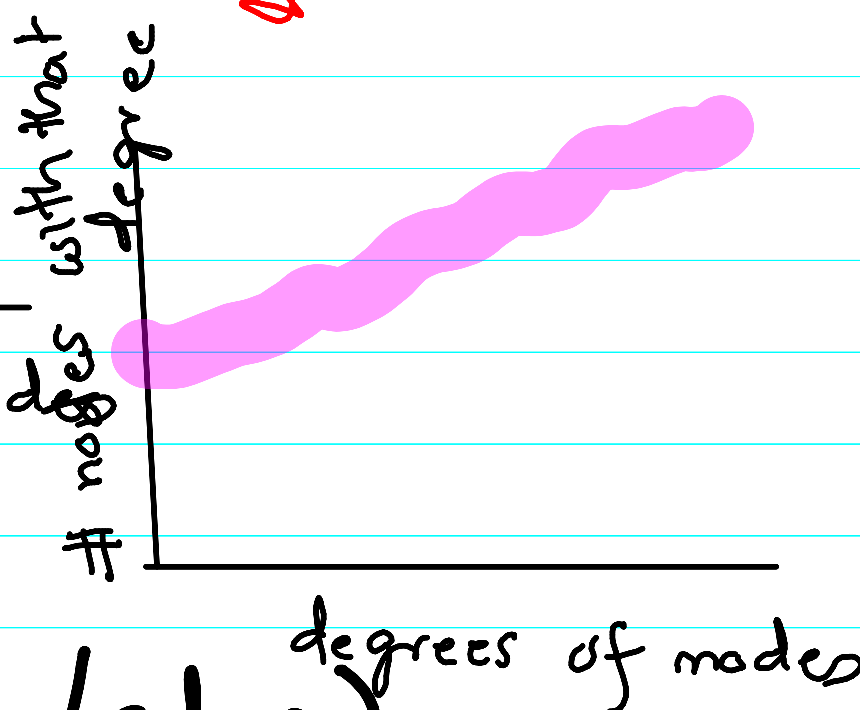
Intro to Graph Kernels

What would be useful properties to consider while computing kernels on graphs?

1. Summaries: (eg: histograms)

Not similar histograms





(eg. # cycles / edges)

2. Reachability of nodes (in p steps)

summary

match nodes

③ Soft matching.

With all this "soft" matching, how do we know we are computing "valid" "kernels"

So far: we came up with a "long" ϕ

Eg: all subsequences of characters or even subsets of characters or subtrees

We would efficiently compute $K(x_1, x_2) = \phi^T(x_1)\phi(x_2)$

unconfirmed

Hereafter:

We will define "intuitively good" kernels $K(x_1, x_2)$

Today we discuss theory that confirms if K is a kernel

In other words:

We have a good application driven K

$$\Downarrow$$

Q: Does $\exists$ a ϕ s.t $K(x_1, x_2) = \phi^T(x_1)\phi(x_2)$?

Part 1.

If $K(x_1, x_2)$ was indeed equal to $\phi^T(x_1)\phi(x_2)$ for some finite $\phi : \mathcal{X} \to \mathbb{R}^n$ (where we will hereafter assume that $x_1, x_2 \in \mathcal{X}$)

space of Graphs | strings etc

and let us say I computed the K matrix (**Gram matrix**) on examples $x_1, x_2 \ldots x_n$

$$K = \begin{bmatrix} K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_n) \\ K(x_2,x_1) & \cdots & \cdots & K(x_2,x_n) \\ \vdots & K(x_i,x_j) & & \vdots \\ K(x_n,x_1) & \cdots & \cdots & K(x_n,x_n) \end{bmatrix}$$

then what properties must K matrix satisfy?

1. $K(x_i,x_j) = \phi^T(x_i)\,\phi(x_j)$

$$= \phi^T(x_j)\,\phi(x_i) = K(x_j,x_i)$$

(Symmetric)

H/w: Verify symmetry & other theory that follows for all examples such as recursive definitions involving $K(x_i, x_j)$ that we discussed so far

② $K(x_i, x_i) \geq 0$ (since $K(x_i, x_i) = \|\phi(x_i)\|^2 \geq 0$)

③ $$K = \begin{bmatrix} \phi^T(x_1)\phi(x_1) & \phi^T(x_1)\phi(x_2) & \cdots & \phi^T(x_1)\phi(x_n) \\ \vdots & \phi^T(x_i)\phi(x_j) & & \vdots \\ \phi^T(x_n)\phi(x_1) & \cdots & \cdots & \phi^T(x_n)\phi(x_n) \end{bmatrix}$$

Let $\Psi = \begin{bmatrix} — & \phi^T(x_1) & — \\ — & \phi^T(x_2) & — \\ \vdots & & \\ — & \phi^T(x_n) & — \end{bmatrix}$

$$\Psi^T = \begin{bmatrix} | & | & & | \\ \phi(x_1) & \phi(x_2) & \cdots & \phi(x_n) \\ | & | & & | \end{bmatrix}$$

$\therefore K = \Psi \Psi^T$

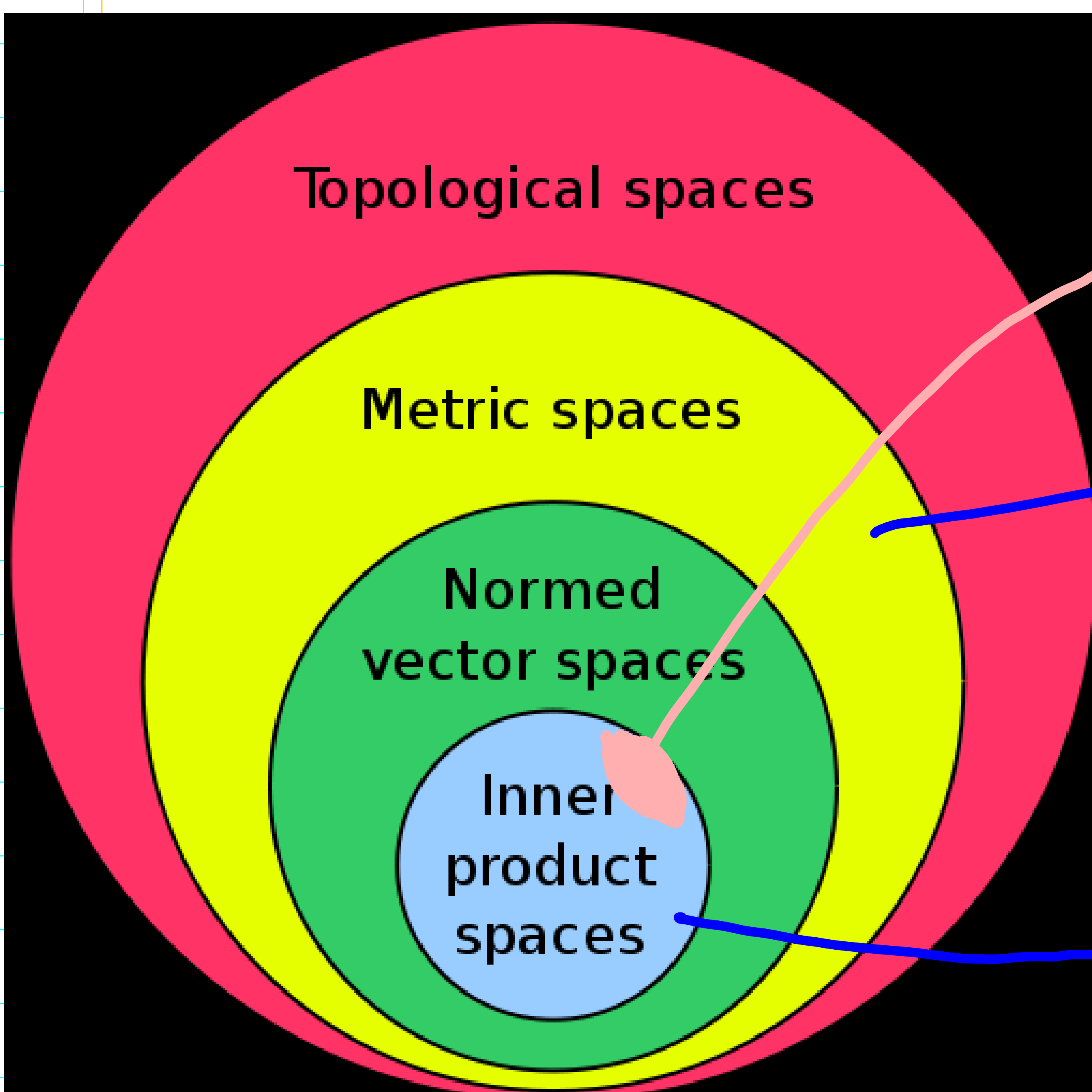
Let $\alpha \in \mathbb{R}^n$. Then

$$\begin{aligned} \alpha^T K \alpha &= \alpha^T \Psi \Psi^T \alpha \\ &= (\Psi^T \alpha)^T (\Psi^T \alpha) \\ &= \|\Psi^T \alpha\|^2 \geq 0 \end{aligned}$$

(Thus the gram matrix is always positive semi-definite p.s.d)

Q: What about infinte dimensional ϕ?

Ans: Hilbert spaces instead of $\mathbb{R}^n$



Hilbert space

Tringle inequality

$\langle v, v \rangle = \|v\|^2$

vector space with an inner prod

Think of Hilbert space as ∞ dimensional extension of euclidean space

(additional property: it should be complete

completing is not a big deal

Eg: $\mathcal{H} = \left\{ [s_1, s_2 \dots s_n \dots] \;\middle|\; \lim_{n\to\infty} \sum_{i=1}^{n} |s_i|^2 < \infty, \; s_1 \dots s_n \dots \in \mathbb{R} \right\}$

is a Hilbert space with inner (dot) product

$$\langle \bar{s}, s \rangle = \lim_{n\to\infty} \sum_{i=1}^{n} s_i \bar{s}_i$$

Verify that it satisfies properties of inner prod

Symmetry

$\langle s, \bar{s} \rangle = 0$ iff s & $\bar{s} = 0$

Cauchy Schwarz

Note the purpose of Hilbert spaces:

If $K(x_1, x_2) = \langle \psi(x_1), \psi(x_2) \rangle$

where $\psi(.) \in$ Hilbert space

& $\langle \ , \ \rangle \in$ Hilbert space,

we can similarly prove that

$$K = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & & \\ k_{n1} & \cdots & k_{nn} \end{bmatrix}$$

$K_{ii} \geq 0$

$K_{ij} = K_{ji}$

K is psd

Defn: $K(x_1, x_2) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

(we can generalize $\mathbb{R}$ to complex)

[Note: most books and papers call the function positive definite & matrix psd]

is called a positive semi-definite (psd) kernel function if

$\forall n$ & $\forall x_1 \ldots x_n \in \mathcal{X}$

the gram matrix $\begin{bmatrix} K_{11} & K_{12} \ldots K_{1n} \\ \vdots & \\ K_{n1} & \ldots K_{nn} \end{bmatrix}$

is positive semi definite

Our goal:

Given a positive semi definite kernel function $K(x_1, x_2) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

we will prove that $\exists$ a Hilbert space $\mathcal{H}$ & an inner product $\langle , \rangle$ on $\mathcal{H}$

s.t $K(x_1, x_2) = \langle \psi(x_1), \psi(x_2) \rangle$

for $\psi \in \mathcal{H}$

Read more abt Hilbert spaces in book on Learning with Kernels

http://en.wikipedia.org/wiki/Hilbert_space & associated pages

Some properties of psd kernel functions (operations giving rise to new psd kernel fns)

① If K is psd kernel fn & $\alpha \geq 0$ then $\alpha K(\cdot,\cdot)$ is also a psd kernel (function)

② If K_1 & K_2 are psd then so is $K_1(x_1,x_2) + K_2(x_1,x_2)$

③ If K_1 & K_2 are psd then so is $K_1(x_1,x_2) \cdot K_2(x_1,x_2)$

④ Polynomial fn of $K(x_1,x_2)$ is psd

provided all coefficients are ≥ 0

5 $e^{K(x_1,x_2)} = \lim_{n\to\infty} \sum_{i=0}^{n} \frac{(K(x_1,x_2))^i}{\lfloor i}$

is also psd

6 if $f(\,) : \mathcal{X} \to \mathbb{R}$

then $f(x_1) K(x_1,x_2) f(x_2)$

is also psd