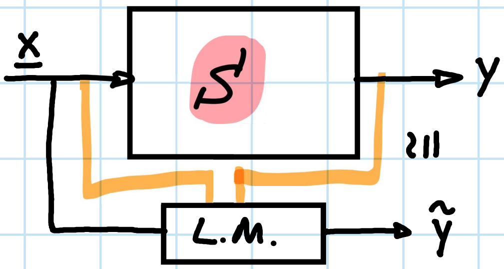
ML AND DATA ANALYSIS	REPRESENTER THEOREM AND KERNELS	October 6 2023



$\underline{x} \in \mathbb{R}^d$

$D_m = \{(\underline{x}_1, y_1), ---, (\underline{x}_N, y_N)\}$

$f(x) = \underline{w}\underline{x}$

$\ell(f(\underline{x}), y) = (y - f(\underline{x}))^2$

$C(f) = \underline{w}' I \underline{w} = \|\underline{w}\|_2^2$ EUCLEADIAN NORM SQUARED

$$\min_{\underline{w}} \|X\underline{w} - y\|^2 + \lambda\|\underline{w}\|^2$$

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_N^T \end{bmatrix} \in \mathbb{R}^{m \times d} \qquad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^m$$

$$\downarrow$$

$$\nabla_w(\|X\underline{w} - y\|^2 + \lambda\|\underline{w}\|^2) = \underline{0}$$

$$\downarrow$$

$$(\underset{d \times m}{X'}\underset{m \times d}{X} + \lambda I)\underline{w} = X'y \quad \text{OR EPUIVALENTLY} \quad \underline{w} = (X'X + \lambda I)^{+} X'y$$

THE SOLUTION OF THIS LINEAR SYSTEM COST ME $O(d^2)$.

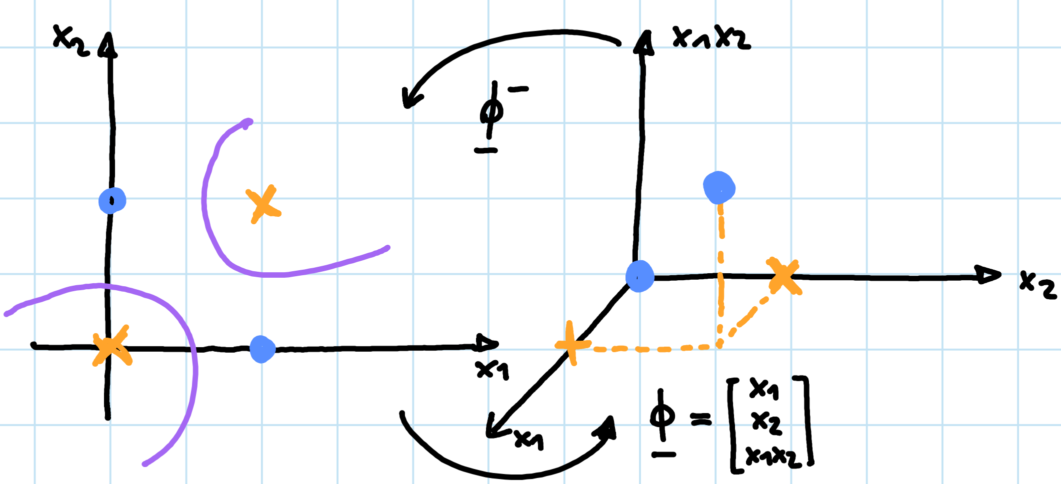
SINCE THE LINEAR HYPOTHESIS IS TOO STRONG WE USE KERNELS

$$K(\underline{x}_i, \underline{x}_j) = \underline{\phi}'(\underline{x}_i)\,\underline{\phi}(\underline{x}_j)$$

YOU DON'T HAVE TO PERFORM THIS TRANSFORMATION, I JUST DO A COMPUTATION IN THE ORIGINAL SPACE.

$$\downarrow$$

$$e^{-\gamma\|\underline{x}_i - x_j\|^2} \rightarrow |\underline{\phi}| = \infty \quad \text{INFINITE DIMENSIONAL SPACE}$$



$$\underline{\phi}^{-}$$

$$\underline{\phi} = \begin{bmatrix} x_1 \\ x_2 \\ x_1x_2 \end{bmatrix}$$

SINCE $\underline{w}$ IS ORIGINATED FROM MY DATA AND SINCE IS A LINEAR SYSTEM, PROBABLY THE SOLUTION OF MY MODEL WILL BE A LINEAR COMBINATION OF THE INFORMATION PRESENT IN MY DATASET.

$$\underset{d\times 1}{\underline{w}} = \underset{d\times m}{X'}\underset{m\times 1}{\underline{\alpha}}$$

SO $\underline{w}$ IS PARALLEL TO THE SPACE INDUCED BY MY SAMPLES.

SO IF PER ABSURDUM $\underline{w}^* = \underline{w}^*_{/\!/} + \underline{w}^*_{\perp}$ DOESN'T LIE ON THE SPACE INDUCED BY MY DATA

$$\|X\underline{w}^* - y\|^2 + \lambda\|\underline{w}^*\|^2 = \|X(\underline{w}^*_{/\!/} + \underline{w}^*_{\perp}) - y\|^2 + \lambda\|\underline{w}^*_{/\!/} + \underline{w}^*_{\perp}\|^2 =$$

$$\|X\underline{w}^*_{/\!/} + \underbrace{X\underline{w}^*_{\perp}}_{=0} - y\|^2 + \lambda\|\underline{w}^*_{/\!/} + \underline{w}^*_{\perp}\|^2 =$$

$$\|X\underline{w}^*_{/\!/} - y\|^2 + \lambda\|\underline{w}^*_{/\!/}\|^2 + \lambda\|\underline{w}^*_{\perp}\|^2 > \|X\underline{w}^*_{/\!/} - y\|^2 + \lambda\|\underline{w}^*_{/\!/}\|^2$$

CONTRADDICTION: THERE IS ANOTHER $\underline{w}^*$ THAT FINDS THE MINIMUM $\Rightarrow$ $\underline{w}^*_{\perp} = \underline{0}$

FOLLOWS $\underline{w}^* = \underline{w}^*_{/\!/}$

$$\underline{w}^* = \text{ARG}\min_{\underline{w}}\left(\|X\underline{w} - y\|^2 + \lambda\|\underline{w}\|^2\right)$$

REPRESENTER THEOREM

THE SOLUTION IS ALWAY A LINEAR COMBINATION OF THE POINTS IN MY DATA

COSTS $O(d^2)$

$$\min_{\underline{w}}\left(\|X\underline{w} - y\|^2 + \lambda\|\underline{w}\|^2\right) = \min_{\underline{\alpha}}\left(\|\underbrace{\overset{m\times d}{X}\overset{d\times m}{X'}}_{m\times m}\underline{\alpha} - y\|^2 + \lambda\underline{\alpha}'XX'\alpha\right)$$

$XX' = Q$ MATRIX $(m\times m)$ where $Q_{ij} = \underset{\text{COMMUTATIVE}}{\underline{x}_i\underline{x}_j} = Q_{ji}$ SAMPLE i (ROW i OF X)

Q IS SIMMETRIC, AND DEFINITE SEMIPOSITIVE $Q > 0$

$$\min_{\underline{\alpha}}\left(\|Q\underline{\alpha} - y\|^2 + \lambda\underline{\alpha}'Q\underline{\alpha}\right) =$$

I CAN SIMPLIFY Q

$$\nabla_{\underline{\alpha}}\left(\underline{\alpha}'Q'Q\underline{\alpha} - 2\underline{y}'Q\underline{\alpha} + \underline{y}'\underline{y} + \lambda\underline{\alpha}'Q\underline{\alpha}\right) = \underline{0}$$

$$\cancel{2Q'}Q\underline{\alpha} - \cancel{2}\cancel{Q}\underline{y} + \cancel{2}\lambda\cancel{Q}\underline{\alpha} = \underline{0}$$

$$Q\underline{\alpha} - \underline{y} + \lambda\underline{\alpha} = \underline{0}$$

$$(Q + \lambda I)\underline{\alpha} = \underline{y} \Rightarrow \underline{\alpha} = (Q + \lambda I)^{+}\underline{y}$$

SOLUTION OF ANOTHER LINEAR SISTEM   COST $O(m^2)$

THE COMPUTATIONAL DEPENDS ON THE NUMBER OF SAMPLES. MUCH EASIER TO SOLVE IF I HAVE A VERY BIG NUMBER OF DIMENSIONS. (DUAL PROBLEM)

PRIMAL:

$$\min_{\underline{w}} \|X\underline{w} - y\|^2 + \lambda\|\underline{w}\|^2 \qquad f(\underline{x}) = \underline{w}\,\underline{x}$$

SOLUTION: $\underline{w}^* = (X'X + \lambda I)^+ X'\underline{y}$ $\qquad O(d^2)$

DUAL:

LINEAR COMBINATION OF ELEMENTS IN MY DATASET

$$\min_{\underline{\alpha}} \|Q\underline{\alpha} - y\|^2 + \lambda\underline{\alpha}'Q\underline{\alpha} \qquad \underline{w} = \sum_{i=1}^{m} \alpha_i \underline{x}_i = X'\underline{\alpha} \Rightarrow f(\underline{x}) = \sum_{i=1}^{m} \alpha_i \underline{x}_i' \cdot \underline{x}$$

SOLUTION: $\underline{\alpha}^* = (Q + \lambda I)^+ \underline{y}$ $\qquad O(m^2)$

$Q_{ij} = \underline{x}_i \underline{x}_j$

D DIMENSION

LET'S CHANGE, NOW WE DO LINEAR MODEL IN A ϕ SPACE THAT WE SUPPOSE TO KNOW

$\phi: \mathbb{R}^d \rightarrow \mathbb{R}^D$

$$\underset{m \times D}{X} = \begin{bmatrix} \phi'(x_1) \\ \vdots \\ \phi'(x_m) \end{bmatrix} \qquad f(\underline{x}) = \underline{w}\,\underline{\phi}(\underline{x})$$

$$\min_{\underline{w}} \|X\underline{w} - y\|^2 + \lambda\|\underline{w}\|^2$$

$$\underline{w}^* = (X'X + \lambda I)^+ X'\underline{y} \qquad O(D^2) \qquad \text{PRIMAL}$$

DUAL:

$$\min_{\underline{\alpha}} \|Q\underline{\alpha} - y\|^2 + \lambda\underline{\alpha}'Q\underline{\alpha} \qquad \underline{w} = \sum_{i=1}^{m} \alpha_i \underline{\phi}(\underline{x}_i) = X'\underline{\alpha} \Rightarrow f(\underline{x}) = \sum_{i=1}^{m} \alpha_i \underbrace{\underline{\phi}'(\underline{x}_i)\phi(\underline{x})}_{K(\underline{x}_i, \underline{x}_j)}$$

$$\underline{\alpha}^* = (Q + \lambda I)^+ \underline{y} \qquad O(m^2)$$

$Q_{ij} = \underline{\phi}'(\underline{x}_i)\phi(\underline{x}_j) = K(\underline{x}_i, \underline{x}_j)$

I AM PROJECTING THE POINTS IN A NEW DIMENSIONAL SPACE.

IN THE DUAL I AM ABLE TO LEARN A MODEL α GENERATED FROM A PROJECTION IN A INF. DIMENSIONAL

SPACE $\rightarrow$ WE ARE ABLE TO DO NON LINEAR MODEL, AS A PRICE I CANNOT USE THE PRIMAL BECAUSE I DON'T KNOW $\underline{\phi}$ $\Rightarrow$ IF I HAVE A LOT OF DATA I STILL TO PAY $O(m^2)$.

AS A PLUS I WORK IN A INFINITE DIMENSIONAL SPACE.

KERNEL RIDGE REGRESSION (KRLS)

$$f(\underline{x}) = \sum_{i=1}^{m} \alpha_i \, K(\underline{x}_i, \underline{x}_j)$$

- THINGS THAT I DON'T KNOW HOW THEY WORK FOR NOW

$$\underline{\alpha} = (Q + \lambda I)^{+} \underline{y}$$

$$Q_{ij} = K(\underline{x}_i, \underline{x}_i) = e^{-\gamma \| \underline{x}_i - \underline{x}_j \|^2}$$

KERNEL

$K(\underline{u}, \underline{v}) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$

WHEN A FUNCTION IS A KERNEL?

THERE IS A MATHEMATICAL DEFINITION (COMPLICATED, THE INTEGRAL OF THE FUNC. AND OF THE KERNEL HAS TO BE FINITE).

$$\langle \underline{u}, \underline{v} \rangle^d , \quad (\langle \underline{u}, \underline{v} \rangle + p)^d , \quad e^{-\gamma \| \underline{u} - \underline{v} \|^2}$$

POLYNOMIAL KERNEL

GAUSSIAN KERNEL (99% USED FOR ML)

WHAT IS THE MEANING OF γ?

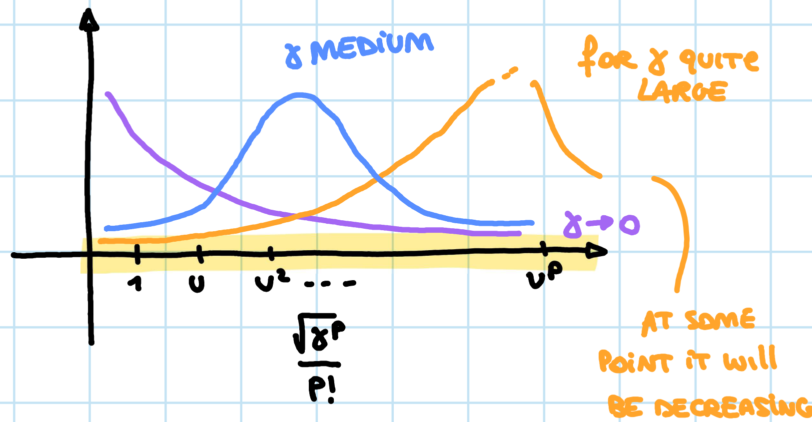
$e^{-\gamma \| \underline{u} - \underline{v} \|^2}$

CONSIDER THE MONODIMENSIONAL CASE $(d=1)$ $\Rightarrow$ $\underline{\phi} : \mathbb{R} \rightarrow \mathbb{R}^{\infty}$

$$e^{x} = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$e^{-\gamma \| \underline{u} - \underline{v} \|^2} = e^{-\gamma \underline{u}^2} e^{-\gamma \underline{v}^2} e^{2\gamma u v} = e^{-\gamma u^2} e^{-\gamma v^2} \sum_{i=0}^{\infty} \frac{(2\gamma u v)^i}{i!} =$$

$$= \underbrace{e^{-\gamma \|\underline{u}\|^2}\begin{bmatrix} 1 \\ \sqrt{2\gamma}\, u \\ \sqrt{\frac{(2\gamma)^2}{2}}\, u^2 \\ \vdots \\ \vdots \end{bmatrix}}_{\underline{\phi}(\underline{u})} \underbrace{e^{-\gamma \|\underline{v}\|^2}\begin{bmatrix} 1 \\ \sqrt{2\gamma}\, v \\ \vdots \\ \sqrt{\frac{(2\gamma)^p}{p!}}\, v^p \\ \vdots \end{bmatrix}}_{\underline{\phi}(\underline{v})}$$



γ REGULATES THE NON LINEARITY OF THE SOLUTION (FILTER)

γ SMALL, SOLUTION IS MORE LINEAR

γ LARGE, SOLUTION IS MORE NON LINEAR

WHAT IS λ?

$$\min_{\underline{w}} \|X\underline{w} - \underline{y}\|^2 + \lambda \|\underline{w}\|^2$$

↓ BECOMES

$\underline{\alpha} = (Q + \lambda I)^{+}\, \underline{y}$ WHY I HAVE TO ADD λ VALUES TO THE DIAGONAL OF Q?

IF $\lambda = 0 \Rightarrow \underline{\alpha} = Q^{+}\underline{y}$ WE LEARN ALL THE THINGS THAT WE CARE FROM THE DATA

IF $\lambda = \infty \Rightarrow \underline{\alpha} = \frac{\underline{y}}{\lambda} = \underline{0}$ WE DON'T CARE OF LEARNING FROM THE DATA, WE JUST CARE TO HAVE A SIMPLE SOLUTION

IF Q IS NOT INVERTIBLE ADDING λI WE HAVE $(Q + \lambda I)$ THAT IS INVERTIBLE $\underline{\alpha} = (Q + \lambda I)^{-1}\underline{y}$

(ONLY ONE SOLUTION)

main.m

```
clc

%%
im = imread('Cat.jpeg','jpg');
im = rgb2gray(im);
im = double(im);

figure, imagesc(im), colormap gray

[U,S,V] = svd(im);
neigketp = 300;
S(1000:end,1000:end) = 0;
im = U*S*V';

figure, imagesc(im), colormap gray
```

WHAT HAPPENS TO THE IMAGE WHEN I START TO REMOVE VALUES FROM THE MATRIX Q.

DON'T TRY TO LEARN ALONG THE AXIS. TRY JUST TO LEARN FROM THE INFORMATION THAT IS MORE STRONG.