# Lec4 - Kernels

Monday, 24 June 2024 16:02

מטרות:

* פונ' מיפוי רב-ממדית
* טריק מתמטי לעבודה עם פונ' מיפוי

הבהרה: יש הרבה אלגברה ליניארית

* מיפוי רב-ממדי

$$\underbrace{\phi(\cdot)}_{\text{פונ' מיפוי}}: R^{\overset{\text{מימד מקור}}{L}} \to R^{\overset{\text{מימד חדש}}{N}}$$

תזכורת:

$\phi(x_k) \to \{1, x_k, x_k^2, \dots, x_k^M\}$

מיצוי לינאר׳ ← $\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^N \\ 1 & x_2 & x_2^2 & \cdots & x_2^N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_M & x_M^2 & \cdots & x_M^N \end{bmatrix}$

נתונים מקוריים

הרבה פולי'

רישום ע"י פונ' מיפוי

$\phi(x_k) = \begin{bmatrix} 1 & x_k & x_k^2 & \cdots & x_k^N \end{bmatrix}$

מספר שורה

* מקרה חד-ממד'

דוגמה:

L=2

$\phi(x_{k_1}, x_{k_2}) = \left[1, x_{k_1}, x_{k_2}, x_{k_1}x_{k_2}, x_{k_1}^2, x_{k_2}^2\right] \to X$ שורות של מטריצה

$M \times N$ (שורות, עמודה)

L=2 N=6

k=1,...,M

מטרה: נתונת ממדים של $X^TX$ ($N \times N$)

$$\mathbf{w}_{opt} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

דוגמה:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} \in \mathbb{R}^{M \times 2} \quad N=2$$

וקטור שורה של נתונים

$\underbrace{x_1^T}_{\substack{1\times M \\ \text{וקטור שורה}}} \cdot \underbrace{x_1}_{\substack{M\times 1 \\ \text{וקטור עמודה}}}$ → $1\times 1$ סקלר

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T\mathbf{x}_1 & \mathbf{x}_1^T\mathbf{x}_2 \\ \mathbf{x}_2^T\mathbf{x}_1 & \mathbf{x}_2^T\mathbf{x}_2 \end{bmatrix}$$

סקלר

## Kernel Trick

$$\mathbf{w} = \Big(\underbrace{\mathbf{X}^T\mathbf{X}}_{N\times N} + \lambda\mathbf{I}\Big)^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\Big(\underbrace{\mathbf{X}\mathbf{X}^T}_{M\times M} + \lambda\mathbf{I}\Big)^{-1}\mathbf{y}$$

ניתוח $\tilde{X}X^T$

עמודה של $X$, מתוך $N$ עמודות

$$\mathbf{x}_k = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{Mk} \end{bmatrix}$$

מספר שורה

מספר עמודה

M- מספר שורות

מימד, בלתי תלוי

בפונ' מיפוי

$$\mathbf{X}\mathbf{X}^T = \begin{bmatrix} \tilde{\mathbf{x}}_1^T \\ \vdots \\ \tilde{\mathbf{x}}_j^T \\ \vdots \\ \tilde{\mathbf{x}}_M^T \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_1 & \vdots & \tilde{\mathbf{x}}_j & \vdots & \tilde{\mathbf{x}}_M \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{\mathbf{x}}_1^T\tilde{\mathbf{x}}_1 & \cdots & \tilde{\mathbf{x}}_1^T\tilde{\mathbf{x}}_M \\ \tilde{\mathbf{x}}_2^T\tilde{\mathbf{x}}_1 & \cdots & \tilde{\mathbf{x}}_2^T\tilde{\mathbf{x}}_M \\ \vdots & \vdots & \vdots \\ \tilde{\mathbf{x}}_M^T\tilde{\mathbf{x}}_1 & \cdots & \tilde{\mathbf{x}}_M^T\tilde{\mathbf{x}}_M \end{bmatrix}$$

$1\times N \quad N\times 1 \to 1\times 1$

שורה של $X$

$$\mathbf{X} = \begin{bmatrix} \boxed{x_{11} \quad x_{12}} \\ \vdots \quad\quad \vdots \\ x_{j1} \quad x_{j2} \\ \vdots \quad\quad \vdots \\ x_{M1} \quad x_{M2} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{x}}_1^T \\ \vdots \\ \tilde{\mathbf{x}}_j^T \\ \vdots \\ \tilde{\mathbf{x}}_M^T \end{bmatrix}$$

$\tilde{\mathbf{x}}_j^T \to 1\times N$

$\tilde{x}_j^T$ ← שלוב של פונ' מיפוי

$$\begin{bmatrix} \tilde{\mathbf{x}}_M^T\tilde{\mathbf{x}}_1 & \cdots & \tilde{\mathbf{x}}_M^T\tilde{\mathbf{x}}_M \end{bmatrix}$$

$1\times M \quad N\times 1 \to 1\times 1$

$x_j \leftarrow$ שילוב של פוני מיפוי

## Kernel

שילוב פוני מיפוי

דוגמה: רישום ע"י סימונים לעיל

$$\phi(\tilde{\mathbf{x}}_j^T) = \left[1, x_{j1}, x_{j2}, x_{j1}x_{j2}, x_{j1}^2, x_{j2}^2\right] = \phi(\tilde{\mathbf{x}}_j)^T$$

הפעלה של פוני מיפוי על מטריצה $\underline{X}$

$$\phi(\mathbf{X}) = \begin{bmatrix} \phi\left(\tilde{\mathbf{x}}_1^T\right) \\ \vdots \\ \phi\left(\tilde{\mathbf{x}}_j^T\right) \\ \vdots \\ \phi\left(\tilde{\mathbf{x}}_M^T\right) \end{bmatrix} \in \mathbb{R}^{M\times N}$$

לאחר מיפוי

$$K(\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_k) = \underbrace{\phi(\tilde{\mathbf{x}}_j)^T\phi(\tilde{\mathbf{x}}_k)}_{\text{סקלר}}$$

של זוג וקטורים

**Kernel matrix** The matrix $\mathbf{K} \in \mathbb{R}^{M\times M}$,

$$\mathbf{K} = \begin{bmatrix} K(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_1) & \cdots & K(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_M) \\ K(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1) & \cdots & K(\tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_M) \\ \vdots & \vdots & \vdots \\ K(\tilde{\mathbf{x}}_M, \tilde{\mathbf{x}}_1) & \cdots & K(\tilde{\mathbf{x}}_M, \tilde{\mathbf{x}}_M) \end{bmatrix}$$

במקום $XX^T$

$$\phi(\mathbf{X})\phi(\mathbf{X})^T = \begin{bmatrix} \phi(\tilde{\mathbf{x}}_1)^T\phi(\tilde{\mathbf{x}}_1) & \cdots & \phi(\tilde{\mathbf{x}}_1)^T\phi(\tilde{\mathbf{x}}_M) \\ \phi(\tilde{\mathbf{x}}_2)^T\phi(\tilde{\mathbf{x}}_1) & \cdots & \phi(\tilde{\mathbf{x}}_2)^T\phi(\tilde{\mathbf{x}}_M) \\ \vdots & \vdots & \vdots \\ \phi(\tilde{\mathbf{x}}_M)^T\phi(\tilde{\mathbf{x}}_1) & \cdots & \phi(\tilde{\mathbf{x}}_M)^T\phi(\tilde{\mathbf{x}}_M) \end{bmatrix}$$

סקלר

מטרה של kernel ← להחליף פוני מיפוי

ולהשתלב בתוך:

$$w = \underbrace{\mathbf{X}^T}_{\phi(X)^T \; (3)} \underbrace{\left(\underbrace{\mathbf{X}\mathbf{X}^T}_{(1)\,K} + \lambda\mathbf{I}\right)^{-1}\mathbf{y}}_{(2)}$$

חזרה ל kernel + מטריצת
Kernel trick

(4) שורה "חדשה" $\tilde{x}_0^T \rightarrow \hat{y}_0$

חיזוי "ליניארי"

$$\hat{y}_0 = \varphi(\tilde{\mathbf{x}}_0)^T\mathbf{w}$$

שלבי חישוב של מקדמי $w$

(1) חישוב מטריצה $\underline{K}$

$$\underset{M\times 1}{\boldsymbol{\alpha}} = (\underset{M\times M}{\mathbf{K}} + \lambda\mathbf{I}_M)^{-1}\underset{M\times 1}{\mathbf{y}} \quad (2)$$

וקטור עמודה עם $M$ שורות

בתום החישוב, אין צורך יותר לשמור של $\underline{K}$

$$\underset{N\times 1}{\mathbf{w}} = \underbrace{\varphi(\mathbf{X})^T}_{N\times M}\underbrace{\boldsymbol{\alpha}}_{M\times 1} \quad (3)$$

חישוב מחדש של $\hat{y}_0$ ללא צורך
בחישוב מקדמי $w$ בצורה מפורשת

$$w = \sum_{j=i}^{M} \alpha_j\varphi\left(\tilde{\mathbf{x}}_j\right)$$

דרך
נוספת
לרישום

$$\hat{y}_0 = \sum_{j=i}^{M} \alpha_j\varphi(\tilde{\mathbf{x}}_0)^T\varphi\left(\tilde{\mathbf{x}}_j\right)$$

$$= \sum_{j=i}^{M} \alpha_j\underbrace{\varphi(\tilde{\mathbf{x}}_0)^T\varphi\left(\mathbf{X}\right)}_{K(\tilde{x}_0, \underline{X})}$$

$$= K\left(\tilde{\mathbf{x}}_0, \mathbf{X}\right)\boldsymbol{\alpha}$$

חישוב בין $\tilde{x}_0$ לבין עמודות של $\underline{X}$ (מקוריות)

סיכום:

1. חישוב ערך של וקטור $\alpha$
2. חישוב $K(\;)$
3. חיזוי ללא צורך בערך $\underline{w}$ בצורה מפורשת

סיכום ממדים

מיפוי של שורה חדשה של מונים

$$\varphi(\tilde{\mathbf{x}}_0)^T \in \mathbb{R}^{1\times N},$$
$$\varphi\left(\mathbf{X}\right)^T \in \mathbb{R}^{N\times M}, \varphi\left(\mathbf{X}\right) \in \mathbb{R}^{M\times N}$$
$$K\left(\tilde{\mathbf{x}}_0, \mathbf{X}\right) \in \mathbb{R}^{1\times M}$$

Kernel ...

# Kernel functions

## Polynomial Kernel

$$K(\mathbf{x},\mathbf{y}) = \left(\mathbf{x}^{\mathsf{T}}\mathbf{y} + c\right)^{d}$$

L עמודות
→ N

c קבוע
d חזקה.

$$N = \binom{d+L-1}{d}$$

$$L = 10, d = 4, N = \binom{13}{4} = 715$$

תזכורת: חישוב מסובב kernel לא דורש
N עמודות באופן מפורש

**Example** for $c = 0, d = 3, L = 2$:
עמודות

$$K(\mathbf{a},\mathbf{b}) = \left(\mathbf{a}^T\mathbf{b}\right)^3 = \left(\underbrace{[a_1, a_2]}_{L\ \text{ערכים}} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\right)^3$$

L ערכים

$$= (a_1b_1 + a_2b_2)^3$$
$$= a_1^3b_1^3 + 3a_1^2b_1^2a_2b_2 + 3a_1b_1a_2^2b_2^2 + a_2^3b_2^3$$
$$= \left[a_1^3, \sqrt{3}a_1^2a_2, \sqrt{3}a_1a_2^2, a_2^2\right] \cdot \left[b_1^3, \sqrt{3}b_1^2b_2, \sqrt{3}b_1b_2^2, b_2^2\right]^T$$
$$= \varphi(\mathbf{a})^T\varphi(\mathbf{b})$$

לא צריך בכלל
מיפוי בצורה מפורשת

דוגמה עבור L שרירותי
$C=0,\ d=2$

$$\varphi(\mathbf{a}) = \langle 1, \sqrt{2}a_1, \sqrt{2}a_1, \ldots, \sqrt{2}a_L, a_1^2, a_2^2, \ldots, a_L^2,$$
$$\sqrt{2}a_1a_2, \sqrt{2}a_1a_3, \ldots, \sqrt{2}a_1a_L, \sqrt{2}a_2a_3, \ldots, \sqrt{2}a_{L-1}a_L\rangle$$

overfitting — קטן b
under fitting — גדול b

## Gaussian Radial Basis Kernel (RBK)

The kernel definition is

Euclidean distance
$(x_1-y_1)^2+(x_2-y_2)^2+\ldots$

$$K(\mathbf{x},\mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2b}\right)$$

Due to Taylor expansion,

$$\exp(x) \cong 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

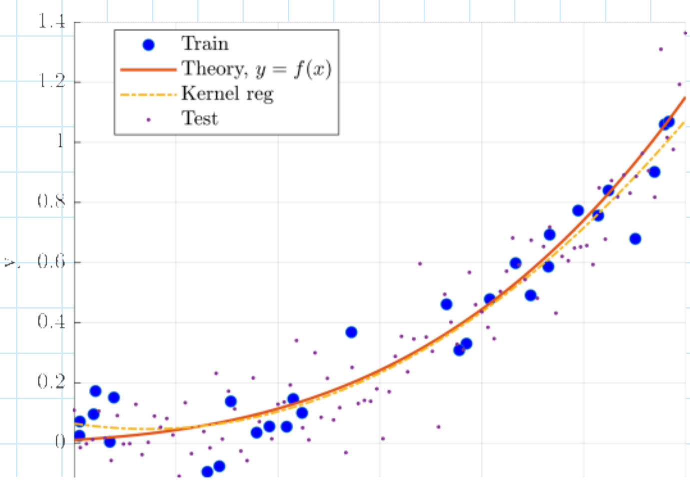
this kernel has $N \to \infty$.

kernel func

וקטורים

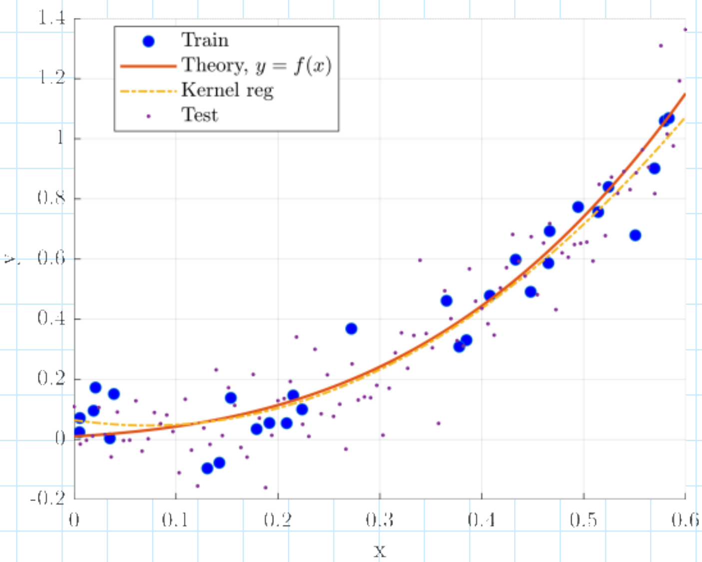
```
%% Kernel regression
lambda = 0.01;
c = 1;
N = 5;
K_func = @(x1, x2) (x1*x2'+c).^N;
K = K_func(x_train,x_train);
alpha = (K + lambda*eye(M_train))\y_train;
yh = K_func(x_test,x_train)*alpha;
```

① מטריצת kernel
②
④




$d=5$
$c=1$
$M_{train}=30$

```
K_rbf = @(x1, x2) exp(-(x1-x2').^2/2/b);
```



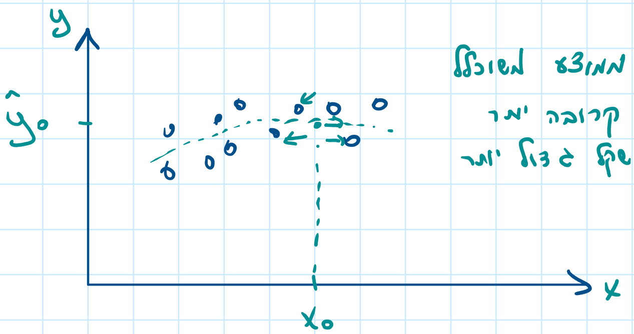

$d=5$
$c=1$
$M_{train}=30$

```
K_rbf = @(x1, x2) exp(-(x1-x2').^2/2/b);
```

**סיכום:**

* N גבוה שרירותית
* היפר-פרמטרים: $\lambda$
Kernel func + פרמטרים שלה
* מגבלה של חישוב עבור מטריצה MxM (מטריצת kernel)
* מקדמי חיזוי $\underline{w}$ לא מחושבים בצורה מפורשת
⇐ פחות ברור השפעה של כל עמודה ב-$\bar{X}$ על המודל.

## הרלקה ע"י Kernel




* עושים ממוצע משוקלל
* נקודה קרובה יותר
⇐ משקל גדול יותר

שיטה: עבור כל הנקודות

$$\hat{y}_0 = \sum_{j=1}^{M} \mathbf{w}_j y_j$$

משקלים

חישוב משקלים:

$$\tilde{\mathbf{w}} = K(\tilde{\mathbf{x}}_0, \mathbf{X})$$

נרמול סכום המשקלים ל-1:

$$w_i = \frac{\tilde{w}_i}{\sum_{j=1}^{M} \tilde{w}_j}$$

סיכום השיטה: * חישוב משקלים

$$\tilde{\mathbf{w}} = K(\mathbf{X}_{\text{test}}, \mathbf{X}_{\text{train}}) \quad \in \mathbb{R}^{M_{\text{test}} \times M_{\text{train}}}$$

* נרמול עמודות ← W

$$\hat{\mathbf{y}}_{\text{test}} = \mathbf{W}\mathbf{y}_{\text{train}}$$

פרמטר של kernel
בוני kernel
קישוב W
נרמול עמודות

```
b = .001;
% b = .00005;
K_rbf = @(x1, x2) exp(-(x1-x2').^2/2/b);
W = K_rbf(x_test,x_train);
W= W./sum(W,2);
yh = W*y_train;
```

עמודות

סיכום:

* יחסית מעט חישובים
* היפר פרמטר: kernel + par
* בג'ג דורש נתונים "צפופים"