Tehnici de Optimizare

Facultatea de Matematica si Informatica

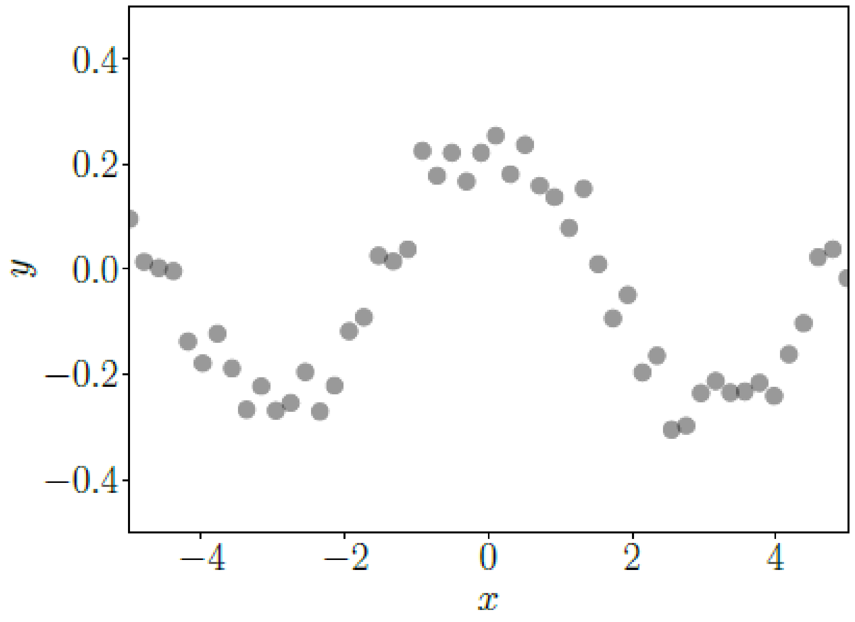
Universitatea Bucuresti

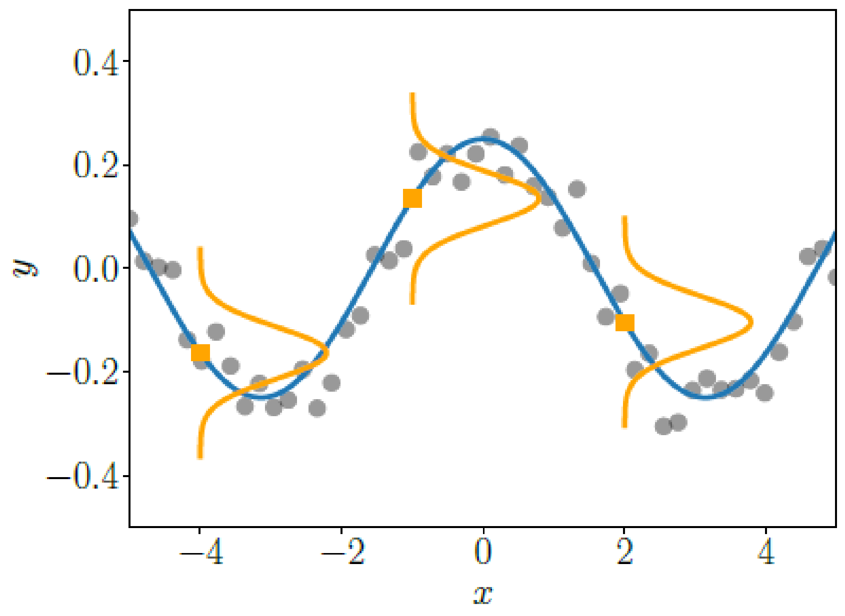
- Department Informatica-

2021

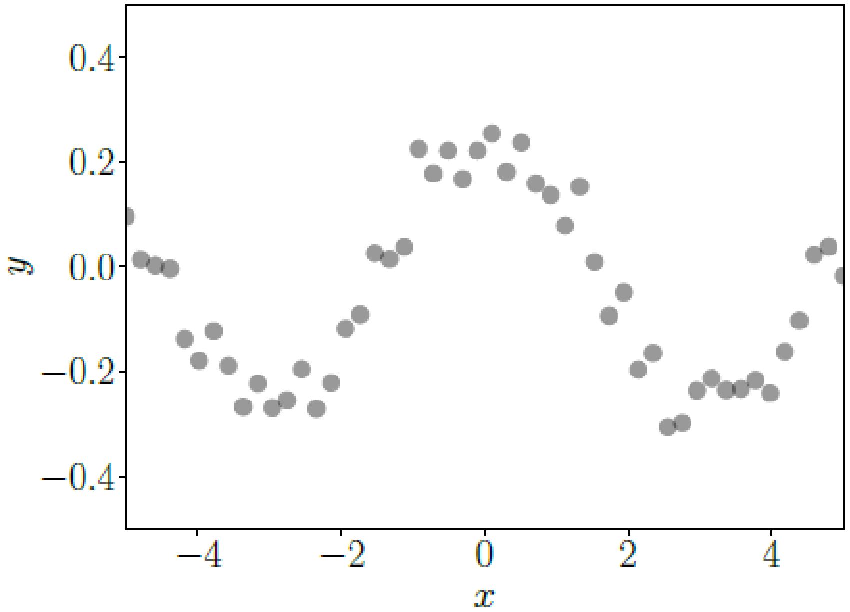
Aplicatii: Regresie si interpolare

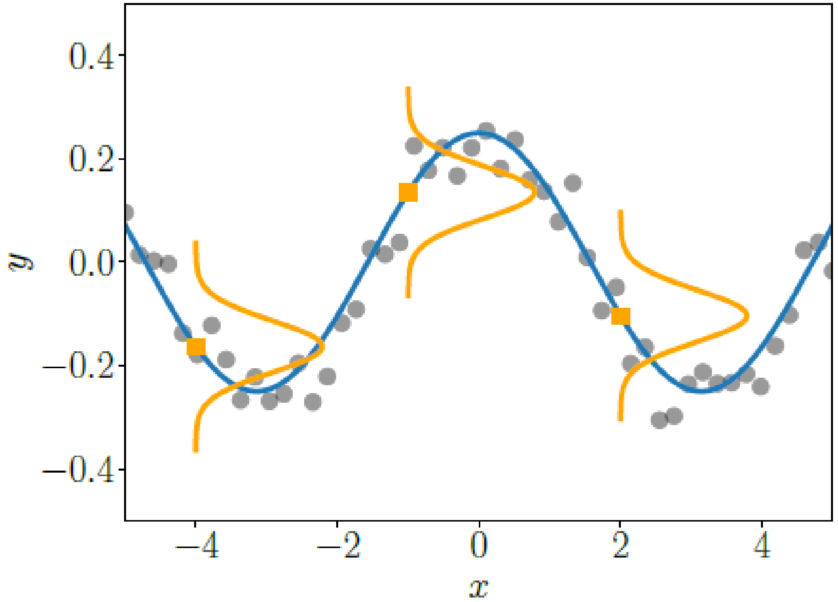
Regresie liniara





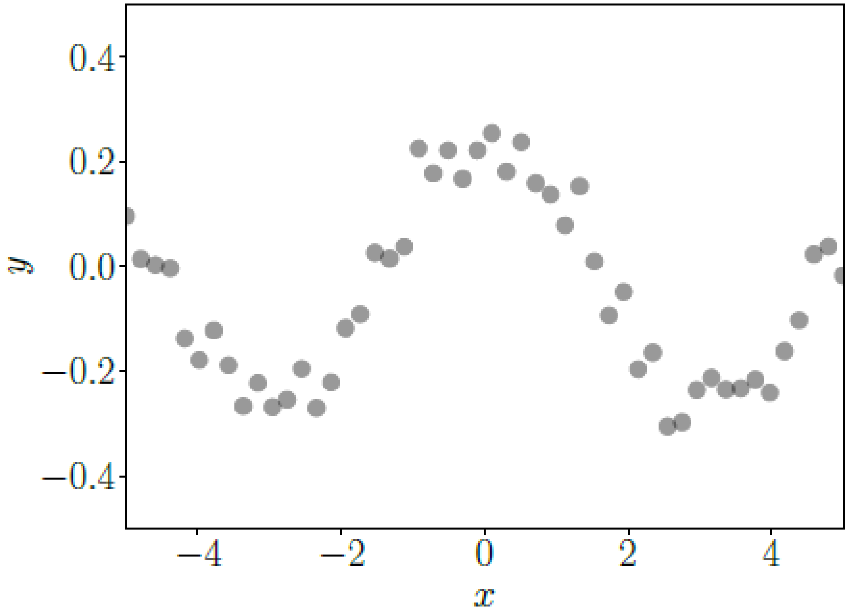
Regresie liniara

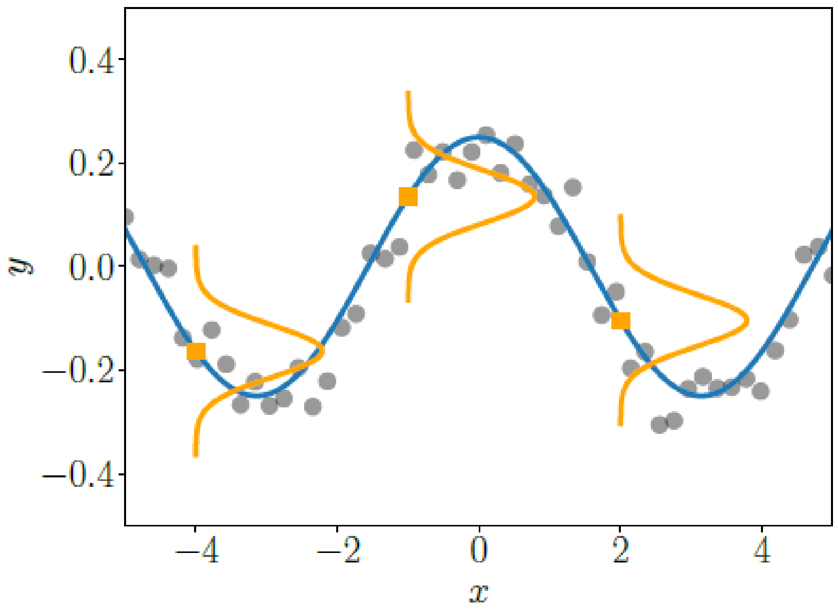




$$y_n = f(x^n) + \epsilon$$

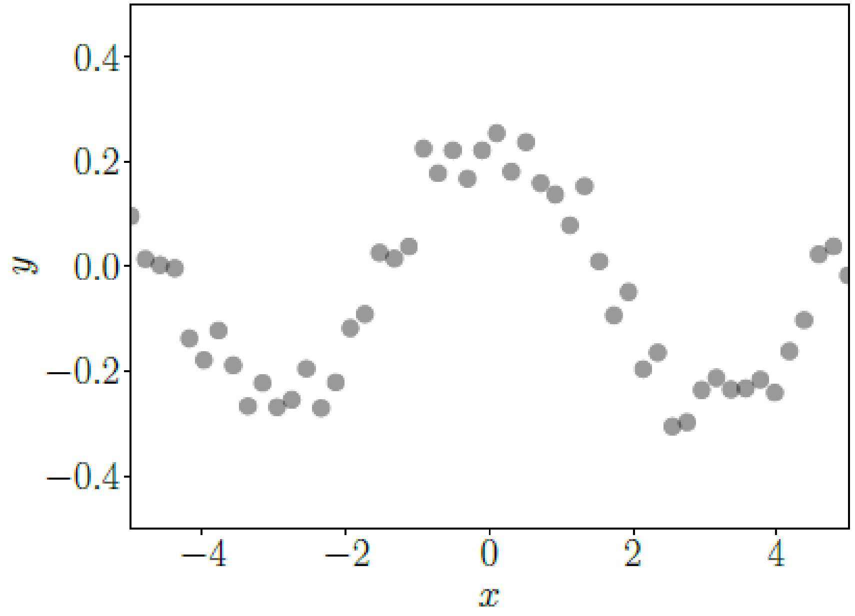
Regresie liniara

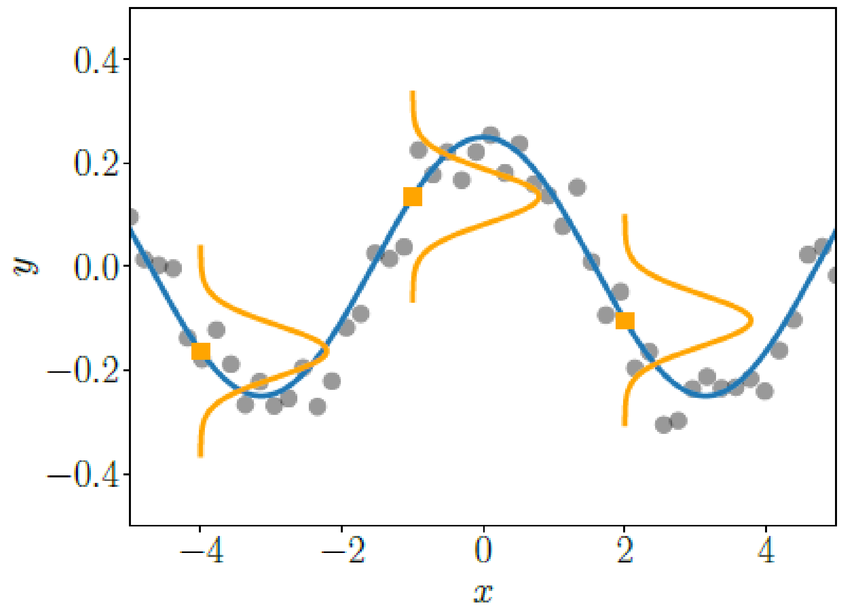




$y_n = f(x^n) + \epsilon \Rightarrow$ presupunem zgomot gaussian de medie 0 si varianta $\sigma^2, i.e. \epsilon \sim N(0, \sigma^2)$

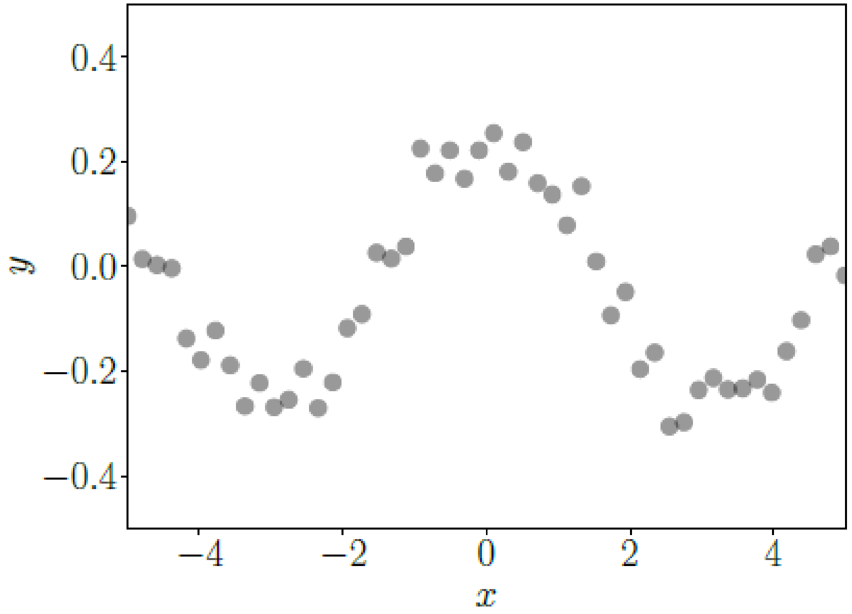
Regresie liniara

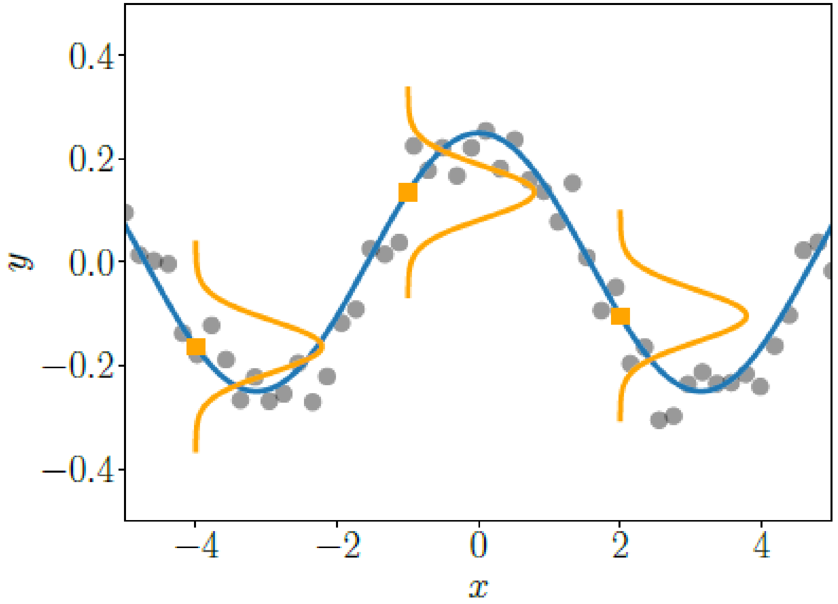




Zgomot Gaussian $\Rightarrow p(\epsilon|0,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right)$

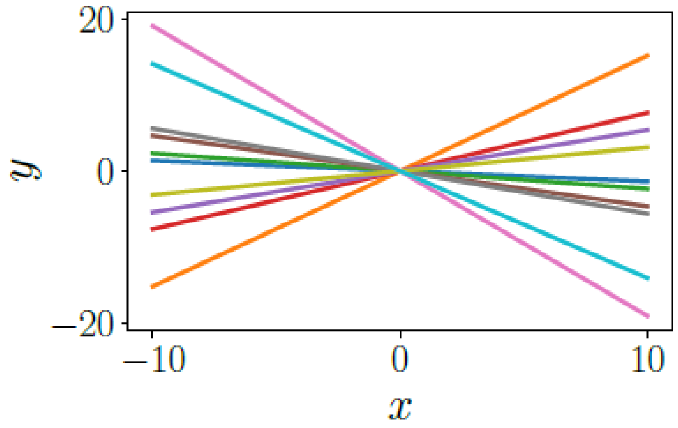
Regresie liniara

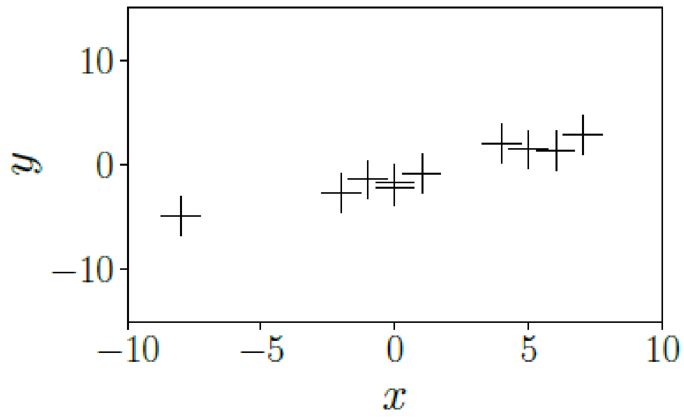


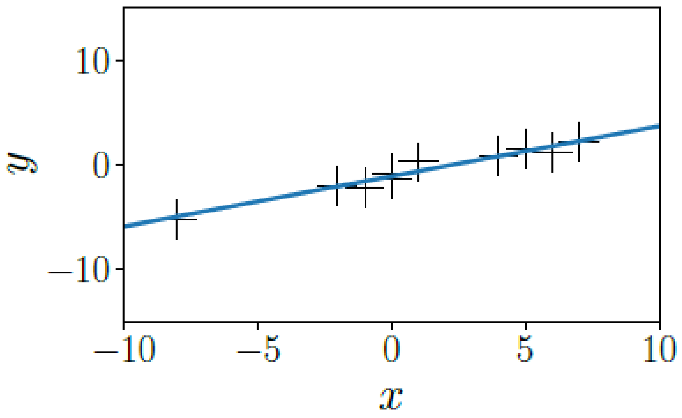


Zgomot Gaussian $\Rightarrow p(\epsilon|0,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) \Leftrightarrow p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-f(x))^2}{2\sigma^2}\right)$

Regresie liniara



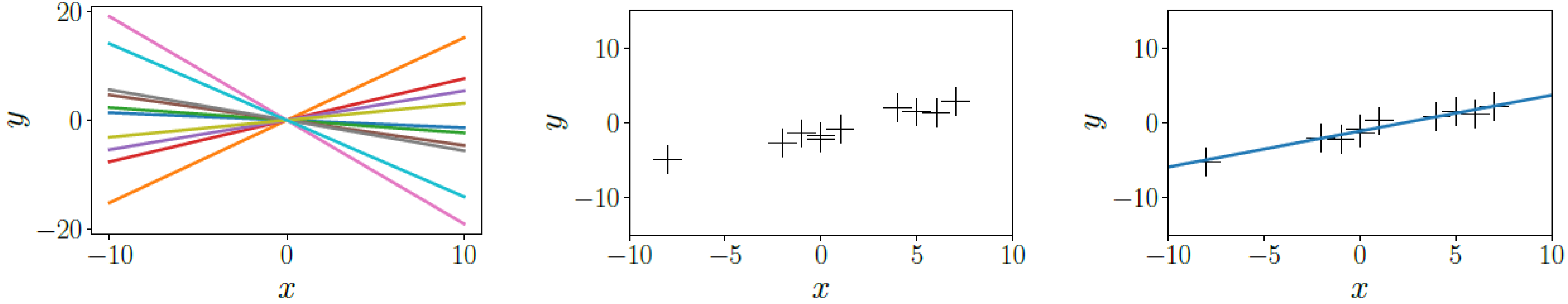




Alegem un model linear: $\mathrm{f(x)} = \theta^T x = \sum_i \theta_i x_i$

$p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\theta^T x)^2}{2\sigma^2}\right)$ (functie de verosimilitate)

Regresie liniara



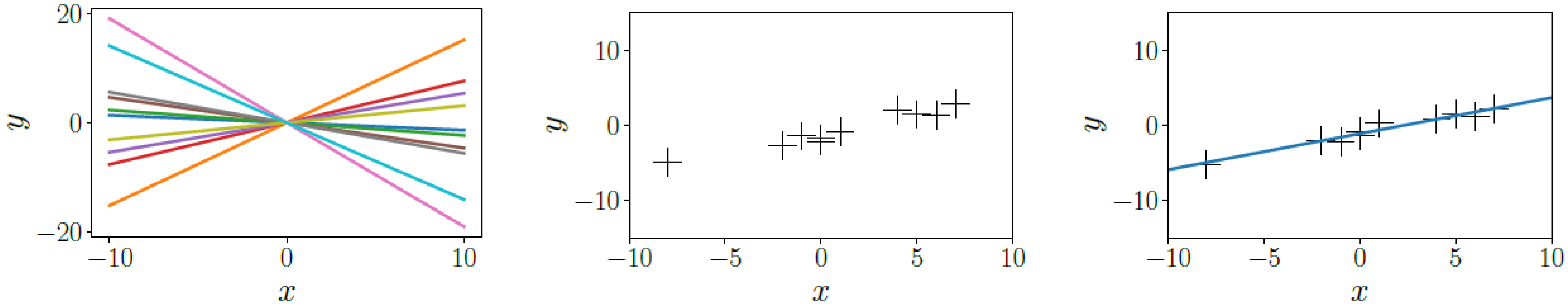
Alegem un model linear: $\mathrm{f(x)} = \theta^T x$

$p(y|x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\theta^T x)^2}{2\sigma^2}\right)$ (functie de verosimilitate)

Set de antrenare: $\{(x^1, y_1), \dots, (x^n, y_n)\}$, masuram cumulative

$$p(Y|X,\theta) = p(y_1, \dots, y_n | x^1, \dots, x^n) = \prod_n p(y_n | x^n, \theta)$$

Regresie liniara

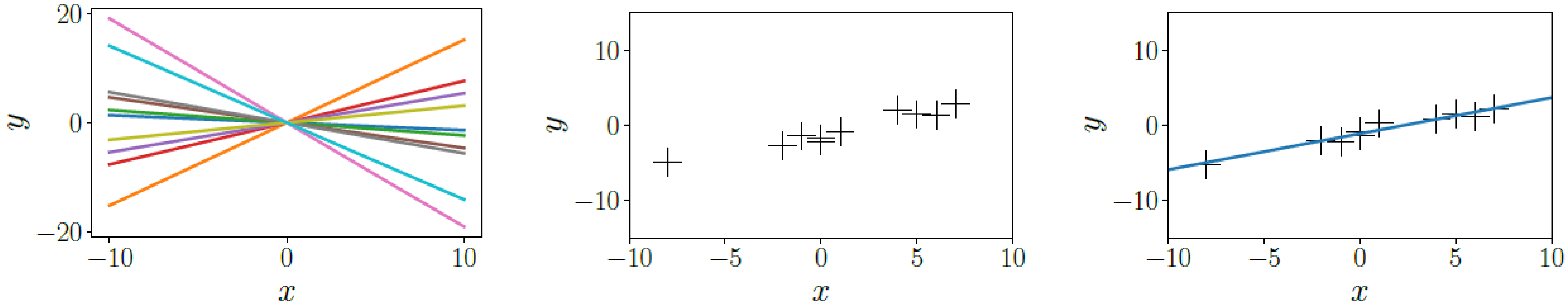


Set de antrenare: $\{(x^1, y_1), \ldots, (x^n, y_n)\}$, masuram cumulative

$$p(Y|X,\theta) = p(y_1, \ldots, y_n | x^1, \ldots, x^n) = \prod_n p(y_n | x^n, \theta)$$

Dorim un set de parametri $\boldsymbol{\theta^*}$ care asigura valoarea maxima a functiei de verosimilitate!

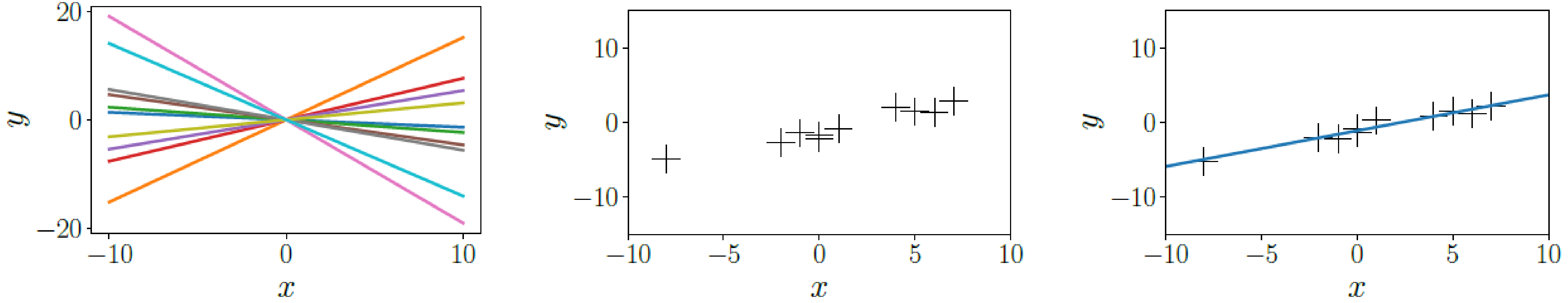
Regresie liniara



Set de antrenare: $\{(x^1, y_1), \ldots, (x^n, y_n)\}$, masuram cumulative

$$\theta_{ML} = argmax_{\theta} \prod_n p(y_n | x^n, \theta) = argmax_{\theta} \log\left(\prod_n p(y_n | x^n, \theta)\right)$$

Regresie liniara

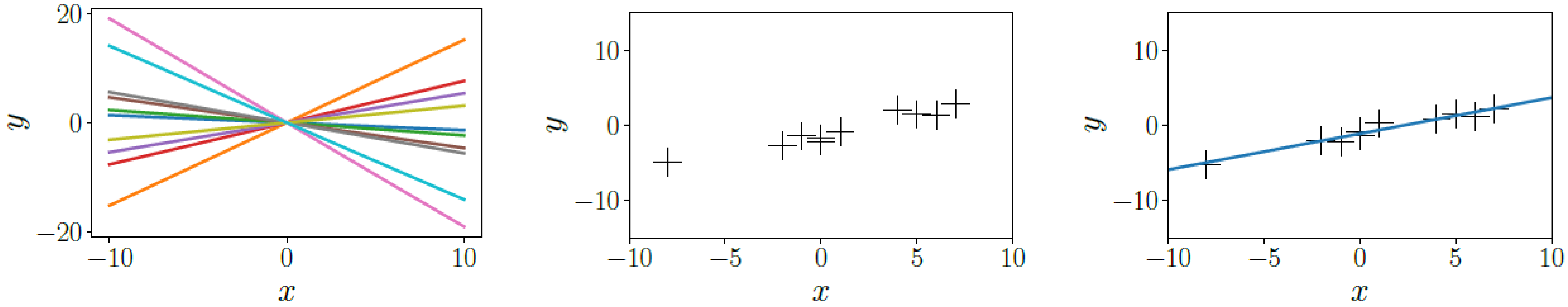


Set de antrenare: $\{(x^1, y_1), \dots, (x^n, y_n)\}$, masuram cumulative

$$\theta_{ML} = argmax_\theta \prod_n p(y_n|x^n, \theta) = argmax_\theta \log\left(\prod_n p(y_n|x^n, \theta)\right)$$

$$= argmax_\theta \sum_i \log p\left(y_i|x^i, \theta\right) = argmax_\theta \sum_i -\frac{1}{2\sigma^2}\left(y_i - \theta^T x^i\right)^2 + const$$

Regresie liniara



Set de antrenare: $\{(x^1, y_1), \dots, (x^n, y_n)\}$, masuram cumulative

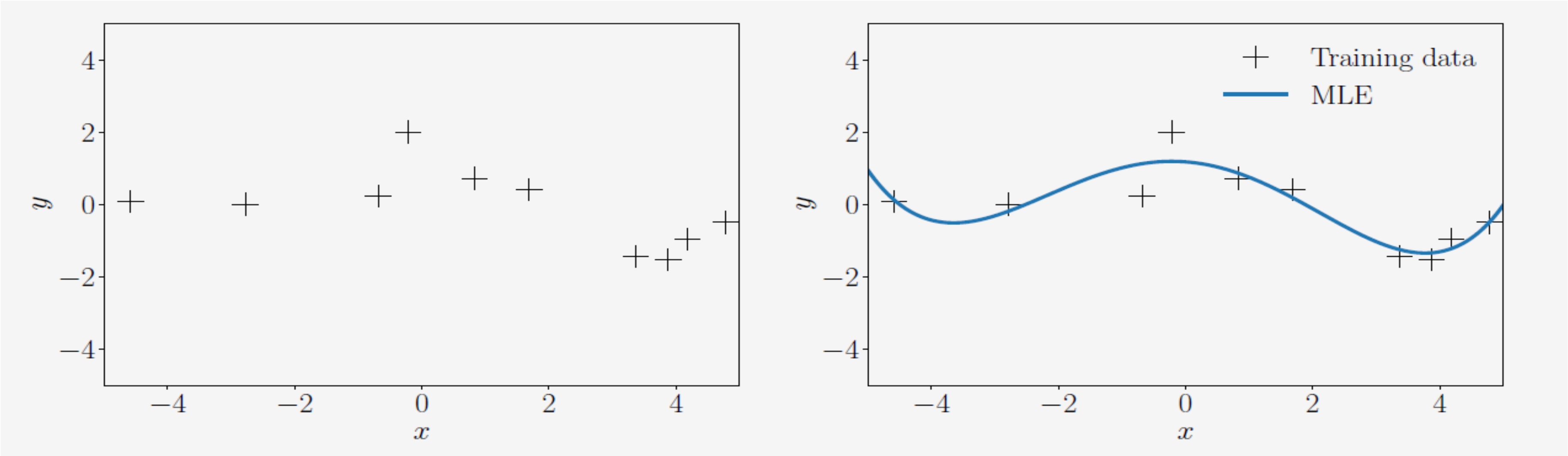
$$\min_{\theta} L(\theta) = \frac{1}{2\sigma^2} \left|\left|X\theta - y\right|\right|_2^2$$

Care este solutia θ_{ML}?

$$\nabla^2 L(\theta) = \frac{1}{\sigma^2} X^T X \succcurlyeq 0 \Rightarrow \nabla \mathrm{L}(\theta^*) = 0 \Leftrightarrow X^T(X\theta^* - y) = 0 \Leftrightarrow \theta_{ML}^* = (X^T X)^{-1} X^T y$$

$$f^*(x) = (\theta_{ML}^*)^T x,\ L^* = \frac{1}{2\sigma^2} \left|\left|X\theta_{ML}^* - y\right|\right|_2^2\ (MSE)$$

Regresie neliniara

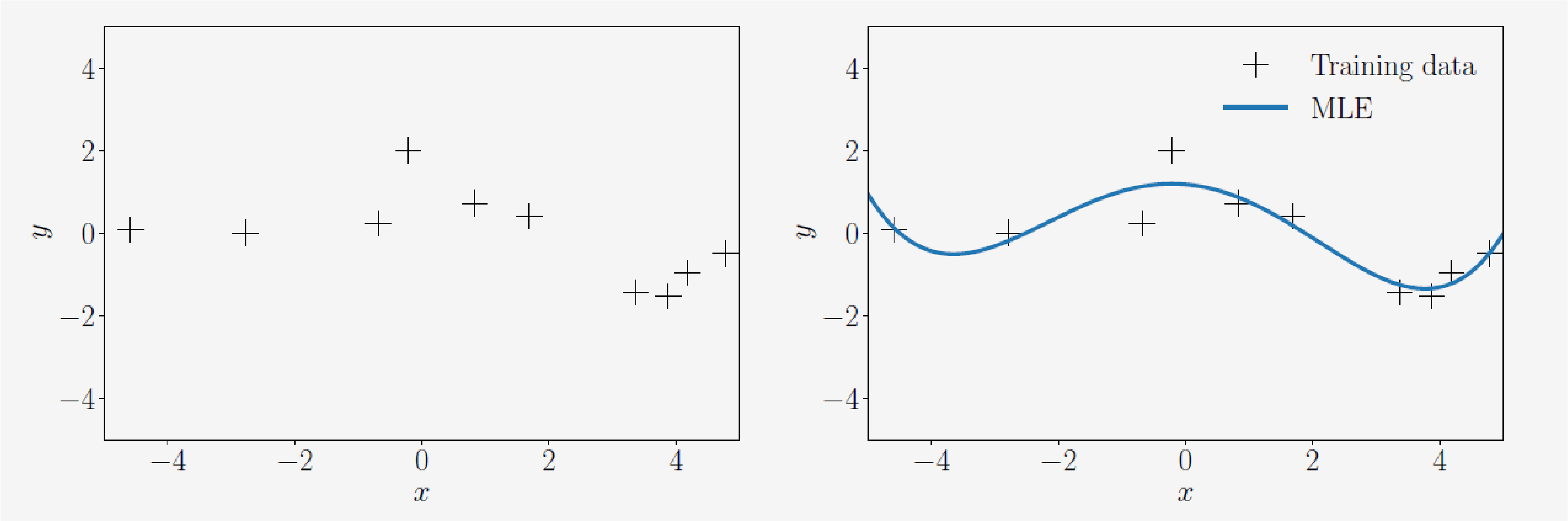


Modelul linear nu este mereu sufficient de expresiv!

Alegem un model nelinear (dar “liniar in parametri”): $\phi: \mathrm{R}^{\mathrm{D}} \rightarrow \mathrm{R}^{\mathrm{N}}, \mathrm{f}(\mathrm{x}) = \sum_i \theta_i \phi(x)_i = \theta^T \phi(x)$

Exemplu: $\phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ \dots \\ x^N \end{bmatrix}, f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_N x^N \Rightarrow \boldsymbol{\phi} = \begin{bmatrix} \phi^T(x^1) \\ \phi^T(x^2) \\ \dots \\ \phi^T(x^n) \end{bmatrix} \in \boldsymbol{R}^{n \times N}$

Regresie neliniara

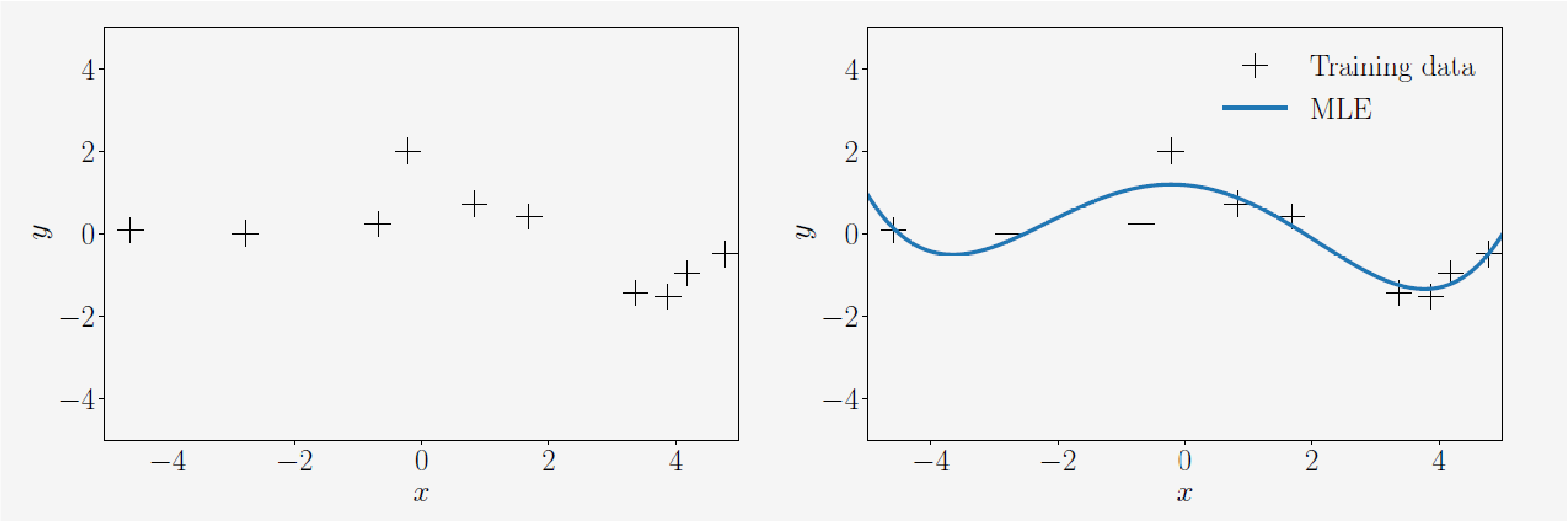


Modelul linear nu este mereu sufficient de expresiv!

Alegem un model nelinear (dar “liniar in parametri”): $\mathrm{f(x)} = \theta^T \phi(x)$

$$\theta_{ML} = argmax_{\theta} - \frac{1}{2\sigma^2} \left|\left|\Phi\theta - y\right|\right|^2 \quad \Rightarrow Solutie\ CMMP!\ \theta^* = (\Phi^T\Phi)^{-1}(\Phi^T y)$$

Regresie neliniara

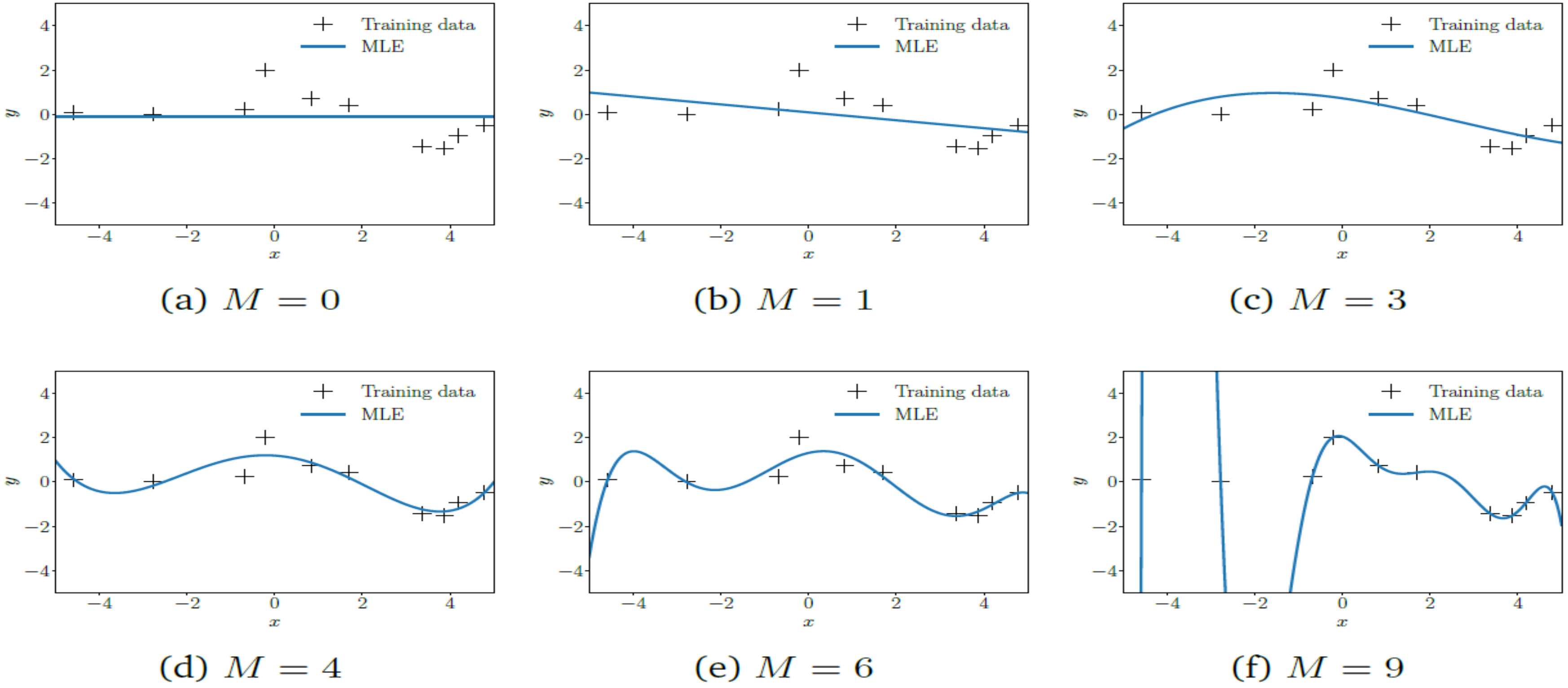


Modelul linear nu este mereu sufficient de expresiv!

Alegem un model nelinear (dar “liniar in parametri”): $f(x) = \theta^T \phi(x)$

$$\theta_{ML} = argmax_{\theta} - \frac{1}{2\sigma^2} ||\Phi\theta - y||^2 \quad \Rightarrow Solutie\ CMMP!$$

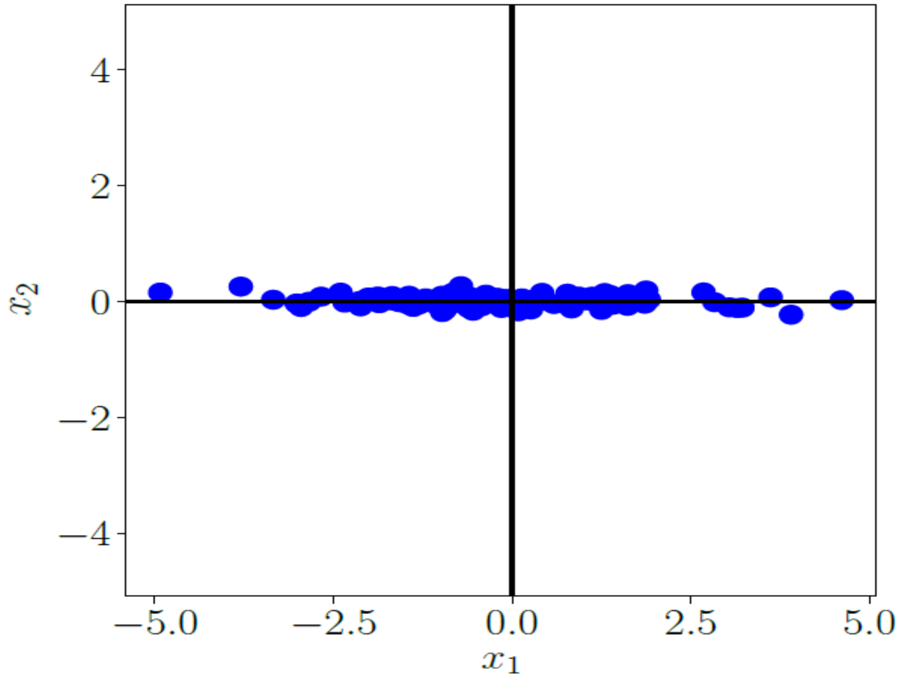
Regresie neliniara

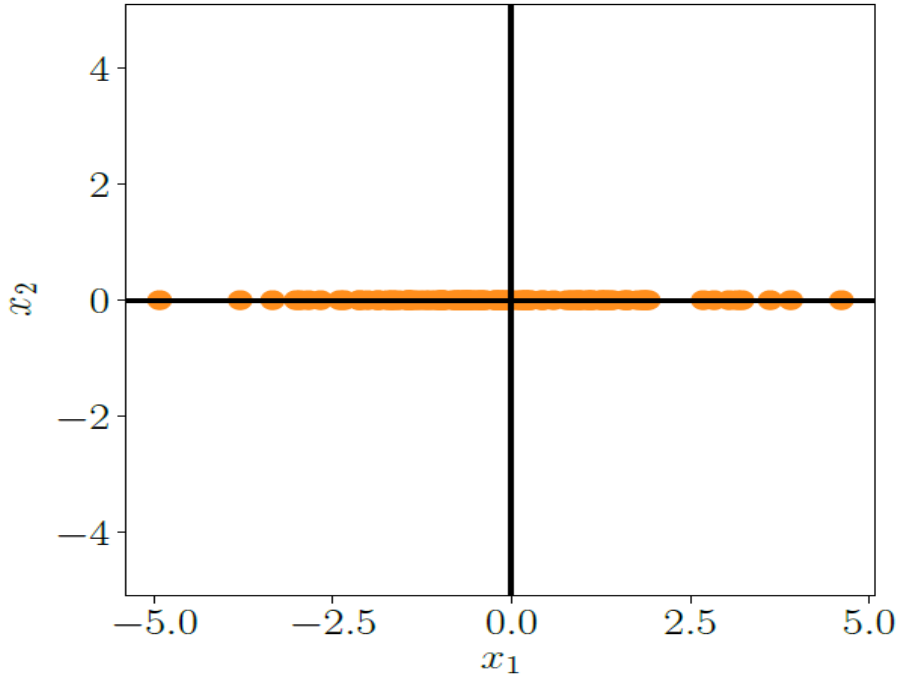


$$f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_N x^N$$

Aplicatii: Analiza componentelor principale (PCA)

Reductie dimensionala





Datele de dimensiuni mari prezinta redundante (dependente) ascunse!

Redundant: $\left\{\begin{bmatrix} x_1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} x_N \\ 0 \end{bmatrix}\right\} \Rightarrow \{x_1, \ldots, x_N\}$

Reductie dimensionala

Datele de dimensiuni mari prezinta redundante (dependente) ascunse!

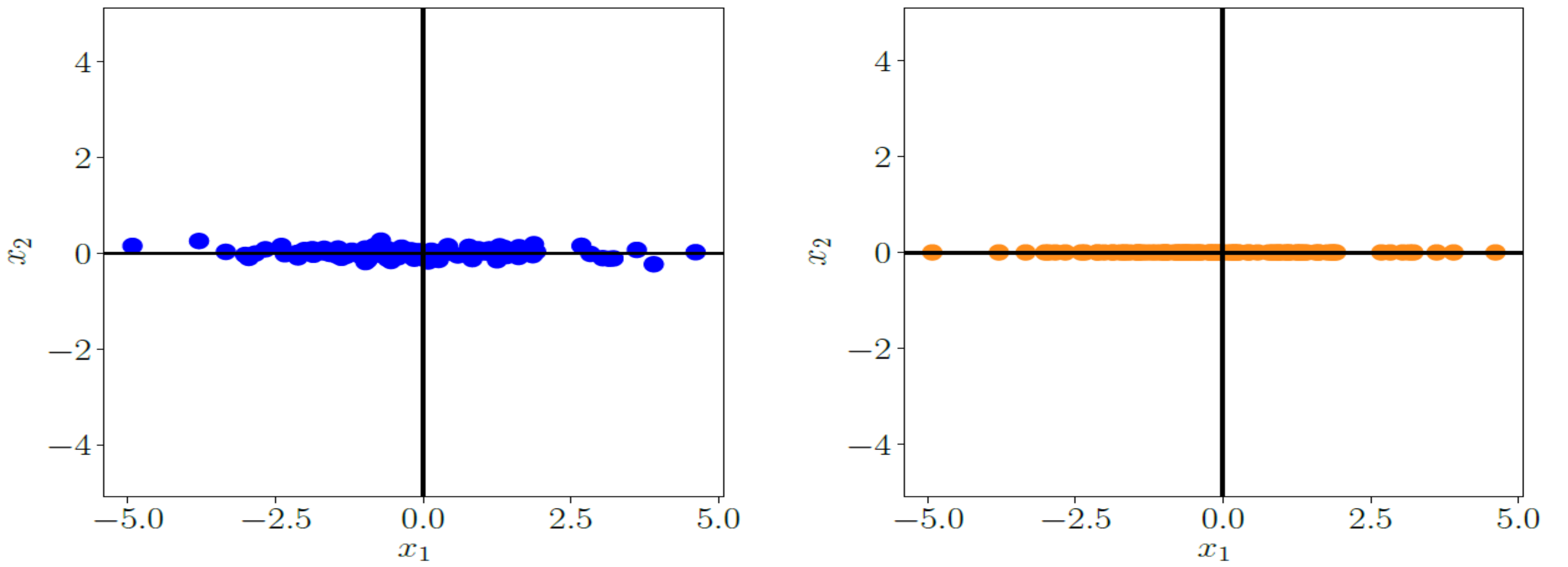
Dependente: $\left\{\begin{bmatrix} x_1 \\ 2x_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ 2x_2 \end{bmatrix}, \dots, \begin{bmatrix} x_N \\ 2x_N \end{bmatrix}\right\}$

$$\left\{\begin{bmatrix} x_1 \\ 2x_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ 2x_2 \end{bmatrix}, \dots, \begin{bmatrix} x_N \\ 2x_N \end{bmatrix}\right\} \Rightarrow \left\{x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix}, x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \dots, x_n \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$$

$$\{x_1 = z_1, x_2, \dots, x_n\} \Rightarrow \left\{\widetilde{x_1} = z_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix}, x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \dots, x_n \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$$

.....

Reductie dimensionala



Reductia dimensionala exploateaza aceste dependente si comprima datele

PCA

Set de date: $\boldsymbol{X} = \{\boldsymbol{x}^{\mathbf{1}}, \dots, \boldsymbol{x}^{\boldsymbol{N}}\} \subset \boldsymbol{R}^{\boldsymbol{D}}$

$$aa^T \succcurlyeq 0$$

Matrice de covarianta:

$$S = \sum_{i=1} x^i (x^i)^T \succcurlyeq 0$$

PCA

Set de date: $\boldsymbol{X} = \{\boldsymbol{x^1}, \dots, \boldsymbol{x^N}\} \subset \boldsymbol{R^D}$

Matrice de convarianta:

$$S = \sum_{i=1} x^i (x^i)^T$$

Presupunem ca exista: $z^i = Bx^i \in R^M$, o codare de dimensiune joasa a datelor (M<<D).

$$B = [b_1 \dots b_M] \in R^{D \times M}, B^T B = I_n, \qquad ||b_i|| = 1, ortonormala$$

Exemplu: baza canonica $e_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, e_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

PCA

Set de date: $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^N\}$

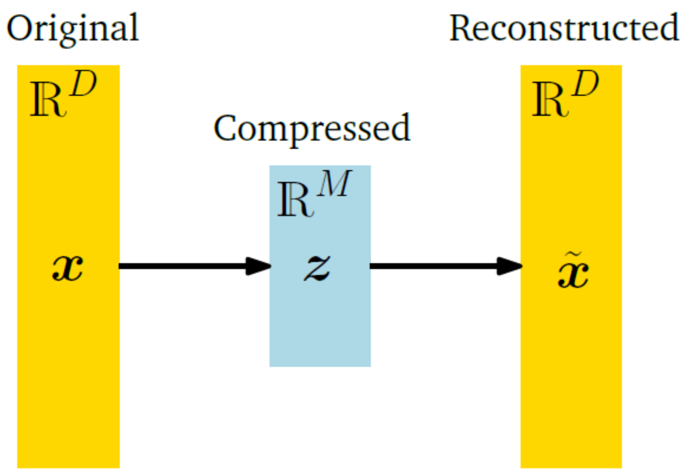
Matrice de convarianta:

$$S = \sum_{i=1} x^i \left(x^i\right)^T$$

Presupunem ca exista: $z^i = Bx^i \in R^M$, o codare de dimensiune joasa a datelor.

$$B = [b_1 \dots b_M] \in R^{D \times M}, \qquad ortonormala$$

Exemplu: baza canonica $e_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, e_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$



z = codul rezultat dupa reductia dimensionala
$\tilde{x}$ = datele reconstruite din z

PCA: Reductie dimensionala cat mai mare,
cu minimizarea pierderilor de informatie (la reconstructie)

PCA

Set de date: $\boldsymbol{X} = \{\boldsymbol{x}^1, \dots, \boldsymbol{x}^N\}$

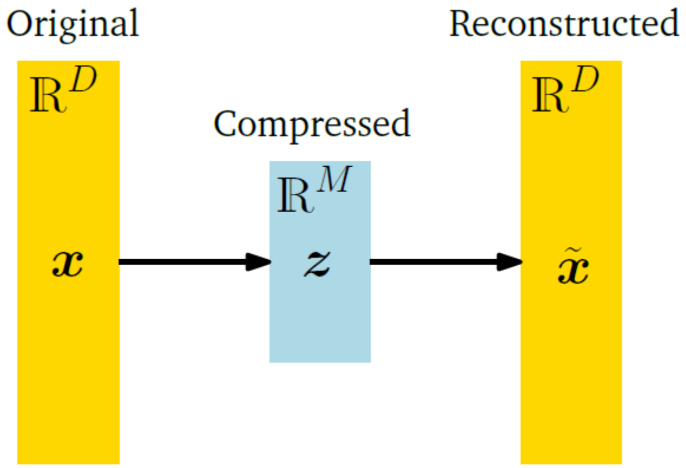
Matrice de convarianta:

$$S = \sum_{i=1} x^i (x^i)^T$$

Presupunem ca exista: $z^i = Bx^i \in R^M$, o codare de dimensiune joasa a datelor.

$$B = [b_1 \dots b_M] \in R^{D \times M}, \qquad ortonormala$$

Exemplu: baza canonica $e_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, e_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$



z = codul rezultat dupa reductia dimensionala

$$\widetilde{x^i} = B^T z^i$$

PCA

Retinere de informatie cat mai mare = Varianta cat mai mare a datelor dupa compresie

Presupunem date cu medie 0:

$$\boldsymbol{V}_z[z] = \boldsymbol{V}_x[B^T(x-\mu)] = \boldsymbol{V}_x[B^T x]$$

Problema: Determinati baza B pentru o varianta maxima.

PCA

Retinere de informatie cat mai mare = Varianta cat mai mare a datelor dupa compresie

Presupunem date cu medie 0:

$$\boldsymbol{V}_z[z] = \boldsymbol{V}_x[B^T(x-\mu)] = \boldsymbol{V}_x[B^T x]$$

Problema: Determinati baza B pentru o varianta maxima.

Procedura in M pasi: primul urmareste maximizarea variantei pe prima componenta z_1^i

$$V_1 = \boldsymbol{V}_z[z_1] = \frac{1}{N}\sum_i (z_1^i)^2$$

PCA

Retinere de informatie cat mai mare = Varianta cat mai mare a datelor dupa compresie

Presupunem date cu medie 0:

$$\boldsymbol{V}_z[z] = \boldsymbol{V}_x[B^T(x-\mu)] = \boldsymbol{V}_x[B^T x]$$

Problema: Determinati baza B pentru o varianta maxima.

Procedura in M pasi: primul urmareste maximizarea variantei pe prima componenta z_1^i

$$V_1 = \boldsymbol{V}_z[z_1] = \frac{1}{N}\sum_i (z_1^i)^2 = \frac{1}{N}\sum_i (b_1^T x^i)^2$$

PCA

Retinere de informatie cat mai mare = Varianta cat mai mare a datelor dupa compresie

Presupunem date cu medie 0:

$$\boldsymbol{V}_z[z] = \boldsymbol{V}_x[B^T(x-\mu)] = \boldsymbol{V}_x[B^T x]$$

Problema: Determinati baza B pentru o varianta maxima.

Procedura in M pasi: primul urmareste maximizarea variantei pe prima componenta z_1^i

$$V_1 = \boldsymbol{V}_z[z_1] = \frac{1}{N}\sum_i (z_1^i)^2 = \frac{1}{N}\sum_i (b_1^T x^i)^2 = \frac{1}{N}\sum_i b_1^T x^i \left(x^i\right)^T b_1$$

$$(a^T b)^2 = a^T b a^T b = a^T b b^T a$$

PCA

Retinere de informatie cat mai mare = Varianta cat mai mare a datelor dupa compresie

Presupunem date cu medie 0:

$$\boldsymbol{V}_z[z] = \boldsymbol{V}_x[B^T(x-\mu)] = \boldsymbol{V}_x[B^T x]$$

Problema: Determinati baza B pentru o varianta maxima.

Procedura in M pasi: primul urmareste maximizarea variantei pe prima componenta z_1^i

$$\max_{b_1} \frac{1}{N} b_1^T \sum_i x^i \left(x^i\right)^T b_1 = b_1^T S b_1 \text{ (convexa)}$$

$$\text{s.l. } \left|\left|b_1\right|\right| = 1$$

PCA

$$\max_{b_1} V_1 = b_1^T S b_1$$
$$\text{s.l. } ||b_1|| = 1$$

Problema convexa?

PCA

$$\max_{b_1} V_1 = b_1^T S b_1$$

$$\text{s.l. } \left||b_1|\right|^2 = 1$$

$$L(b_1, \lambda) = b_1^T S b_1 + \lambda(\left||b_1|\right|^2 - 1)$$

Sistem Kuhn-Tucker?

PCA

$$\max_{b_1} V_1 = b_1^T S b_1$$
$$\text{s.l. } ||b_1||^2 = 1$$

$$L(b_1, \lambda) = b_1^T S b_1 + \lambda(||b_1||^2 - 1)$$

Sistem Kuhn-Tucker:

$$\nabla_{\mathrm{b}} \mathrm{L}(\mathrm{b}_1, \lambda) = 0$$
$$b_1^T b_1 = 1$$

PCA

$$\max_{b_1} V_1 = b_1^T S b_1$$
$$\text{s.l. } \left|\left|b_1\right|\right|^2 = 1$$

$$L(b_1, \lambda) = b_1^T S b_1 + \lambda\left(1 - \left|\left|b_1\right|\right|^2\right)$$

Sistem Kuhn-Tucker:

$$Sb_1 = \lambda b_1$$
$$b_1^T b_1 = 1$$

PCA

$$\max_{b_1} V_1 = b_1^T S b_1$$
$$\text{s.l. } \left||b_1|\right|^2 = 1$$

$$L(b_1, \lambda) = b_1^T S b_1 + \lambda(1 - \left||b_1|\right|^2)$$

Sistem Kuhn-Tucker:

$$Sb_1 + \lambda b_1 = 0 \Rightarrow \mathrm{Sb}_1 = \lambda b_1 \Rightarrow b_1^T S b_1 = \lambda_{max}$$
$$b_1^T b_1 = 1$$

PCA

$$\max_{b_1} V_1 = b_1^T S b_1$$
$$\text{s.l. } \left|\left|b_1\right|\right|^2 = 1$$

$$L(b_1, \lambda) = b_1^T S b_1 + \lambda(1 - \left|\left|b_1\right|\right|^2)$$

Sistem Kuhn-Tucker:
$\mathrm{Sb}_1 = \lambda b_1 \Rightarrow b_1^*$ vector propriu, λ^* valoarea proprie asociata
Pentru ca provide dintr-o problema de maximizare, (b_1^*, λ^*) este v.p. maximal.

PCA

$$\max_{b_1} V_1 = b_1^T S b_1$$

$$\text{s.l. } \left|\left|b_1\right|\right|^2 = 1$$

$$L(b_1, \lambda) = b_1^T S b_1 + \lambda(1 - \left|\left|b_1\right|\right|^2)$$

Sistem Kuhn-Tucker:
$\mathrm{Sb}_1 = \lambda b_1 \Rightarrow b_1^*$ vector propriu, λ^* valoarea proprie asociata

Solutia b_1^* se numeste **prima componenta principala**.

PCA

Solutia b_1^* se numeste **prima componenta principala**.

Daca ne limitam doar un singur pas de PCA, atunci

$$\tilde{x}^i = b_1^* z_1^i = b_1^* b_1^{*T} x^i$$

Reconstructia datelor bazata pe prima componenta principala!

PCA

La pasul k, extragem efectul primilor $k-1$ pasi prin:

$$\hat{X} = X - \sum_{i=1}^{k-1} b_i b_i^T X = X - B^{k-1} X,$$

Reprezinta proiectia datelor X pe subspatiul generat de $(b_1, \ldots, b_{k-1})$

Noua covarianta: $\hat{S} = \frac{1}{N} \hat{X} \hat{X}^T$

PCA

La pasul k, extragem efectul primilor $k-1$ pasi prin:

$$\hat{X} = X - \sum_{i=1}^{k-1} b_i b_i^T X = X - B^{k-1} X,$$

Reprezinta proiectia datelor X pe subspatiul generat de $(b_1, .., b_{k-1})$

Noua covarianta: $\hat{S} = \frac{1}{N} \hat{X} \hat{X}^T$ (**se arata usor ca are acelasi spectru ca $\boldsymbol{S}$**)

Problema de la pasul k:

$$\max_{||b_k||=1} b_k^T \hat{S} b_k$$

PCA

La pasul k, extragem efectul primilor $k-1$ pasi prin:

$$\hat{X} = X - \sum_{i=1}^{k-1} b_i b_i^T X = X - B^{k-1} X,$$

Reprezinta proiectia datelor X pe subspatiul generat de $(b_1, .., b_{k-1})$

Noua covarianta: $\hat{S} = \frac{1}{N} \hat{X} \hat{X}^T$ (**se arata usor ca are acelasi spectru ca $\boldsymbol{S}$**)

Problema de la pasul k:

$$\max_{||b_k||=1} b_k^T \hat{S} b_k = \max_{||b_k||=1, b_k \perp B^{k-1}} b_k^T S b_k$$

Cu solutia: b_k al $k - lea$ vector propriu al lui S.

PCA

Eroarea de reconstructie: puncte reconstruite $\tilde{x} = BB^T x$

$$E(B) = \frac{1}{N} \sum_i ||x^i - \tilde{x}||\text{^}2$$

Cum variaza cu B?

PCA

Proprietatile subsetului de imagini (din MNIST) cu cifra “8”.