Tehnici de Optimizare

Facultatea de Matematica si Informatica

Universitatea Bucuresti

- Department Informatica-

2021

Tehnici de Optimizare

- Curs + seminar/laborator;
- Vom folosi platforma Teams pentru toate orele
- Modul de evaluare va fi prezentat in **cursul viitor**
- Materiale pe Moodle si Teams

Optimizare

- Regasim actul de “optimizare” in multe dintre domeniile fundamentale si chiar actiunile umane (constiente sau inconstiente)

- Scurta lista de aplicatii:
  - Minimizarea riscului (e.g. invatarea automata)
  - Prelucrarea imaginilor (e.g. eliminare zgomot, tomografie)
  - Alocare de resurse
  - Detectie de anomalii

Predictie (Regresie)

Andrew doreste sa aprecieze (prin rating) un hypermarket (e.g. Lidl) inspirandu-se din experienta colegilor sai.

	Market	Alex	Paul	Andrew	
x^1	Cora	1.5	2	2	(y_1)
x^2	Carrefour	3	1	1.5	(y_2)
x^t	Lidl	4	2		?

Bazat pe rating-urile precedente x^1, x^2 ale lui Paul si Alex si pe propriile aprecieri ale acestora, se doreste predictia rating-ului lui Andrew pentru magazinul Lidl in acest context.

Predictie (Regresie)

Aproximam ipoteza h cu o **functie liniara**:

$$h(x) \coloneqq h_\theta(x) = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

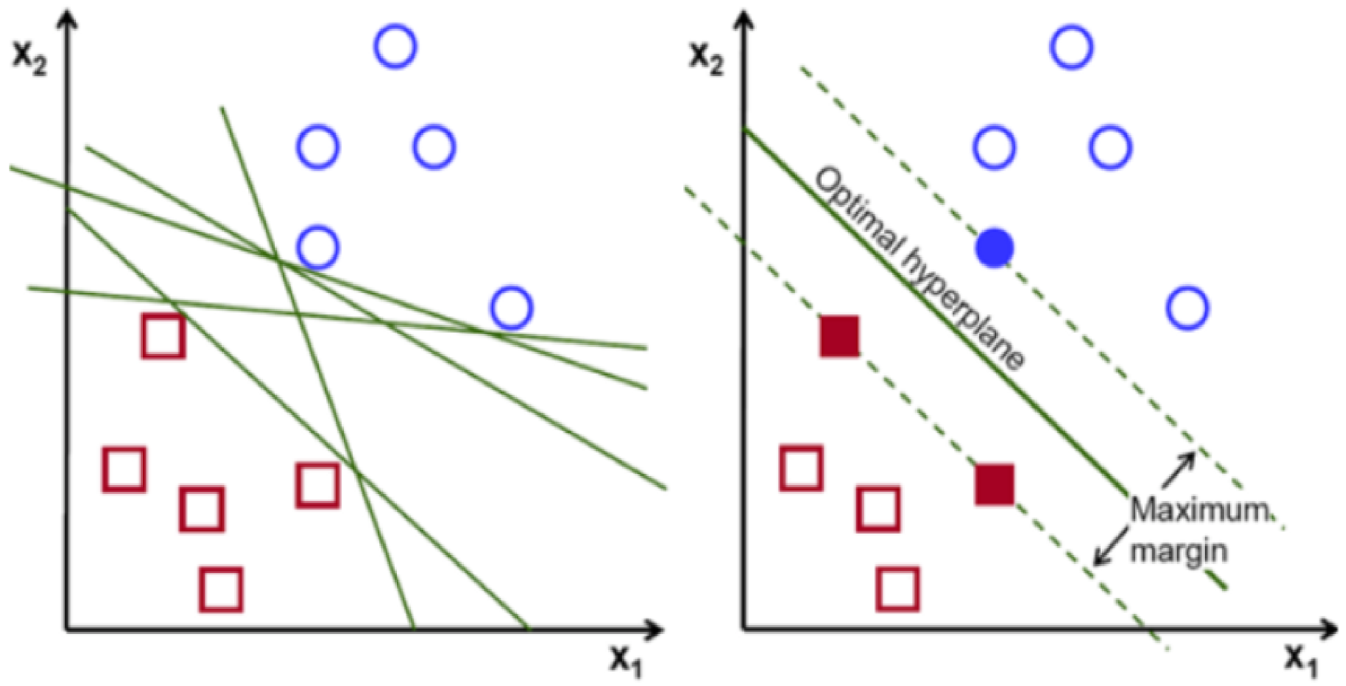
unde θ reprezinta vectorul parametrilor.

Modelele de regresie liniara se reduc la gasirea parametrilor optimali care minimizeaza eroarea dintre valoarea functiei-decizie $h(x)$ si valorile de iesire reale y_i:

$\min\limits_{\theta} \frac{1}{2m} \sum \left(h_\theta\left(x^i\right) - y_i\right)^2 = \min\limits_{\theta} \ \frac{1}{2m} \|X\theta - y\|_2^2$ (||Ax-b||^2)

- Ce putem spune despre vectorul optim de parametri? $X^T(X\theta - y) = 0 \Rightarrow \theta^* = (X^T X)^{-1} X^T y$
- Accesibil pentru dimensiuni mari?

Support Vector Machine



ps://towardsdatascience.com/support-vector-machine-vs-logistic-regression-94cc2975433f

Hiperplan:

- In 2D este o dreapta
- In 3D este un plan
- In nD, o multime de forma: $H = \{x \in R^n : w^T x = b\}$

Separare:

- Daca $w^T x < b$ atunci *x se afla in clasa 1*
- Daca $w^T x > b$ atunci *x se afla in clasa 2*

Problema: Se da setul de date compus din exemple ale celor doua clase. Determinati hiperplanul care da ***cea mai ferma separare*** intre doua clase.

Support Vector Machine

$$\min_{w,b} \quad ||w||^2$$
$$s.t. \quad y_i(w^T x_i - b) \geq 1 \qquad i = 1, \dots, m$$

- y_i reprezinta etichetele exemplului x_i; *w* si *b* reprezinta variabilele de cautare
- Pentru m mare, avem un numar mare de constrangeri
- Functie de minimizat se numeste **cost** sau **functie obiectiv**
- Relatiile la care se supun variabilele de cautare se numesc **constrangeri** (sau restrictii) si formeaza **multimea fezabila**

Detectie anomalii

Intr-un set dat de tranzactii financiare, se realizeaza detectia transferurilor anormale de bani intr-un grup limitat de client.

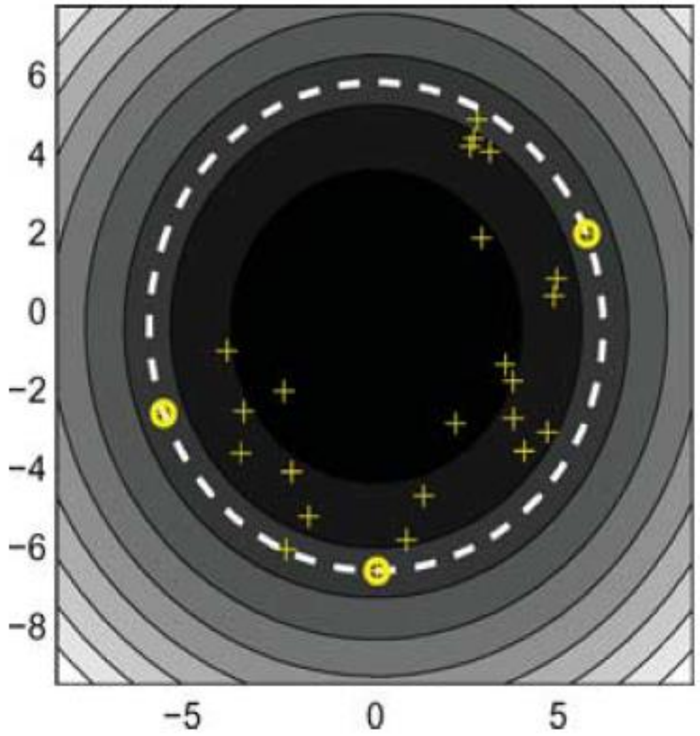
Source Client	Destination Client	Amount	Date
A	B	200000	01.01.2020
Z	F	1050	02.01.2020
B	A	200000	07.01.2020
M	F	50	02.01.2020
A	B	700000	01.01.2020
N	W	70	01.01.2020
B	A	700000	07.01.2020

In contextul transactional dat, transferurile dintre clientii A si B devin **suspecte de frauda** datorita circuitului inainte-inapoi si, de asemenea, sumelor mari.

Problema: Fie setul de date $\{x^i\}$, care contine colectia de tranzactii, sa se evidentieze (ne)supervizat tranzactiile suspecte.

SVDD (Support Vector Data Description)

Tehnica pentru clasificare cu o singura clasa (one-class classification):



$$\min_{c,R} R^2 + sum(z)$$
$$s.t.\ \left\|x^i - c\right\| \leq R + z_i \quad i = 1, \ldots 7$$

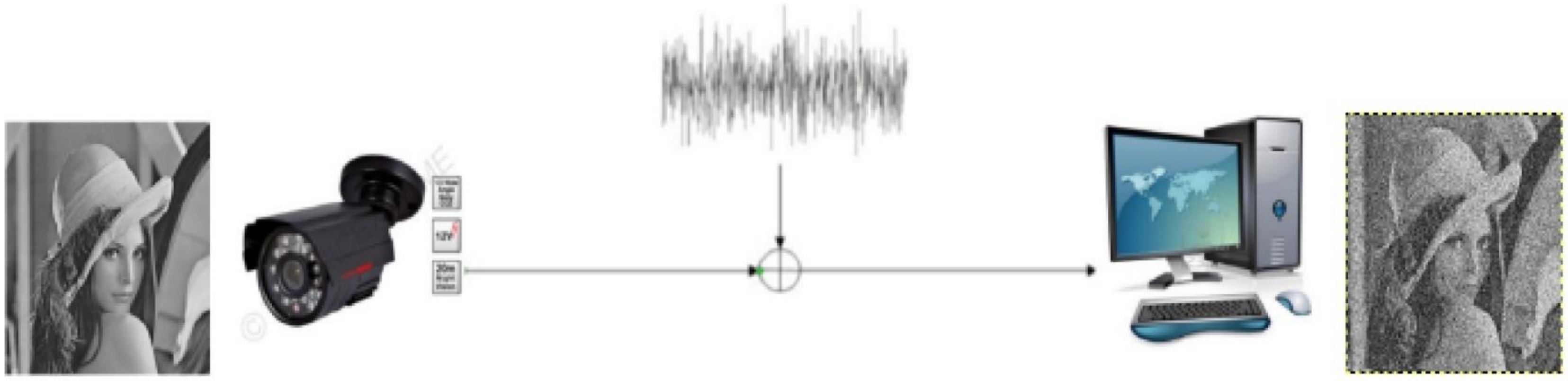
Normal: ||x^j – c||<R
Anormal: ||x^j – c||>R

Problema determina centrul si raza unei (hiper)sfere cu frontiera in jurul **majoritatii** datelor.

Exemplele ramase in afara sferei (outliers) se considera **anomalii**.

Filtrare de semnal (Signal denoising)

Procesele ingineresti implica achizitia/prelucrarea si comunicatia semnalelor.



La achizitie/comunicatie apar modificari nedorite ale semnalelor (zgomot).

Problema: Modelati problema si “filtrati” semnalul pentru a elimina zgomotul (ne)gaussian.

Filtrare de semnal (Signal denoising)

$$\min_{Y \in \mathbb{R}^{m \times n}} \frac{1}{2}\|Y - X\|^2 + TV(Y)$$

- TV = "total variation" functie cuantificare variatiei totala
- $TV(y) = \sum_{i,j} |y_{i+1,j} - y_{i,j}| + |y_{i,j+1} - y_{i,j}|$;

Rezultate (noisy image/denoised image):




Optimizare matematica

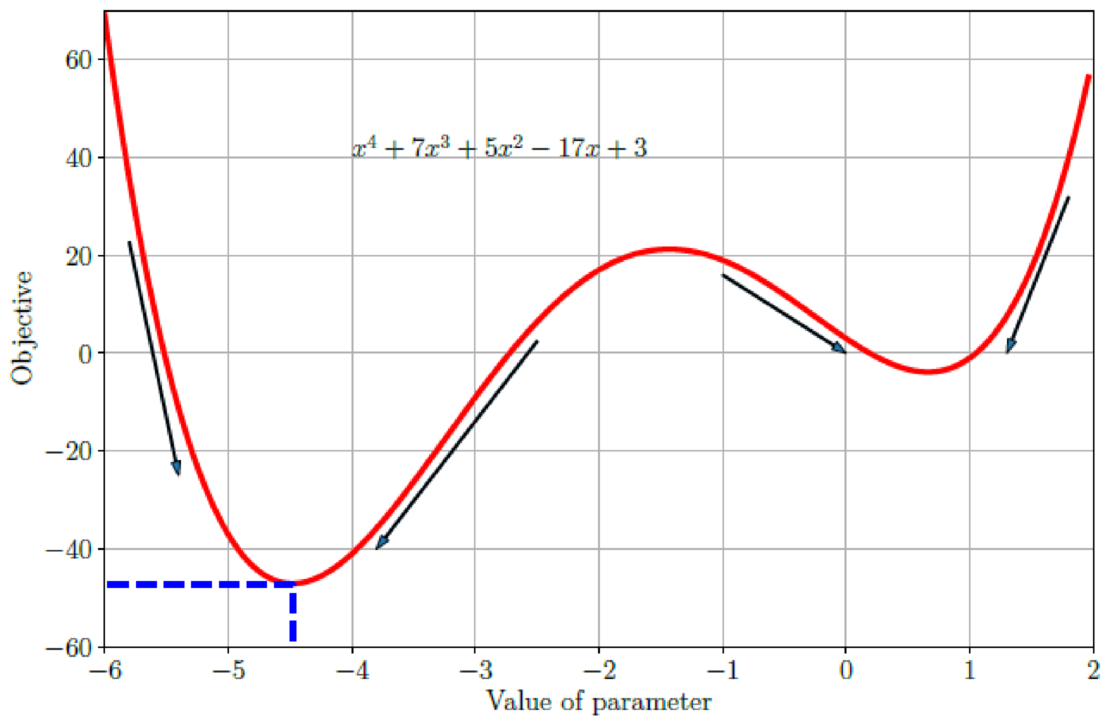
$$\min_{x} \quad f(x)$$
$$s.t. \; g_1(x) \leq 0, \dots, g_m(x) \leq 0,$$
$$h_1(x) = 0, \dots, h_p(x) = 0$$

- Functie de minimizat f se numeste **cost** sau **functie obiectiv**
- Functiile g_i, h_i definesc constrangerile (restrictiile). Q reprezinta multimea punctelor ce satisfac constrangerile (i.e. **multimea fezabila)**

- **Domeniul optimizarii isi propune analiza si rezolvarea de modele de optimizare cu ajutorul algoritmilor iterativi**
- **Evident, este necesara o intelegere a notiunilor de baza folosite pentru analiza; deci reamintim diferentiabilitate, gradient, Hessiana**

- **Rezolvarea modelelor generice este extrem de dificila!**

Optimizare matematica



- Punctul $x^* \in \mathbb{R}^n$ este un *punct de minim global* (adesea denumit *minim global*) daca si numai daca $x^* \in X$ si $f(x^*) \leq f(x)$ oricare ar fi $x \in X$.
- Punctul $x^* \in \mathbb{R}^n$ este *minim local* daca si numai daca $x^* \in X$. si exista o vecinatate $\mathcal{N}$ a lui x^* (e.g. o bila deschisa cu centrul în x^*) astfel încât $f(x^*) \leq f(x)$ oricare ar fi $x \in X \cap \mathcal{N}$.

Optimizare matematica

$O\ functie\ f: R^n \rightarrow R\ de\ n\ variabile\ este\ diferentiabila\ in\ punctul\ x\ daca\ exista\ a\ \in R^n\ astfel\ incat, pentru\ orice\ d\ \in R^n$:

$$f(x+d) = f(x) + \ a^T d + o(d),$$

$o(r) este\ o\ functie\ de\ r \geq 0\ astfelincat \lim_{r \to 0} \frac{o(r)}{r} = 0\ si\ o(0) = 0.$

$$Gradientul\ functiei\ f\ in\ x:\ \nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

$$Matricea\ Hessiana\ functiei\ f\ in\ x:\ \nabla^2 f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_n \partial x_n} \end{bmatrix}$$

Exemple

Functia liniara: $f(x) = a^T x + b \; = \; \sum a_i x_i + b$;

$Exemplu$: $f(x_1, x_2) = x_1 + x_2, deci\ in\ acest\ caz\ \nabla f(x) = [1\ 1]^T$;

In general, $\nabla f(x) = a, \nabla^2\ f(x)$ = 0.

Functia patratica: $f(x) = \frac{1}{2} x^T H x + a^T x + b$

Exemplu: $f(x_1, x_2) = 0.5(x_1^2 + x_2^2) = \frac{1}{2} [x_1\ x_2] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$;

In general, $\nabla f(x) = Hx + a, \nabla^2\ f(x)$ = H

Alte notatii

In cadrul cursului vom utiliza urmatoarele notatii:

- Vectori (considerati intotdeauna vector coloana) cu litere mici, i.e. $x \in \mathbb{R}^n$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$
- Produs scalar in spatiul Euclidian: $\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i$
- Norma Euclidiana standard $\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \cdots + x_n^2}$
- Multimi cu litere mari: $S, Q, U \subseteq \mathbb{R}^n$ ($\mathbb{R}^n_+$ - orthantul nenegativ, S^n_+ - multimea matricelor pozitiv semidefinite)
- Matrice cu litere mari: $A, B, C, H \in \mathbb{R}^{n \times m}$
- Norma spectrala a unei matrici $\|A\| = \sqrt{\lambda_{max}(A^T A)}$
- Matrice pozitiv definita: $A \succ 0$, si pozitiv semidefinita $A \succeq 0$