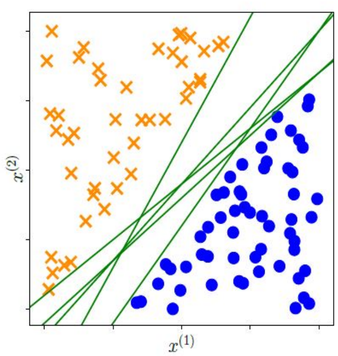
Tehnici de Optimizare
Laborator 4
– Probleme de optimizare cu restricţii –

1 Introducere

Problema de clasificare SVM: Fie seturile de "obiecte" $X^+ = \{x_1^+, x_2^+, \cdots, x_m^+\}, X^- = \{x_1^-, x_2^-, \cdots, x_m^-\}$. Determinaţi hiperplanul $H(w, b)$ de *margine maximă* care separă (distinge) cele două seturi.

$$w^T x_i^+ + b > 0 \quad i = 1, \cdots, m$$
$$w^T x_i^- + b < 0 \quad i = 1, \cdots, m.$$



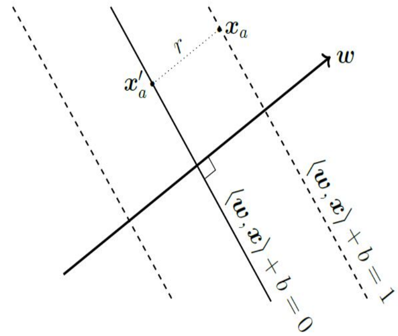
În cazul separabil, printr-o simplă schimbare de variabilă, problema se reduce la găsirea perechii $(\hat{w}, \hat{b})$ astfel încât:

$$y_i(\hat{w}^T x_i + \hat{b}) \geq 1 \quad \forall x_i \in X^+ \cup X^-,$$

unde $y_i = +1$ dacă $x_i \in X^+$ sau $y_i = -1$ dacă $x_i \in X^-$.
Observaţie: *Pentru simplitate, vom folosi mai departe, în loc de* $(\hat{w}, \hat{b})$, *notaţiile originale* (w, b).

Distingem hiperplanele auxiliare de separare $H(w, 1 - b)$ şi $H(w, -1 - b)$ (figurate cu linie punctată):
Pentru forma analitică a marginii (distanţa pâna la $H(w, \pm 1 - b)$) observăm:

$$x_a = x'_a + r\frac{w}{\|w\|} \Rightarrow w^T x_a + b = 1$$
$$\Rightarrow w^T(x'_a + r\frac{w}{\|w\|}) + b = 1$$
$$\Rightarrow r = 1/\|w\|.$$



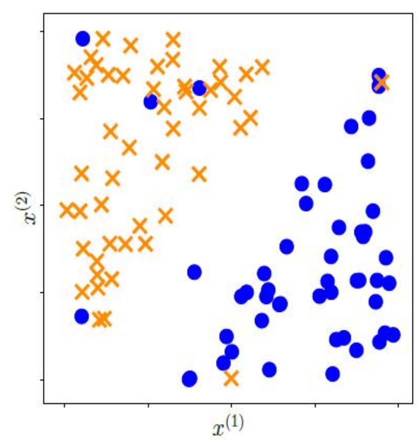
De aceea, asigurarea marginii maxime de separare se atinge prin rezolvarea problemei:

$$\max_{w,b} \frac{1}{\|w\|}$$
$$\text{s.l. } y_i(w^T x_i + b) \geq 1 \qquad \forall i = 1, \cdots, m$$

Maximizarea $\frac{1}{\|w\|}$ este echivalentă cu minimizarea $\|w\|$ sau $\|w\|^2$, de aceea rescriem:

$$\texttt{(SVM separabil):} \quad \min_{w,b} \|w\|^2$$
$$\text{s.l. } y_i(w^T x_i + b) \geq 1 \qquad \forall i = 1, \cdots, m.$$

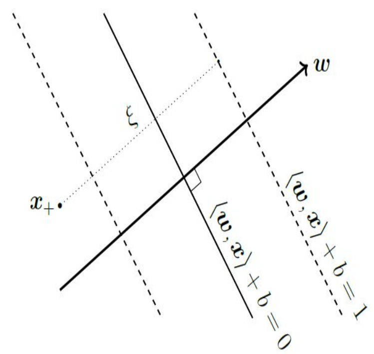
În cele mai realiste situaţii (e.g. clasificarea de documente) ipoteza de separabilitate nu este verosimilă, de aceea, se urmăreşte clasificarea cu "eroare" minimă.



Modelăm eroarea printr-un set de variabile virtuale $\xi_i = 1 - y_i(w^T x_i - b)$ nenegative. Natural, ne dorim la optim ca valoarea variabilelor de eroare sa fie cât mai mici posibil.

În problema de `SVM separabil` identificăm $\xi_i = 1 - y_i(w^T x_i - b) \geq 0$ şi penalizăm în cost valoarea cumulată a variabilelor de eroare $\sum_i \xi_i$:

$$\texttt{(SVM neseparabil):} \quad \min_{w,b,\xi} \|w\|^2 + \rho \sum_i \xi_i$$
$$\text{s.l. } y_i(w^T x_i + b) \geq 1 - \xi_i \qquad \forall i = 1, \cdots, m$$
$$\xi \geq 0.$$



Numărul mare de constrângeri face problema `SVM neseparabil` greu de abordat în forma primală. Deoarece intervin doar constrângeri liniare de inegalitate, utilizăm argumentul dualităţii pentru a simplifica problema.

$$\mathcal{L}(w, b, \xi, \lambda, \gamma) = \frac{1}{2}\|w\|^2 + \rho \sum_i \xi_i - \sum_{i=1}^m \lambda_i [y_i(w^T x_i + b) - 1 + \xi_i] - - \sum_{i=1}^m \gamma_i \xi_i.$$

Condiţiile Kuhn-Tucker de optimalitate primal-duală:

$$\nabla_w \mathcal{L}(w, b, \xi, \lambda, \gamma) = w - \sum_{i=1}^m \lambda_i y_i x_i = 0$$
$$\nabla_b \mathcal{L}(w, b, \xi, \lambda, \gamma) = -\sum_{i=1}^m \lambda_i y_i = 0$$
$$\nabla_{\xi_i} \mathcal{L}(w, b, \xi, \lambda, \gamma) = \rho - \lambda_i - \gamma_i.$$

Din prima relaţie observăm la optim:

$$w^* = \sum_{i=1}^m \lambda_i^* y_i x_i. \tag{1}$$

Datele x_i pentru care $\lambda_i^* = 0$ nu contribuie la soluţia problemei `SVM neseparabil`, iar cele pentru care $\lambda_i^* > 0$ se numesc **vectori suport** şi satisfac $w^T x^i + b = y_i$ (se poziţionează pe hiperplanele auxiliare). După eliminarea variabilelor primale, obţinem problema duală pentru `SVM neseparabil`:

$$\texttt{(SVM dual):} \quad \max_{\lambda} \; -\frac{1}{2}\lambda^T Q^T Q \lambda + e^T \lambda$$
$$\text{s.l. } y^T \lambda = 0, \; 0 \le \lambda_i \le \rho.$$

unde $Q = \begin{bmatrix} y_1 x^1 & y_2 x^2 & \cdots & y_m x^m \end{bmatrix}$. După rezolvarea problemei `SVM dual`, soluţie primală se determină pe baza (1) şi, în plus, $b = y_i - w^T x^i$ pentru orice index i corespunzător vectorilor suport.

2 Probleme propuse

1. Generaţi m date aleatoare separabile în $\mathbb{R}^n$. Determinaţi hiperplanul de margine maximă rezolvând `SVM separabil`, folosind CVXPY. Testaţi optimalitatea soluţiei.

2. Generaţi m date aleatoare (ne)separabile în $\mathbb{R}^n$.

a. Reformulaţi problema `SVM dual` pentru a elimina constrângerea de egalitate, adăugând în obiectiv funcţia de penalitate corespunzătoare.

b. Implementaţi Metoda Gradientului Dual (MGD) pentru rezolvarea problemei duale modificate de la punctul a.

c. Calculaţi soluţia primală (w^*, b^*) pornind de la multiplicatorii Lagrange furnizaţi de MGD de la punctul b..

2.1 Ghid Python

Importare pachete necesare:

```
import numpy as np
from numpy import linalg
from numpy.linalg import norm
import numpy.random as random
import matplotlib.pyplot as plt
```

Exemplu generare vector aleator:

```
s = random.rand(n)
```

Exemplu generare intreg aleator între 0 şi m:

```
i = random.randint(m, size = 1)
```