

# Istraživanje podataka

## Vežbe 10

23. april 2021

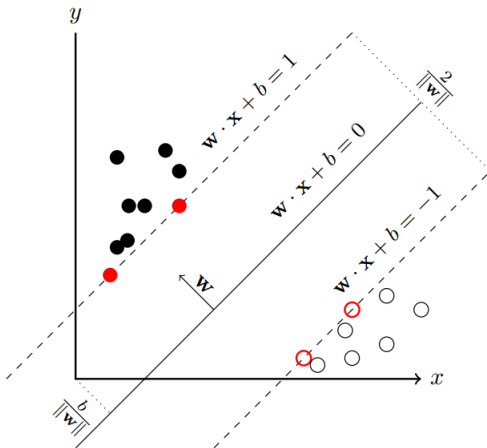
# Outline

- 1 SVM - Metod podržavajućih (potpornih) vektora
- 2 Zadatak za samostalan rad

# Outline

- 1 SVM - Metod podržavajućih (potpornih) vektora
- 2 Zadatak za samostalan rad

# Metod podržavajućih (potpornih) vektora - linearno separabilan skup



# Metod podržavajućih (potpornih) vektora - linearno separabilan skup

- Pretpostavka: podaci su linearno razdvojivi
- Klase: 1 i -1
- Cilj: naći optimalnu hiper-ravan, tj. hiper-ravan sa maksimalnom marginom koja razdvaja instance klase 1 i instance klase -1

$$w * x + b = 0$$

- Instance na margini - podržavajući (potporni) vektori

# Metod podržavajućih (potpornih) vektora - linearno separabilan skup

- Ograničenja margina

$$w * x + b = 1$$

$$w * x + b = -1$$

- Ograničenja za instance

- za  $y_i = 1$   $w * x_i + b \geq 1$
- za  $y_i = -1$   $w * x_i + b \leq -1$
- ili  $\forall i$  iz skupa  $y_i * (w * x_i + b) \geq 1$

# Metod podržavajućih (potpornih) vektora - linearno separabilan skup

- Rastojanje od hiper-ravni:  $\frac{|w \cdot x + b|}{\|w\|}$
- Rastojanje od hiper-ravni do podržavajućih vektora:  $\frac{1}{\|w\|}$
- Margina:  $\frac{2}{\|w\|}$
- Zahtev

$$\max_w \frac{2}{\|w\|}$$

uz ograničenja  $y_i * (w * x_i + b) \geq 1, i = 1, 2, \dots, N$

- Ili

$$\min_w \frac{\|w\|^2}{2}$$

uz ograničenja  $y_i * (w * x_i + b) \geq 1, i = 1, 2, \dots, N$

- Klasifikacija instance  $z$ :  $\text{sign}(w * z + b)$

## Metod podržavajućih (potpornih) vektora - linearno separabilan skup

- Optimizacioni problem može da se reši korišćenjem metoda Lagranžovi množiaci
- $L_p = \frac{\|w\|^2}{2} - \sum_{i=1}^N \lambda_i * [y_i * (w * x_i) + b - 1]$
- nenegativni Lagranžovi množiaci  $\lambda_i$  se povezuju sa ograničenjima



# Metod podržavajućih (potpornih) vektora - linearno separabilan skup

- Važi

$$w = \sum_{i=1}^l y_i \alpha_i * (x_i)$$

$$\sum_{i=1}^l y_i \alpha_i = 0$$

- Važi  $\lambda_i = 0$  osim  $y_i * (w * x_i) + b = 1$
- Ako je  $\lambda_i > 0$ ,  $x_i$  je potporni vektor

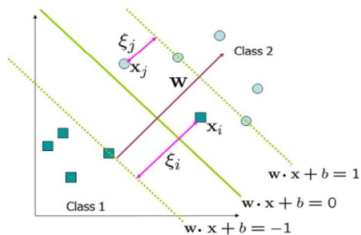
# Metod podržavajućih (potpornih) vektora - linearno separabilan skup

- dualni problem (transformacija Lagranžijana u funkciju od Lagranžijanovih množioca)

$$\max_{\lambda_i} L_D = \sum_{i=1}^N \lambda_i - \frac{1}{2} * \sum_{i=1}^N \sum_{j=1}^N \lambda_i * \lambda_j * y_i * y_j * x_i * x_j$$

- Klasifikacija instance  $z$ :  $\text{sign}(\sum_{i=1}^N \lambda_i * y_i * x_i * z + b)$

# Metod podržavajućih (potpornih) vektora - meka margina



## Metod podržavajućih (potpornih) vektora - meka margina

- Ako instance skupa nisu linearno razdvojive, uvode se promenljive u ograničenja za svaku instancu kako bi se olabavila
- Ograničenja za instance
  - za  $y_i = 1$   $w * x_i + b \geq 1 - \xi_i$
  - za  $y_i = -1$   $w * x_i + b \leq -1 + \xi_i$

# Metod podržavajućih (potpornih) vektora - meka margina

- Ciljna funkcija

$$\min \frac{\|w\|^2}{2} + C * \sum_{i=1}^N \xi_i^k$$

- $C$  i  $k$  (1 ili 2) su konstante
- $\sum_{i=1}^N \xi_i^k$  - gubitak
- dualni problem  $L_D = \sum_{i=1}^N \lambda_i - \frac{1}{2} * \sum_{i,j} \lambda_i * \lambda_j * y_i * y_j * x_i * x_j$

## Nelinearni metod podržavajućih (potpornih) vektora

- Za skupove koji nisu linearno razdvojivi, potrebno je odrediti funkciju  $\Phi$  koja će skup transformisati u prostor sa linearno razdvojivim instancama
- Traži se  $w * \Phi(x) + b = 0$
- Klasifikacija instance  $z$ :  $sign(\sum_{i=1}^N \lambda_i * y_i * \Phi(x_i) * \Phi(z) + b)$
- Kernel trik: računanje  $\Phi(x_i) * \Phi(z)$  u transformisanom prostoru koristeći originalne atribute

# Nelinearni metod podržavajućih (potpornih) vektora

- Zahtev

$$\min \frac{\|w\|^2}{2} + C * \sum_{i=1}^N \xi_i^k$$

- Ograničenja za instance

$$y_i * (w * \Phi(x_i) + b) \geq 1 - \xi_i, i = 1, 2, \dots, N$$

$$\xi_i \geq 0$$

# Metod podržavajućih (potpornih) vektora u biblioteci scikit-learn

- `sklearn.svm.SVC`
- parametri
  - $C$  - kazna za grešku: default 1
  - kernel funkcija
    - *linear*:  $\langle x, x' \rangle$
    - *poly*:  $(\gamma \langle x, x' \rangle + r)^d$ ,  $\gamma$  se zadaje parametrom *gamma*:  $d$  parametrom *degree*, a  $r$  sa *coef0*
    - *rbf*:  $\exp(-\gamma \|x - x'\|^2)$   $\gamma$  se zadaje parametrom *gamma* i mora biti pozitivna
    - *sigmoid*:  $\tanh((\gamma \langle x, x' \rangle + r))$



# Metod podržavajućih (potpornih) vektora u biblioteci scikit-learn

- *degree* - stepen za kernel funkciju *poly*, default=3
- *gamma* - koeficijent za kernel funkcije: *rbf*, *poly*, *sigmoid*, default='auto', tj.  $1/\text{broj\_atributa}$
- *coef0* - nezavisan član u kernel funkcijama: *poly*, *sigmoid*,

# Metod podržavajućih (potpornih) vektora u biblioteci scikit-learn

- atributi
  - *support\_* - indeksi podržavajućih vektora
  - *support\_vectors\_* - podržavajući vektori
  - *n\_support\_* - broj podržavajućih vektora za svaku klasu
  - *dual\_coef\_* - koeficijenti podržavajućih vektora  $y_i\alpha_i$ .  
Ukoliko postoji više klasa, postoje koeficijenti za sve *1-vs-1* klasifikatore.
  - *intercept\_* - konstante u funkciji odlučivanja

# Neuronska mreža sa skrivenim slojevima u biblioteci scikit-learn

- metode
  - $fit(X, y)$  - za treniranje modela
  - $predict(X)$  - za predviđanje klasa

# Outline

- 1 SVM - Metod podržavajućih (potpornih) vektora
- 2 Zadatak za samostalan rad

## Zadatak za samostalan rad

Preuzeti skup podataka *car.csv* o klasama automobila. Koristeći IBM SPSS Modeler i SVM izvršiti klasifikaciju nad datim skupom. Atribut *class* sadrži informaciju kojoj klasi pripada automobil. Primeniti PCA na skup i izvršiti klasifikaciju nad transformisanim skupom. Koji broj atributa ste izabrali i zašto? Diskutovati dobijen model.

## Zadatak za samostalan rad

Skup *car* sadrži podatke o automobilima. Atributi skupa su:

- class - klasa automobila
- cylinders - broj cilindara
- displacement - zapremina motora
- horsepower - konjska snaga
- weight - težina
- acceleration - ubrzanje