

## Tema 2

### SVM

Scopul aceste teme este lucrul cu SVM-uri pe diferite date de test si evaluarea parametrilor in scopul cresterii acuratetei de clasificare. Pentru implementarea SVM-urilor se va folosi bibliotecak LIBSVM [1].

Se considera seturile de date:

1. Skin\_nonskin [2]
2. News20 [3]

Sa se implementeze un SVM care sa clasifice fiecare set de date. Testarea se va realiza utilizand diferite tipuri de kernel [4]:

- a) kernel liniar,
- b) kernel polinomial,
- c) kernel Gaussian
- d) kernel sigmoid

Pentru clasificarea datelor intr-un numar de clase  $> 2$  se vor implementa doua modalitati [5] (2.1, 2.2), [6] (pentru acestea nu se vor folosi functii de biblioteca):

1. Metoda one versus one
2. Metoda one versus all

Se vor calcula:

- a) Se vor varia parametri ce definesc un kernel si se va afisa grafic eroarea clasificarii in functie de parametrul variat
- b) Pentru fiecare SVM se va calcula matricea de confuzie [7] pentru setul de antrenare si cel de testare
- c) Pentru fiecare set de date se va crea un fisier .pdf in care se vor prezenta rezultatele obtinute:
  - i) Marime set de date de test
  - ii) Tip kernel
  - iii) Numar vectori suport
  - iv) Numar iteratii
  - v) Parametri specifici fiecarui kernel
  - vi) Acuratete
  - vii) Eroarea calculata la punctul a)
  - viii) Matricea de confuzie set de antrenare, set de testare (calculata la punctul b)
- d) In fisierul .pdf se vor indica cele mai bune valori obtinute pentru fiecare SVM corespunzator fiecarui set de date

**Bonus:** Se va implementa k-fold-cross-validation [8], [9] si se vor prezenta rezultatele de la punctul c) (nu se vor folosi functii existente).

## Bibliografie

- [1] <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [2] [https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#skin\\_nonskin](https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#skin_nonskin)
- [3] <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/multiclass.html#news20>
- [4] <http://crsouza.com/2010/03/17/kernel-functions-for-machine-learning-applications/#gaussian>
- [4] [http://www.keerthis.com/multiclass\\_mcs\\_kaibo\\_05.pdf](http://www.keerthis.com/multiclass_mcs_kaibo_05.pdf)
- [5] <https://courses.media.mit.edu/2006fall/mas622j/Projects/aisen-project/>
- [6] <https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>
- [7] <https://machinelearningmastery.com/k-fold-cross-validation/>
- [8] <https://towardsdatascience.com/cross-validation-explained-evaluating-estimator-performance-e51e5430ff85>