Proiect Inteligenta Artificiala

Istrate Dorian-Mihai

Facultatea de Matematica si Informatica

Calculatoare si Tehnologia Informatiei

Anul 3, Grupa 364

Cuprins

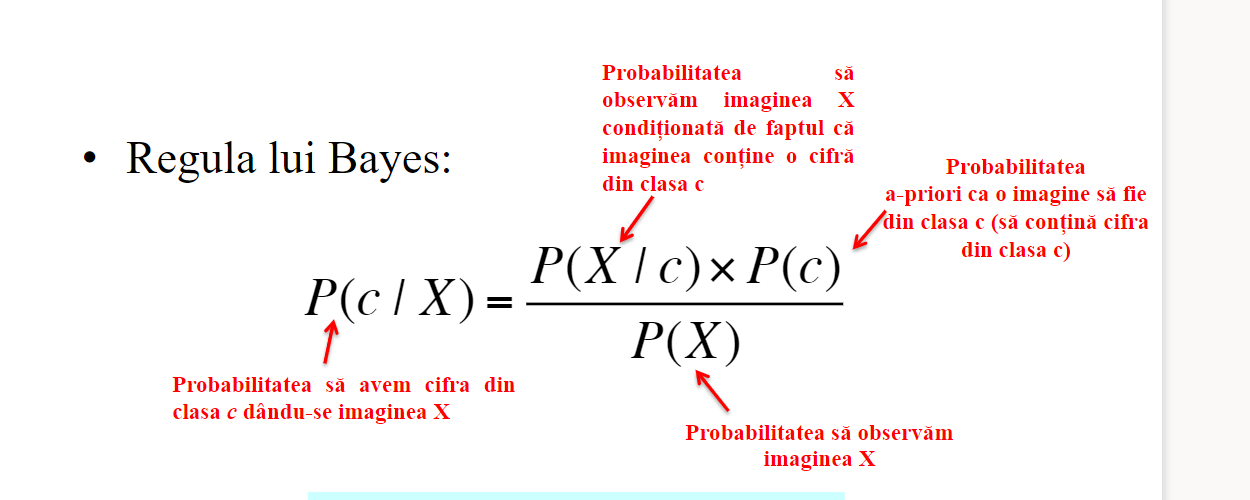
1 Multinomial Naïve Bayes

* 1. Caracteristici folositi
  2. Hiperparametrii si antrenarea acestora
  3. Durata antrenare

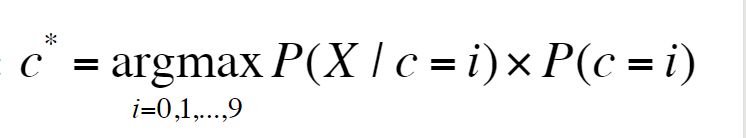
1. Linear SVC
   1. Caracteristici folositi
   2. Hiperparametrii si antrenarea acestora
   3. Durata antrenare

3 Concluzie

Clasificatorul Naïve Bayes reprezinta o familie de clasificatori probabilistici simpli, bazati pe aplicarea teoremei lui Bayes. Aceasta presupune calcularea sanselor unor evenimente pe baza unor informatii a priori.



Clasificatorul va alege clasa c\* care sa maximizeze numaratorul in prezicerea etichetei.

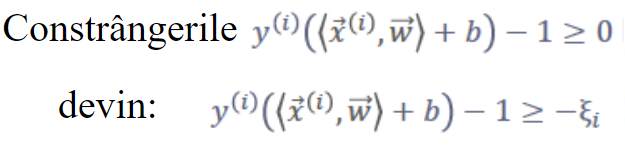
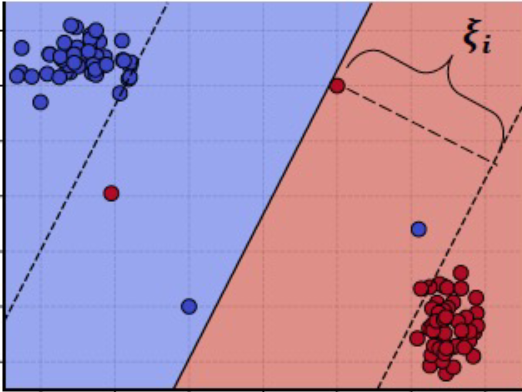


Durata de antrenare + predictie: 20 de secunde

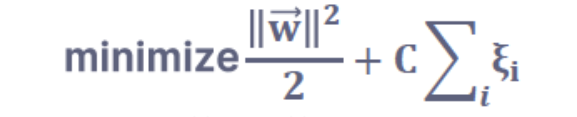
Un SVM (support vector machine = mașină cu vectori suport) este un clasificator liniar binar nonprobabilist.

* clasificator: metoda de invatare supervizata care are ca scop predictia de clase
* liniar: frontiera de decizie este un hiper plan in n-dimensiuni
* binar: invata sa discrimineze intre 2 clase (clasa + si clasa -)
* nonprobabilist: rezultatul unui SVM nu este limitat, nu poate fi interpretat ca o probabilitate

Am folosit un model Soft-margin SVM, deoarece permite ca anumite exemple de antrenare sa fie misclasificate (clasificate gresit), cu un cost (penalitate). Variabilele ξi >= 0, cuantifica cat de mult exemplul de antrenare x(i) este de partea cealalta a marginii.



Problema se reduce la:



**Parametri:**

Hiperparametrul C controleaza tradeoff-ul dintre maximizarea marginii si acuratetea modelului.

Al doilea parametru este kernel trick, care consta in folosirea unei functii kernel (nucleu) ϕ care mapeaza datele intr-un alt spatiu in care datele sunt liniar separabile. Am folosit kernelul liniar. 

SVM fiind un clasificator binar (prin constructie) iar etichetele distincte in numar de 3, am ales metoda One-versus-one (OVO). Acesta presupune antrenarea a n(n-1)/2 clasificatori, cate unul pentru fiecare pereche de clase. In SKLearn, se realizeaza acest lucru printr-o singura instantiere.

**Durata de antrenare:** 45 de minute, antrenare + validare.

Concluzie

Acest proiect m-a determinat sa inteleg ca nu exista o singura metoda corecta de a clasifica date. Toate functioneaza si returneaza un rezultat concret, insa numai cateva metode au o acuratete optima sau un timp de executie minim. Nu exista reguli dupa care sa poti sti din avans ca acuratetea modelului va creste, totul trebuie testat empiric, si modificarile trebuie facute pe baza observatiei si a teoriei algoritmilor. Inteligenta Artificiala pentru mine reprezinta una din cele mai practice experiente a informaticii, intrucat fiecare submisie am evaluat-o ca pe un experiment.