1 Introducere

Recent, rețelele neuronale artificiale sunt din ce în ce mai întâlnite în învățarea automată, însă acestea nu sunt nicidecum tehnici noi, ele fiind propuse imediat după Cel de-al Doilea Război Mondial. Mai exact, prima rețea neuronală a fost construită în anul 1948 și a încercat să propună un model matematic pentru modul în care funcționează neuronii biologici. Marile piedici pentru *deep learning*[1] în acea perioadă erau complexitatea computațională în procesul de antrenare și necesitatea unui nivel mare de date de antrenament pentru a obține o performanță bună. Astfel, rețelele neuronale și-au pierdut atractivitatea, fiind preferate alte metode de clasificare precum SVM (Support Vector Machines) sau clasificatori liniari. Odată cu creșterea în popularitate a internet-ului, a crescut și nivelul de date distribuite public, facând tehnicile de deep learning viabile în contextul actual, cu performanțe chiar mai bune decât metodele clasice.

Domeniul *Computer Vision* [2] a cunoscut un adevărat progres în jurul anului 2012 atunci când Alex Krizhevsky, Ilya Sutskever și Geoffrey Hinton au construit ceea ce s-a numit AlexNet.[1] O rețea neuronală convoluțională [3] prin care au obținut o eroare de 15.3% în cadrul competiției ImageNet LSVRC-2010 care constă în clasificarea a 1.2 milioane de imagini de rezoluție înaltă într-una din 1000 de clase. Acest rezultat a fost cu 10.8% mai bun decăt precedentul, arătând că tehnicile de deep learning au un potențial enorm în problema recunoșterii de imagini. La data scrierii acestei lucrări, eroarea în cadrul competiției este $\approx 2.9\%$.

Prima rețea convoluțională a fost creată în anul 1998 de către Yann LeCun, numindu-se LeNet-5. [2] Scopul ei a fost clasificarea cifrelor scrise de mână, fiind inspirată de anumite descoperiri din biologie care s-au referit la faptul că, în cortexul vizual, există neuroni care răspund individual la regiuni mici dintr-un anumit stimul, creierul neprocesând o imagine ca un tot unitar.

[1] Familie de algoritmi de învățare automată ce au la bază rețelele neuronale

[2] Domeniu al inteligenței artificiale ce își propune înțelegerea imaginilor și a video-urilor de către calculator

[3] Rețea neuronală folosită în recunoașterea de imagini cu ajutorul operației de convoluție

În cadrul acestei lucrări vom studia folosința rețelelor neuronale convoluționale, a celor *obișnuite*, cât și a altor tehnici de învățare automată pentru a asigna independent unei instanțe, definită în problemă ca o mulțime de fotografii dintr-un reastaurant, fiecare dintre următoarele clase:

1. bun pentru prânz (good for lunch)
2. bun pentru cină (good for dinner)
3. acceptă rezervări (takes reservations)
4. are sejur în aer liber (outdoor seating)
5. este scump (restaurant is expensive)
6. oferă alcool (has alcohol)
7. are serviciu de masă (has table service)
8. atmosfera este rustică (ambience is classy)
9. bun pentru copii (good for kids)

Această problemă a fost propusă în anul 2015 de cei de la Yelp prin platforma Kaggle.[4] Motivația din spate a fost că răspunsurile pentru întrebările de mai sus reprezintă un factor important în sistemele lor de recomandări, însă utilizatorii nu le oferă foarte des. În acest caz, un model de învățare automată care primește ca date de intrare fotografii dintr-un restaurant și oferă o valoare din mulțimea $\{0, 1\}$ pentru fiecare clasă ar fi de folos.

Structura datelor este următoarea:

- 234842 de imagini de antrenament în format .jpg și .png.
- 237152 de imagini de test folosite pentru a determina scorul în cadrul competiției.

- ***train_photo_to_biz_ids.csv***: tabel ce asociază fiecare fotografie de antrenament la restaurantul din care provine folosind id-uri.

- ***train.csv***: tabel ce conține 2 coloane $\{business\_id, labels\}$ semnificând etichetarea unui restaurant (prin id-ul său) cu o submulțime din $\{1, 2, ..., 9\}$ ce reprezintă cele 9 clase ilustrate mai sus.

- ***test_photo_to_biz_ids.csv***: tabel ce asociază fiecare fotografie de test la restaurantul din care provine folosind id-uri.

În total sunt 1996 de restaurante în setul de antrenament și 10000, ce trebuie clasificate, în setul de testare.

2 Problema clasificării. Tipuri de clasificare

2.1 Învățare supervizată

În contextul învățării supervizate ne este pus la dispoziție un set de date format din perechi (x_i, y_i), x_i numindu-se *instanță* și y_i *etichetă*. Obiectivul unui algoritm de învățare automată este de a învăța o funcție care să modeleze cât mai bine relația dintre x_i și y_i ce reiese din date. După natura variabilei y, se pot defini 2 tipuri de probleme:

- Clasificare (variabila y este discretă)
- Regresie (variabila y este continuă)

În această lucrare vom aborda doar subiectul clasificării.

2.2 Clasificare binară

Problema clasificării binare ne cere ca pentru o instanța $x \in \mathbb{R}^p$ $p \in \mathbb{N}^*$ dată ca input și o clasă C, să returnăm o valoare $y \in \{0, 1\}$ astfel încât $y = 1$ dacă x face parte din clasa C, sau $y = 0$ în caz contrar. Mai formal, trebuie să gasim o funcție $h : \mathbb{R}^p \to \{0, 1\}$ definită astfel:

$$h(x) = \begin{cases} 1, & \text{x} \in \text{C} \\ 0, & \text{altfel.} \end{cases}$$

Această funcție se mai numește model sau ipoteză și scopul unui algoritm de învățare automată este de a oferii un model cât mai bun.
Un exemplu de clasificare binară este ca pentru o imagine să decidem dacă în aceasta se află sau nu o pisică. În acest caz, instanța x va fi un vector ce conține valoarea reală a fiecarui pixel și output-ul va fi $y = 1$ dacă în imagine apare o pisică sau $y = 0$ în caz contrar.

2.3 Clasificare cu clase multiple

La fel ca în cazul clasificării binare, instanța este $x \in \mathbb{R}^p$ $p \in \mathbb{N}^*$, însă, în loc de o singură clasă C, avem o mulțime $\{C_1, C_2, ..., C_k\}$ $k \in \mathbb{N}^*$ disjuncte, iar obiectivul este să găsim o

funcție $h : \mathbb{R}^p \to \{0, 1, 2, ..., k\}$ astfel încât:

$$h(x) = \begin{cases} 1, & \text{x} \in C_1 \\ 2, & \text{x} \in C_2 \\ \vdots & \vdots \\ \text{k}, & \text{x} \in C_k \\ 0, & \text{altfel.} \end{cases}$$

2.4 Clasificare cu etichete multiple

Pentru o instanță definită la fel ca mai sus $x \in \mathbb{R}^p$ $p \in \mathbb{N}^*$ și o mulțime $C = \{C_1, C_2, ..., C_k\}$ $k \in \mathbb{N}^*$ vom construi o ipoteză $h : \mathbb{R}^p \to \{0, 1\}^k$ astfel încât pentru o predicție $y = h(x)$ să avem:

$$y_i = \begin{cases} 1, & \text{x} \in C_i \\ 0, & \text{altfel.} \end{cases} \quad i = \overline{1, k}$$

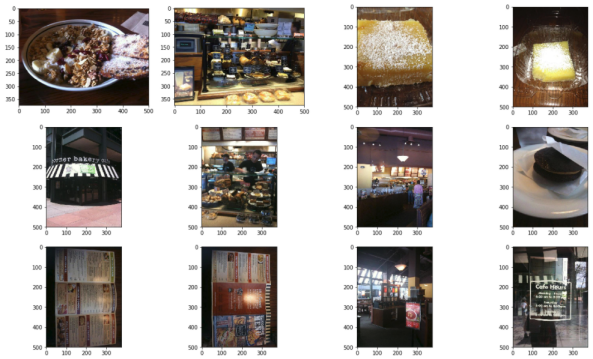
Apare des o confuzie între clasificarea cu etichete multiple și cea cu clase multiple. Diferența fundamentală este ca în prima o instanță poate să aparțină mai multor clase, spre deosebire de cealaltă, în care o instanță este asignată unei singure clase. Totodată, clasificarea binară este un caz particular a clasificării cu etichete multiple atunci când $k = 1$. Detectarea obiectelor face parte din acest tip de clasificare, unde, pentru o imagine, vom marca prezența mai multor obiecte (clase) și nu a unui singur element.

2.5 Învățare multi-instanță

În acest caz, o instanță nu mai este reprezentată de un singur vector de numere reale, ci de o mulțime de astfel de vectori. Fie $X = \{x_1, x_2, ..., x_k\}$ $k \in \mathbb{N}^*$ $x_i \in \mathbb{R}^{p_i}$ $p_i \in \mathbb{N}^*$ o instanță. O ipoteză h determinată de un algoritm de învățare automată va clasifica întreaga mulțime X și nu fiecare componentă x_i independent. [3]

Problema abordată în această lucrare este una de învățare multi-instanță și de clasificare cu etichete multiple. Un restaurant este reprezentat de o mulțime de imagini și rezultatul

este un vector y cu 9 componente, fiecare marcând prezență sau absență celor 9 clase descrise în capitolul introductiv.



$$\longrightarrow \quad y = [1, 0, 0, 1, 1, 0, 1, 1, 0]$$

O instanță formată din 12 fotografii și vectorul asociat

Bibliografie

[1] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton *ImageNet Classification with Deep Convolutional Neural Networks* 2012.

[2] Yann LeChun, Leon Bottou, Yoshua Bengio, Patrick Haffner *Gradient-Based Learning Applied to Document Recognition* 1998.

[3] Jaume Amores *Multiple Instance Classification: review, taxonomy and comparative study* Computer Vision Center, Computer Science Department, UAB, Spain 2013.

[4] `https://www.kaggle.com/c/yelp-restaurant-photo-classification`