

Математичка гимназија

МАТУРСКИ РАД

- из предмета програмирање и програмски језици -

Обрада звучне датотеке - машинско учење

Ученик:
Ђорђе Стефановић IVд

Ментор:
Јелена Хаџи-Пурић

Београд, јун 2020.

Садржај

1	Увод	1
2	Људска перцепција звука	2
2.1	Људско чуло слуха	2
2.2	Карактеристике звука	2
2.3	Особине слуха	3
3	Дигитализација звука	6
3.1	Приказивање звука	6
3.1.1	Аудио сигнал	6
3.1.2	Краткорочна Фуријеова трансформација	6
3.1.3	Мел спектрограм	7
3.1.4	Constant-Q трансформација	7
3.1.5	Хромаграм	7
3.2	Меморисање звука	8
3.3	Формати звучних датотека	9
3.4	Обрада звучних датотека	10
3.5	Машинско учење у обради звука	10
4	Машинско учење	11
4.1	Увод	11
4.2	Неуралне мреже	12
4.2.1	Обрада података у неуралној мрежи	12
4.2.2	Учење у ацикличној неуралној мрежи	13
4.2.3	Учење у рекурентној неуралној мрежи	15
4.2.4	Активациона функција	17
5	Обрада звука	19
5.1	Врсте слојева и њихове примене	19
5.1.1	Густо повезани слој	19
5.1.2	Конволуциони слој	20

САДРЖАЈ

5.1.3	Рекурентни слој	20
5.2	Примене у свакодневном животу	20
6	Закључак	21
	Литература	21

1

Увод

Људи су одувек идеје за своје изуме добијали посматрајући природу и откривајући како она ради. Те механизме који проверено раде су преточили у апарате који олакшавају живот. Видевши да се камен котрља брже него што се клиза, изумљен је точак. Видевши како мачка лови миша, изумљена је мишоловка. Посматрајући како функционише људски мозак, добијене су идеје за неуралне мреже.

Чуло слуха је једно од основних људских чула којима се опажа околина. Развој технологије је омогућио да се звук брже и прецизније снима, анализира и обрађује.

2

Људска перцепција звука

2.1 Људско чуло слуха

Звук представља осцилацију која може надражити чуло слуха. Звук се, као и сваки други талас, простира тако што честице до којих је талас стигао почну да осцилују и та осцилација покреће суседе којима се преда енергија и који сами настављају процес. Када тај талас доспе до унутрашњег уха, стварају се електрични сигнали који се у мозгу трансформишу у оно што људи опажају као звук. Људско ухо се састоји из три дела: спољашњег уха, средњег уха и унутрашњег уха. Спољашње ухо скупља таласе из средине и каналише их ка средњем уху, које их потом шаље ка унутрашњем уху. У унутрашњем уху се налази специјални орган који носи име пуж. У њему се налазе длачице које вибрирају на различите начине у зависности од фреквенције. Под утицајем тих вибрација генеришу се електрични импулси који се преко слушног нерва преносе до мозга где се разазнају као звук.

2.2 Карактеристике звука

Звук, као и сваки талас, може се описати следећим особинама:

- (1) фреквенција/таласна дужина
- (2) амплитуда/ звучни притисак/ интензитет
- (3) брзина простирања
- (4) правац и смер простирања

Људи не опажају ове физичке величине, већ њихове последице.

- (1) Висина - последица циклично понављајуће природе звука. Људски слушни механизам анализира звук у потрази за понављајућим секвенцама, што проузрокује јединствену висину која се чује у датом тренутку.
- (2) Јачина - последица броја надражених рецептора по времену. Број надражених рецептора зависи од интензитета звука, његовог трајања и сложености. Талас већег интензитета ће надражити више рецептора, талас који се кратко простире ће бити опажен као тиши, а сложен талас ће надражити више различитих рецептора због чега ће се тај сигнал чути као гласнији.
- (3) Боја - последица интерференције таласа. Већина извора звука производи спектар таласа, и човек повезује извор са његовим звуком. Због овога, на пример, врло се лако разликују различити говори и препознају особе.
- (4) Дужина - време које протече од тренутка када човек опази звук до тренутка када опази да га нема или да се променио.
- (5) Просторна локација - човек, као и већина бића која имају чуло слуха, има два органа за пријем звучних сигнала. Разлика у временима регистравања и јачинама таласа садрже довољно информација за приближно одређивање локације извора звука.

Како је звук промена положаја честица средине, мерењем промене притиска кроз време могу се добити све потребне информације за репродукцију. Опсег промена које могу да се примете је веома широк, због чега се уводи нова скала, децибели.

$$L_p = 10 \log_{10} \left(\frac{p^2}{p_{ref}^2} \right) Db.$$

L_p – ниво звучног притиска у децибелима, p – средња квадратна вредност притиска средине, p_{ref} – референтна вредност притиска, за ваздух је стандардна вредност $20\mu Pa$. На пример, просечан људски говор је интензитета $60Db$.

2.3 Особине слуха

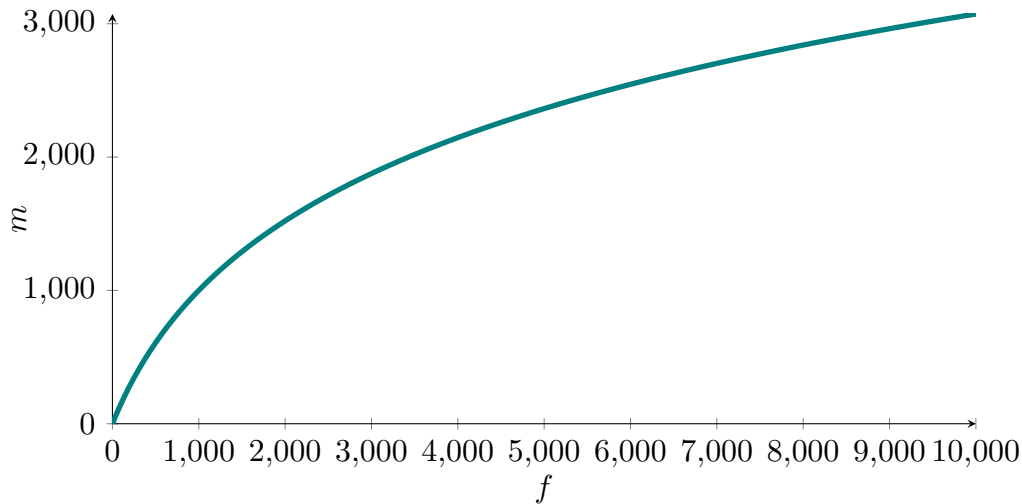
Просечно људско ухо може да чује тонове фреквенције од $20Hz$ до $20KHz$. Минимална разлика две фреквенције које се чују као различите је око $3.6Hz$,

иако се и мање разлике могу приметити. Интерференција два тона блиских таласних дужина ће се чути као промена јачине звука, а та појава се назива пулсирање.

Висина тона и његова фреквенција нису линеарно зависни, и еквидистантне фреквенције неће увек имати еквидистантне висине. Због тога је одређена скала која приближно трансформише фреквенције у нову јединицу, мел, која мери разлике у висинама тонова.

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

где је m висина звука у мелима, а f фреквенција у херцима.



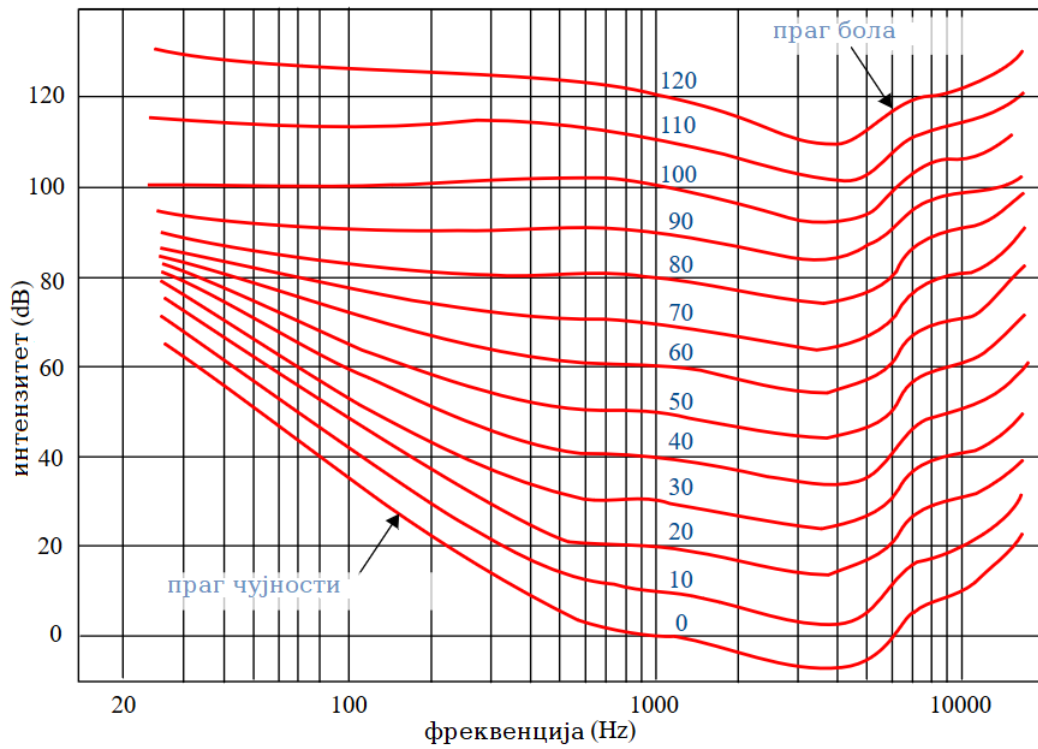
Дужина звука који се чује није увек пропорционална његовој стварној дужини. На пример, у бучном окружењу испрекидани звук ће се чути као континуалан, јер прекиде у његовом опажању може привремено да замени позадински шум. Такође, да не би обрађивао превише информација, мозак може да игнорише одређене звукове који у нису од превеликог значаја, или да се прилагоди навикавањем на драж која се монотono понавља.

Тоновни који настају стојећим таласима се јављају у виду сложеног таласа који се састоји од целобројних умножака основног хармоника. Ако је основна фреквенција осциловања f , тон ће настати интерференцијом са вишим хармоницима који осцилују фреквенцијама $2f$, $3f$, ... Људско ухо је навикло на овакав тип таласа и тонове којима фали основни хармоник и даље региструје на исти начин.

Надражљивост зависи од фреквенције. Минимални интензитет звука који може да се опази зове се праг чујности. Минимални интензитет звука који ће повредити ухо се назива праг бола.

Механизам за пријем и обраду звука у људском мозгу не репродукује увек

надражај директно пропорционалан спољашњој дражи, већ најпогоднији за преживљавање.



Слика 2.1: Зависност прага чујности од фреквенције

3

Дигитализација звука

3.1 Приказивање звука

Звук се најчешће приказује као дводимензионални график, спектрограм, чија је једна оса време, а друга фреквенција. Ове репрезентације се могу сматрати као слике и обрађивати се као такве, мада постоје разлике у неким карактеристикама. Код слика, суседни пиксели најчешће имају сличне особине, док се на спектрограмима корелације јављају по хармоницима на оси која симболизује фреквенцију. Постоји више различитих метода које могу да забележе звучни сигнал.

3.1.1 Аудио сигнал

Ова метода не мења добијени сигнал, него га приказује у изворном облику. Састоји се од исечака који одређују амплитуду у датом временском интервалу.

3.1.2 Краткорочна Фуријеова трансформација (енг. Short-Time Fourier Transform/STFT)

Приказује се промена фреквенције кроз време са линеарно раздвојеним фреквенцијама. Захваљујући брзој Фуријеовој трансформацији, овакав начин приказивања се најбрже генерише. Линеарно раздвојене фреквенције нису увек најпогодније за анализу звука, јер нису компатибилне људском слуху. Предност ове методе је што се може вратити назад у аудио сигнал.

3.1.3 Мел спектрограм

Ова дводимензионална репрезентација звука је конструисана и оптимизирана према људском чулу слуха и компресује фреквенције према меловој скали, чиме смањује меморијско оптерећење са очувањем најзначајнијих информација. Мел спектрограм приказује само енергије таласа дате фреквенције из чега се не може добити њихова амплитуда, због чега процес није повратан (не може се добити аудио сигнал).

Постоје и друге методе које функционишу по истом принципу, оптимизоване за људски слух, као што је на пример барк скала.

3.1.4 Constant-Q трансформација (енг. Constant-Q Transform/CQT)

CQT представља скуп класа фреквенција, са логаритамски скалираним централним фреквенцијама.

$$f_k = f_0 \cdot 2^{\frac{k}{b}}$$

f_k - фреквенција k -тог центра, f_0 - најмања фреквенција која се анализира, b - број центара по октави. Да би се ови центри додиривали, сваки мора имати своју ширину Δ_k .

$$\Delta_k = f_{k+1} - f_k = f_k(2^{\frac{1}{b}} - 1)$$

$$Q = \frac{f_k}{\Delta_k} = \frac{1}{2^{\frac{1}{b}} - 1}$$

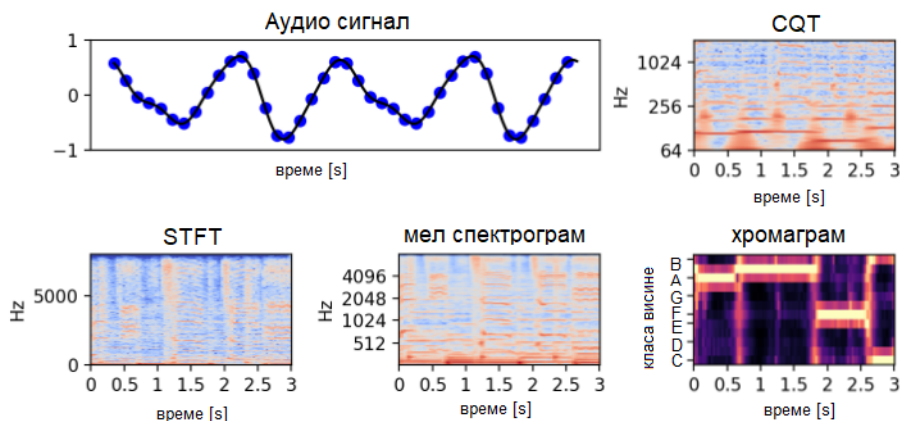
Однос фреквенција и дужина Q је константан, по чему је ова метода добила име. Ако се за b узме број 12, а за f_0 фреквенција прве ноте солмизације, сви центри ће одговарати одређеним нотама, што олакшава анализу музике.

3.1.5 Хромаграм

Хромаграм представља распоред енергије према класама висине, најчешће представљених нотама солмизације. Може се посматрати као CQT испресавијан тако да се сви центри у истим класама преклапају. Ова метода је такође погодна за обраду музике.

Ово нису све методе које се користе за визуелно приказивање звука, али се најчешће примењују.

Слика 3.1: Различите врсте приказа звука



3.2 Меморисање звука

Као и човек, да би рачунар могао да сачува аналогни звучни сигнал, он га некако мора прво пребацити у електрични, који ће се касније записати у неки низ нула и јединица. Микрофони се најчешће користе да аналогни сигнал пребаце у дигитални. Они функционишу слично као и људско ухо – пребацују механичку енергију у електричну. Постоји танка мембрана која вибрира и покреће навојницу која је намотана на стални магнет. Пошто вибрације померају положај навојнице у односу на магнет, мења се магнетно поље које индукује електричну струју, чија фреквенција и интензитет зависе од фреквенције и интензитета звука. Звучници раде по истом принципу, али у обрнутом смеру. Тако добијен електрични сигнал треба се сачувати. Метода за чување се зове импулсна кодна модулација (енг. Pulse code modulation/PCM). Сигнал се мери на кратким интервалима одређене дужине и амплитуда се заокружује на најближу дигиталну вредност.

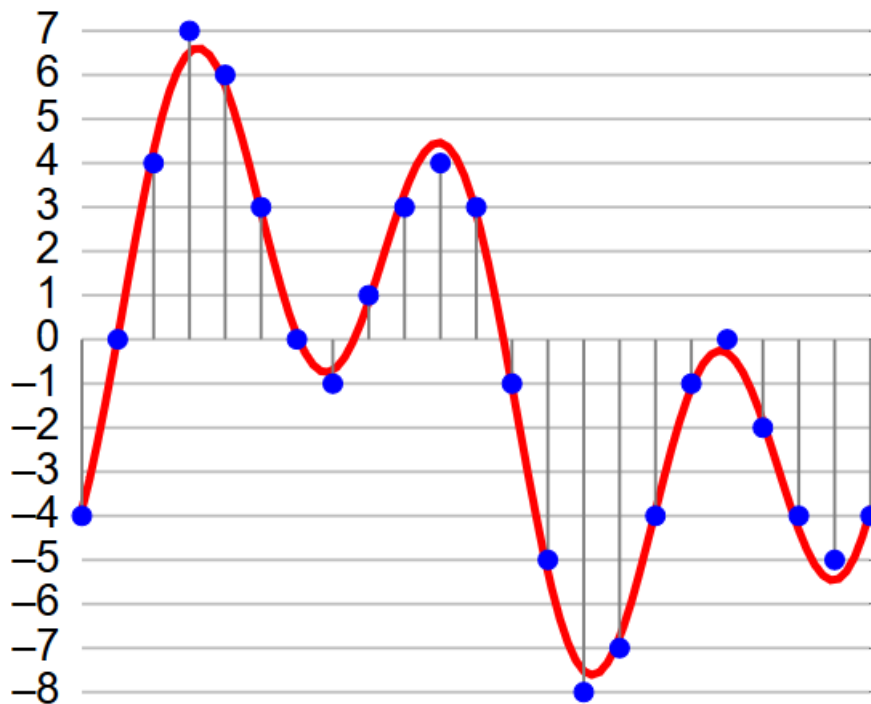
Као што филмови нису континуалан ток и прави покрети него низ статичних слика који се смењују довољно брзо, тако и овај низ дискретних амплитуда ће људском уху звучати као континуалан сигнал. Потребно је извршити три корака да би се извршила ова метода:

- (1) Одмеравање – дискретизација континуалног сигнала. Узима се вредност измерена за једнаке интервале времена
- (2) Квантизација – опсег у коме се мери се подели на коначан број интервала квантизације са својим вредностима, нивоима квантизације. Вредност

измерена у датом тренутку се класификује у најближи интервал квантизације и заокружује на његову вредност.

- (3) Кодирање – пошто се неке амплитуде јављају чешће од других и опсег није увек фиксан, обично линеарно кодирање није најефикасније решење. Вредности нивоа квантизације узимају се као функција амплитуде по A-law алгоритму или μ -law алгоритму.

Слика 3.2: импулсна кодна модулација



3.3 Формати звучних датотека

Звучне датотеке се могу чувати на различите начине. Основна разлика између формата је компресија. Неки формати, попут .wav или .aiff уопште не користе никакву компресију него чувају сигнал у изворном облику. Компресија служи да смањи меморијско оптерећење рачунара. Компресија може бити са или без губитака у информацији. Оба су применљива, зависи да ли је потребнији квалитет или штедња меморије. Формати попут .flac или .alac су без губитака док су .mp3 или .aac са губицима. Губици су такви да човек што

теже примети разлику. Људско ухо не разликује тонове блиских фреквенција, већ слушни систем на њих одреагује исто. Алгоритми компресије углавном предвиђају вредност тона у зависности од његових суседа. Такође се врло често сложен сигнал разлаже на суму синусних/косинусних функција користећи брзу Фуријеову трансформацију или дискретну косинусну трансформацију.

3.4 Обрада звучних датотека

Као и свака друга датотека, и звучне датотеке се могу уређивати у програмима за обраду текста. Међутим, иако овај начин обраде делује једноставно, смислене звучне датотеке је врло тешко добити, јер је потребно знати сва унапред дефинисана заглавља, како се физички записују секвенце звука у меморији и како се оне виде у програмима за обраду текста. Као и код других типова датотека, иако постоји бијекција између машинског кода и човеку смисленог сигнала, она није једноставна за тумачење и зато постоје многобројни програми за манипулацију звучних датотека. У њима је могуће изводити велики број операција, од једноставних попут мењања дужине и јачине, обртања, спајања, до комплексних попут уклањања шума, појачавања/смањивања интензитета одређеног инструмента или допуњавања.

3.5 Машинско учење у обради звука

Искусни музичари врло лако могу да разликују инструменте и да препознају њихове карактеристике. После многобројних часова проведених свирањем и слушањем различитих инструмената, они су истренирали своје ухо да препознаје суптилне разлике између тонова. Зашто исти принцип не би могао да се примени и на рачунаре? Они имају много бржу оперативну меморију од људи и способни су да извршавају стотине хиљада операција у секунди. Програми који користе алгоритме машинског учења нашли су примену и у овом пољу.

4

Машинско учење

4.1 Увод

Идеје о интелигентним машинама су почеле да се јављају са развојем дигиталних машина средином двадесетог века, али су тек недавно постале више од визионарске маште. Машинско учење је процес у коме програм без директних команди свог творца доноси одлуке на основу претходно обрађених примера. Циљ машинског учења је генерализација на основу искуства, односно да програм даје тачно решење и на примерима које није претходно видео. Такви програми постају све чешћи због своје практичности и флексибилности. Постоје различити приступи учења рачунара да решава проблеме за које не постоји конкретан алгоритам. Уобичајена је подела на три категорије:

- (1) Супервизирано учење - рачунару се предаје низ улазних података и низ жељених излазних података за сваки унос. Циљ је да алгоритам пронађе правило које повезује улазе са излазима и на примерима који нису претходно задати. Овакви алгоритми се најчешће користе за решавање проблема класификације.
- (2) Несупервизирано учење - рачунару се предају само улазни подаци и алгоритам има циљ да их препозна и класификује према сопственој процени. Овакви алгоритми су корисни у препознавању шаблона у подацима или за решавање проблема груписања.
- (3) Учење са појачавањем - програм интерагује са динамичким окружењем и извршава одређене радње које мењају стање система. За успешно одрађене акције програм добија „награду”, односно поене, и има циљ да

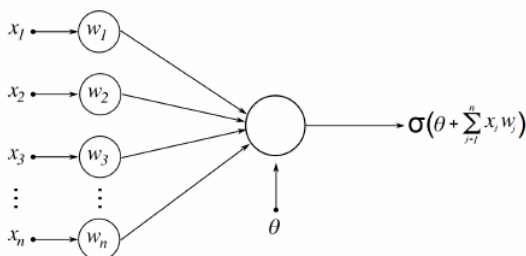
максимизује број поена. Овакви алгоритми налазе примену у интерактивним проблемима као што су аутоматизована возња аутомобила или играње игрица.

4.2 Неуралне мреже

Цела грана машинског учења и његових подврста настала је инспирисана механизмом људског учења и покушавањем његове имитације. Неуралне мреже нису изузетак - њихов узор представља нервни систем. Састоје се из чворова, вештачких неурона, који су међусобно повезани, као синапсе између неурона. Сваки чвор може да обради и проследи сигнале својим суседима. Као и код неурона у мозгу, један чвор може имати више улаза, али само један излаз. Неуралне мреже су подврста супервизираниог учења.

4.2.1 Обрада података у неуралној мрежи

У неуралним мрежама, сигнал који се шаље од чвора до чвора је неки реалан број.



Слика 4.1: Пример чвора

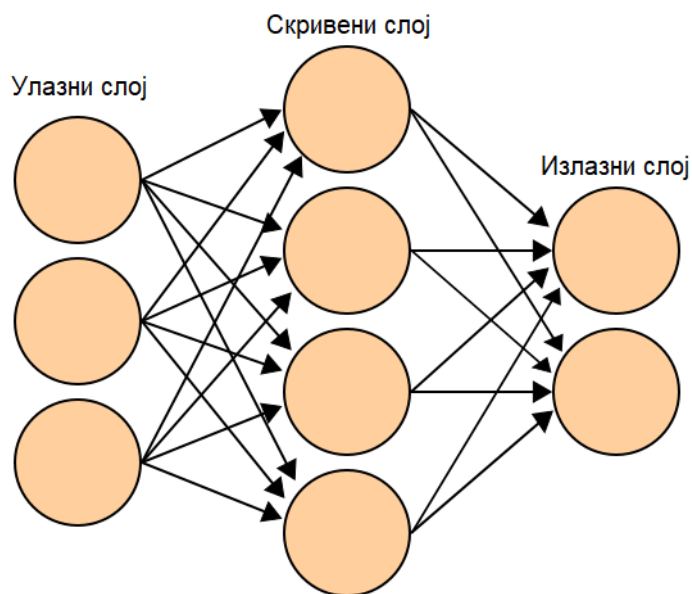
Један чвор има n улаза (x_1, x_2, \dots, x_n) са својим тежинама (w_1, w_2, \dots, w_n) . Сваки чвор има и свој слободан члан θ , који је користан за случајеве када је улаз 0. На резултат се примењује активациона функција σ и тај резултат се шаље даље у мрежи.

Мрежа се састоји из три дела: улазног, излазног и скривеног слоја. Мреже са више скривених слојева се називају дубоке неуралне мреже. Према начину пропације сигнала, дели се на две основне врсте:

- (1) ацикличне мреже - сигнал се преноси са слоја на слој почевши од улаза, а везе увек преносе информацију са нижег слоја на виши. Рад сваког

слоја започиње након завршетка претходног. Овакве мреже сваки улаз обрађују независно од осталих.

- (2) рекурентне мреже - везе између чворова могу правити циклусе, и обрађена информација се може вратити у чвор из кога је потекла. Овакве мреже могу да обрађују више различитих улаза истовремено и да мешају њихове утицаје што је веома значајно код обраде звука где се информације о једном низу тонова не могу преставити само једним улазним параметром.



Слика 4.2: Неурална мрежа

4.2.2 Учење у ацикличној неуралној мрежи

Учење је главна разлика по којој се алгоритми машинског учења разликују од осталих алгорита. Код неуралних мрежа се то постиже подешавањем тежина и слободних чланова. Функција губитка $J(w)$ представља разлику између добијеног резултата и предвиђеног резултата у зависности од тежина. Током процеса учења алгоритам има циљ да смањи функцију губитка пробавањем различитих комбинација тежина и слободних чланова. Нове вредности се узимају тако да функција губитка најбрже опада.

$$w_{k+1} = w_k - \eta \nabla J(w) \quad (4.1)$$

w - скуп тежина, k број итерације, η предетерминисана стопа учења, $\nabla J(w)$ градијент функције губитка

Да би се ово израчунало, потребно је знати вредности парцијалних извода $\frac{\partial J(w)}{\partial w_{ij}}$ где је w_{ij} тежина везе између чвора i и j .

Нека је in_i сума свих вредности која долазе у чвор i , а out_i вредност коју израчуна.

$$in_i = \theta_i + \sum_k w_{ki} \cdot out_k \quad (4.2)$$

$$out_i = \sigma(in_i) \quad (4.3)$$

Применом правила о изводу сложене функције важи:

$$\frac{\partial J(w)}{\partial w_{ij}} = \frac{\partial J(w)}{\partial in_i} \cdot \frac{\partial in_i}{\partial w_{ij}} \quad (4.4)$$

$$\frac{\partial in_i}{\partial w_{ij}} = \frac{\partial \left(\theta_i + \sum_k w_{ki} \cdot out_k \right)}{\partial w_{ij}} = out_i \quad (4.5)$$

Нека је $\frac{\partial J(w)}{\partial in_i} = \delta_i$.

$$\frac{\partial J(w)}{\partial w_{ij}} = \delta_i \cdot out_i \quad (4.6)$$

Параметре δ_i можемо одредити пропацијом уназад, почевши од излазног слоја ка улазном. За излазни слој поново се примењује правило о изводу сложене функције:

$$\delta_i = \frac{\partial J(w)}{\partial out_i} \cdot \frac{\partial out_i}{\partial in_i} \quad (4.7)$$

Како је $out_i = \sigma(in_i)$ и зависи само од in_i , парцијални извод $\frac{\partial out_i}{\partial in_i}$ је обичан извод функције σ у тачки in_i .

$$\frac{\partial out_i}{\partial in_i} = \sigma'(in_i) \quad (4.8)$$

Функције губитка се обично уводе као суме разлика добијених и очекиваних вредности за сваки чвор. Стога се $\frac{\partial J(w)}{\partial out_i}$ може израчунати поређењем добијених и очекиваних вредности излазних чворова.

За чворове који нису у излазном слоју промена вредности у њему ће утицати на све чворове који су повезани са њим. Вредности δ_i се могу израчунати преко већ добијених параметара.

$$\delta_i = \sum_k \frac{\partial J(w)}{\partial in_k} \cdot \frac{\partial in_k}{\partial in_i} = \sum_k \delta_k \frac{\partial in_k}{\partial in_i} \quad (4.9)$$

$$\frac{\partial in_k}{\partial in_i} = \frac{\partial \left(\theta_k + \sum_j w_{jk} \cdot \sigma(in_j) \right)}{\partial in_i} = w_{ik} \sigma'(in_i) \quad (4.10)$$

$$\delta_i = \sigma'(in_i) \sum_k w_{ik} \delta_k \quad (4.11)$$

Да би се смањила функција губитка, треба да се замене тежине.

$$w_{ij} := w_{ij} - \eta \frac{\partial J(w)}{\partial w_{ij}} \quad (4.12)$$

односно

$$w_{ij} := w_{ij} - \eta \delta_i out_j \quad (4.13)$$

аналогно једначини (4.1).

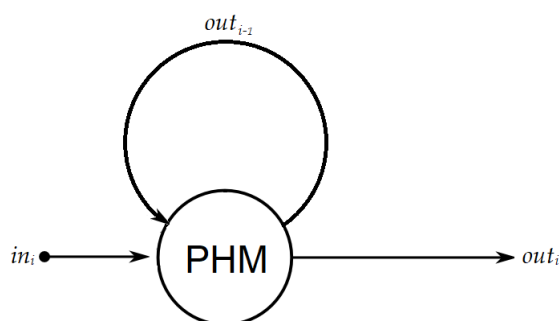
Вредности слободних чланова се ажурирају по истом принципу.

$$\theta_i := \theta_i - \eta \delta_i \quad (4.14)$$

Стопа учења је неки реалан број од кога зависи брзина и прецизост учења неуралне мреже. Већи бројеви ће као резултат имати брже учење, док ће мањи имати мању грешку.

4.2.3 Учење у рекурентној неуралној мрежи

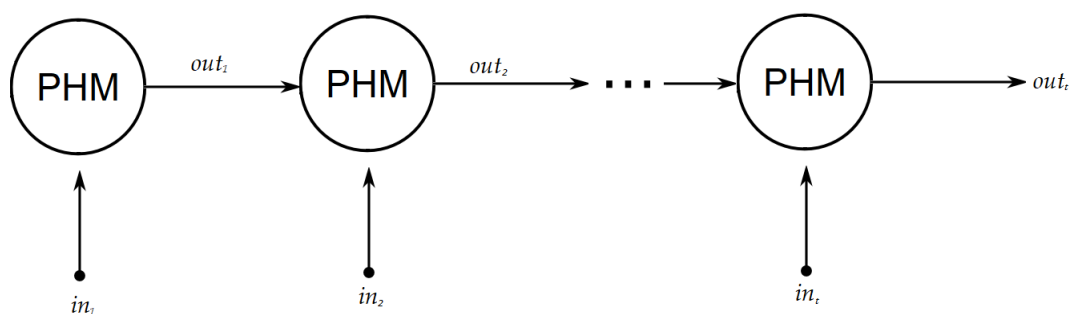
Основна мана ацикличних неуралних мрежа је немогућност обрађивања више улаза паралелно. Људи приликом слушања не обрађују информацију



Слика 4.3: Рекурентна неурална мрежа

само на основу последњег саслушаног тона, већ на основу неког дужег сегмента. На пример, када људи разговарају, да би се разумело шта је речено мора се обрадити реченица од почетка до краја. Због тога се уводе рекурентне неуралне мреже које један улаз могу обрађивати у истом чвору више пута омогућавајући да претходно донете одлуке утичу на нове.

Рекурентна неурална мрежа, иако има циклусе и враћања, у сваком тренутку има јединствено стање. Због тога се, развијањем кроз време, рекурентна неурална мрежа може представити као скуп повезаних ацикличних неуралних мрежа.



Слика 4.4: Развој рекурентне неуралне мреже кроз време

Алгоритам пропагације уназад који је служио за тренирање ацикличне неуралне мреже не може се директно применити на рекурентне, али се може применити на рекурентну неуралну мрежу развијену кроз време.

4.2.4 Активациона функција

Активациона функција се бира у зависности од проблема и улазних података, мада се најчешће користе:

(1) Идентитет: $\sigma(z) = z$

Извод: $\sigma'(z) = 1$

(2) Логистичка функција: $\sigma(z) = \frac{1}{1+e^{-z}}$

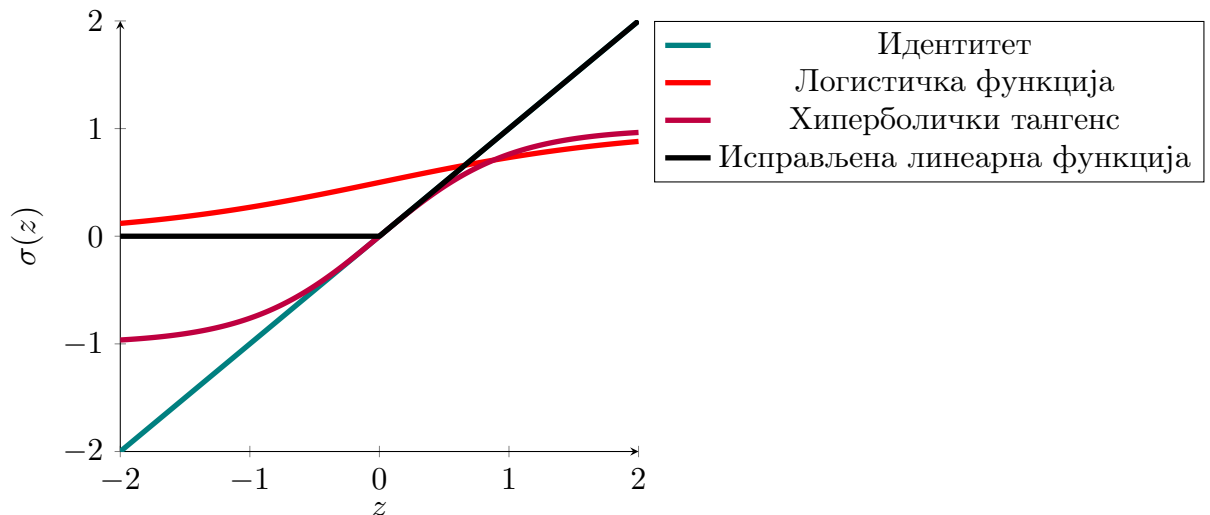
Извод: $\sigma'(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} = \sigma(z) \cdot (1 - \sigma(z))$

(3) Хиперболички тангенс: $\sigma(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$

Извод: $\sigma'(z) = \frac{(e^z + e^{-z})^2 - (e^z - e^{-z})^2}{(e^z + e^{-z})^2} = 1 - \sigma(z)^2$

(4) Исправљена линеарна функција: $\sigma(z) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$

Извод: $\sigma'(z) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$



Како проблеми који су потребни да се решавају нису увек линеарни, уведене су логистичка функција и хиперболички тангенс као активационе функције. Због свог облика који подсећа на латинично слово *s* добиле су назив сигмоидне функције. Међутим, коришћењем сигмоидних функција градијент

$\nabla J(w)$ може значајно да се смањи. У рекурентним или дубоким неуралним мрежама разлика тежина у две итерације зависи од производа извода активационе функције за све слојеве.

$$w_{ij} := w_{ij} - \eta \delta_i \text{out}_i$$

$$\delta_i = \sigma'(in_i) \sum_k w_{ik} \delta_k$$

Како је извод сигмоидних функција увек мањи од један, овај број може да буде веома мали и тежине се неће мењати. Овај проблем се зове проблем нестајућег градијента, и због овога је уведена исправљена линеарна функција. Она такође боље опонаша људски нервни систем који импулс преноси уколико има довољно велики интензитет.

5

Обрада звука

Алгоритми машинског учења су пронашли своје место у обради звука јер постоји доста проблема за које је ово најефикаснији начин да се реше. Неки од њих су алгоритми за препознавање темпа, тонова, инструмената, песама, алгоритми за одређивање жанра и њихову класификацију.

Све неуралне мреже, невезано за њихов контекст, раде по сличном принципу који повезује улазне податке са одговарајућим излазима. За обраду звука, улазни подаци би били неки од звучних репрезентација (аудио сигнал, краткорочна фуријеова трансформација, мел-спектрограм...). Излазни подаци зависе од типа проблема који се решава.

5.1 Врсте слојева и њихове примене

Иако је принцип исти, неуралне мреже се могу конструисати и повезивати на различите начине.

5.1.1 Густо повезани слој

Ако је сваки чвор једног слоја повезан са сваким чвором следећег, ти слојеви су густо повезани. Улазне и излазне параметре можемо посматрати као векторе димензија n где је n број чворова у слоју. Пошто број чворова у излазном и улазном слоју не мора да буде исти, овакве везе су корисне за смањивање, односно повећавање димензија овог вектора. Код обраде звука овакви слојеви се најчешће користе за мапирања, која ће, на пример, из спектрограма извући вектор фреквенција који ће се касније лакше обрађивати.

5.1.2 Конволуциони слој

Главна разлика ових слојева је што се за активацију не користи обичан производ, него операција конволуције. Ова операција је клизајући скаларни производ - скаларни производ транслиран за неку константу.

$$(f * g)[n] = \sum_{i=0}^N f[i] \cdot g[(i + n)_{\text{mod}N}]$$

$$\text{out}_i = \sigma \left(\theta_i + \sum_k w_{ki} * \text{out}_k \right)$$

Овде су и улаз и излаз дводимензионални, а w_{ki} су конволуциона језгра која повезују повезују k -ти канал улаза са i -тим каналом улаза. Дводимензионална конволуциона језгра деле тежине, што као последицу има много мање параметара за тренирање. Резултат овога је мапа карактеристика локалних активирања. За разлику од других слојева, конволуциони слој задржава дводимензионалну архитектуру података. Често се користе за удруживање. Слој за удруживање смањује мапе карактеристика користећи неку функцију, најчешће максимум. Користећи ову функцију претпоставља се да су најбитнији делови на мапи они који су се највише активирали. Понекад се користе и функције које рачунају средњу вредност.

Код обраде звука овакви слојеви су корисни да разазнају информације и одвоје битно од небитног.

5.1.3 Рекурентни слој

Ово је слој у коме се јавља веза са претходним слојем и ствара се циклус. Пошто информације циркулишу кроз слој и више делова информације остају истовремено у њему могу се користити за препознавање понављајућих карактеристика и њихово издвајање.

5.2 Примене у свакодневном животу

Многе апликације су успешно примениле алгоритме машинског учења који помажу у свакодневном животу. Неки од најпознатијих примера су апликације за препознавање говора попут Skype-а или „Siri”. Такође, програми за препознавање песама попут Shazam-а или SoundHound-а су стекли популарност решавајући наизглед једноставне проблеме које људи могу да ураде без превеликог мучења.

6

Закључак

Бројне задатке које човек може врло лако и једноставно да уради рачунар има проблем да изврши. Иако машина може брже да рачуна и обрађује податке, нека поља захтевају процене својствене људима. Препознавање, сналажљивост у новим ситуацијама, способност учења неки су од фактора који су убрзали развој човечанства. Вођени овим идејама, осмишљени су алгоритми који приближавају способности машина карактеристикама човека. Велике компаније и организације као што су Google, FedEx, NASA или BMW увелико користе овакве програме. У инжењерству неуралне мреже имају веома велики значај и пружају подршку у областима попут контроле лета, вожње аутомобила, контроле медицинских система или електрана. Алгоритми постају све једноставнији за имплементацију и примену, што их чини све употребљивијим у свакодневном животу.

На крају рада желео бих да се захвалим мом ментору, Јелени Хаџи-Пурић, на обезбеђивању литературе, уложеном труду и подршци.

Литература

- [1] Alex Smola: Introduction to Machine Learning, Cambridge University Press 2008. Доступно на: <http://alex.smola.org/drafts/thebook.pdf> [Прегледано 20.05.2020.]
- [2] Keunwoo Choi, György Fazekas, Kyunghyun Cho, Mark Sandler: A Tutorial on Deep Learning for Music Information Retrieval, arXiv:1709.04396v2 [cs.CV] 3 May 2018 Доступно на: <https://arxiv.org/pdf/1709.04396.pdf> [Прегледано 20.05.2020.]
- [3] Wikipedia: Machine learning. Доступно на: https://en.wikipedia.org/wiki/Machine_learning [Прегледано 20.05.2020.]
- [4] Wikipedia: Sound. Доступно на: <https://en.wikipedia.org/wiki/Sound> [Прегледано 20.05.2020.]
- [5] Wikipedia: Psychoacoustics. Доступно на: <https://en.wikipedia.org/wiki/Psychoacoustics> [Прегледано 20.05.2020.]
- [6] Wikipedia: Artificial neural network. Доступно на: https://en.wikipedia.org/wiki/Artificial_neural_network [Прегледано 20.05.2020.]
- [7] Wikipedia: Digital audio. Доступно на: https://en.wikipedia.org/wiki/Digital_audio [Прегледано 20.05.2020.]