ԵՐԵՎԱՆԻ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆ

ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԿԻՐԱՌԱԿԱՆ
ՄԱԹԵՄԱՏԻԿԱՅԻ ՖԱԿՈՒԼՏԵՏ

Թվային անալիզի և մաթեմատիկական մոդելավորման ամբիոն

«ԹՎԱՅԻՆ ԱՆԱԼԻԶ ԵՎ ՄԱԹԵՄԱՏԻԿԱԿԱՆ
ՄՈԴԵԼԱՎՈՐՈՒՄ» ԿՐԹԱԿԱՆ ԾՐԱԳԻՐ

ՄԻՆԱՍՅԱՆ ԳԵՎՈՐԳ ՄԱՆՎԵԼԻ

ՄԱԳԻՍՏՐՈՍԱԿԱՆ ԹԵԶ

ՏՐԱՆՍՖԵՐԱՅԻՆ ՈՒՍՈՒՑՄԱՆ ՈՐՈՇԱԿԻ ՄԵԹՈԴԻ ԸՆԴՀԱՆՐԱՑՄԱՆ ՍԽԱԼԱՆՔԻ ԳՆԱՀԱՏՄԱՆ ՄԱՍԻՆ

«Ինֆորմատիկա և կիրառական մաթեմատիկա» մասնագիտությամբ ինֆորմատիկայի և կիրառական մաթեմատիկայի մագիստրոսի որակավորման աստիճանի հայցման համար

ԵՐԵՎԱՆ 2019

Ուսանող՝ ______________________________ **Մինասյան Գ.Մ.**

ստորագրություն

Գիտական ղեկավար՝ ֆիզ.-մաթ. գիտությունների թեկնածու

______________________________ **Դանոյան Հ.Է.**

ստորագրություն

«Թույլատրել պաշտպանության»

Ամբիոնի վարիչ՝ ֆիզ.-մաթ. գիտությունների դոկտոր, պրոֆեսոր

______________________________ **Հակոբյան Յու.Ռ.**

ստորագրություն

31 մայիսի 2019 թ.

Համառոտագիր

Տրանսֆերային ուսուցման որոշակի մեթոդի ընդհանրացման սխալանքի գնահատման մասին

On Generalization Error Bound Estimation of Certain Transform Learning Method

Մեքենայական ուսուցման դասական մոդելում ենթադրվում է, որ օրինակները տրված են ըստ որևէ ֆիքսված բայց անհայտ հավանականային բաշխումից՝ մեկը մյուսից անկախ: Իրական կյանքին առնչվող բազմաթիվ դրվագներում (որոնցից են բիոինֆորմատիկան, ռոբոտաշինությունը և այլն) բավարար ճշգրտություն ապահովող քանակով օրինակների ձեռքբերումը կարող է լինել ոչ արդյունավետ, երբեմն նաև անհնար: Նման դեպքերում խնդիր է առաջանում օգտագործելով այլ խնդիրներում կուտակված փորձը (օրինակներ, մոդել) կիրառելով քիչ օրինակներով տրված խնդրի համար կառուցել բավարար ճշգրտություն ունեցող մոդել: Մեքենայական ուսուցման այս ոլորտը հայտնի է տրանսֆերային ուսուցում անվամբ:

Աշխատանքը վերաբերվում է տրանսֆերային ուսուցման առկա մոտեցումներից մեկի՝ նեյրոնային ցանցի շերտերի միջոցով տվյալների ներկայացմանը (fine tuning) մաթեմատիկական վերլուծությանը: Մասնավորապես, աշխատանքում ստացվել է համեմատաբար փոքր քանակով օրինակներով խնդրի և նախկին փորձի հիման վրա կառուցված մոդելի ընդհանրացման սխալանքների կապն արտահայտող գնահատական: Աշխատանքում ստացված արդյունքը թույլ է տալիս մաթեմատիկորեն մեկնաբանել նշված տրանսֆերային ուսուցման մեթոդի պրակտիկ արդյունավետության տիրույթը:

Բովանդակություն

Ներածություն

Վերջին տարիներին մեքենայական ուսուցման հատկապես «խորը» ուսուցման ալգորիթմների օգտագործմամբ բազմաթիվ կիրառական խնդիրներ հնարավոր է լուծել բարձր արդյունավետությամբ: Խորը ուսուցման ալգորիթմները հնարավորություն են տալիս ոչ կառուցվածքային տվյալների համար նոր ներկայացումներ ստանալ և փորձնական ճանապարհով պարզվում է, որ այդ ներկայացումները նմանատիպ առաջադրանքներում օգտագործելու դեպքում բարելավվում է լուծման արդյունավետությունը: Փոխարինելով տվյալի x օրինակը $f(x)$ ներկայացման նկարագիրների վեկտորով՝ կրճատվում է պիտակավորված տվյալների անհրաժեշտությունը նոր դասակարգման առաջադրանքներում: Հաշվի առնելով որ բազմաթիվ խնդիրներում պիտակավորված տվյալների հավաքագրումը դժվար է և ծախսատար, շատ կարևոր է լուծման ցանկալի ճշգրտությունն ապահովել՝ օգտագործելով ավելի քիչ պիտակավորված տվյալներ: Անբավարար տվյալների դեպքում մեքենայական ուսուցման հիմնական խնդիրները լուծելու համար «տրանսֆերային» ուսուցումը շատ կարևոր գործիք է:

Նկարների համար լայնորեն օգտագործվող ներկայցումների ստացումը իրականացվում է խորը փաթույթային նեյրոնային ցանցերի միջոցով, որոնց ճարտարապետությունը մշակվել է՝ վարժեցնելով միլլիոնից ավել նկարների վրա: Ներկայացումները ստացվում են՝ վերցնելով նեյրոնային ցանցի նախավերջին շերտի ներքին ներկայացումները, որոնց օգտագործումը էապես լավացնում է նոր դասակարգման խնդրի ճշգրտությունը և կրճատում պիտակավորված տվյալների անհրաժեշտությունը: Մշակված հայտնի ճարտարապետություն ունեցող ցանցեր են $AlexNet$-ը, VGG-ն և այլն, իսկ ավելի խորը ցանցերի օրինակներ են $Inception$ և $ResNet$ տեսակի ճարտարապետությունները:

Բնական լեզվի մշակման մեջ «տրանսֆերային» ուսուցումը իրականացվում է բառերի, նախադասությունների կամ տեքստերի ներդրված վեկտրոնների միջոցով, որոնց ստացումը հիմնականում իրականացվում է ոչ վերահսկվող ուսուցման ալգորիթմների միջոցով: Բառերի ներդրված վեկտորների առավել հայտնի մեթոդներ են $word2vec$ մեթոդների ընտանիքը, $glove$ մատրիցային վերլուծության վրա հիմնված մեթոդը և այլն:

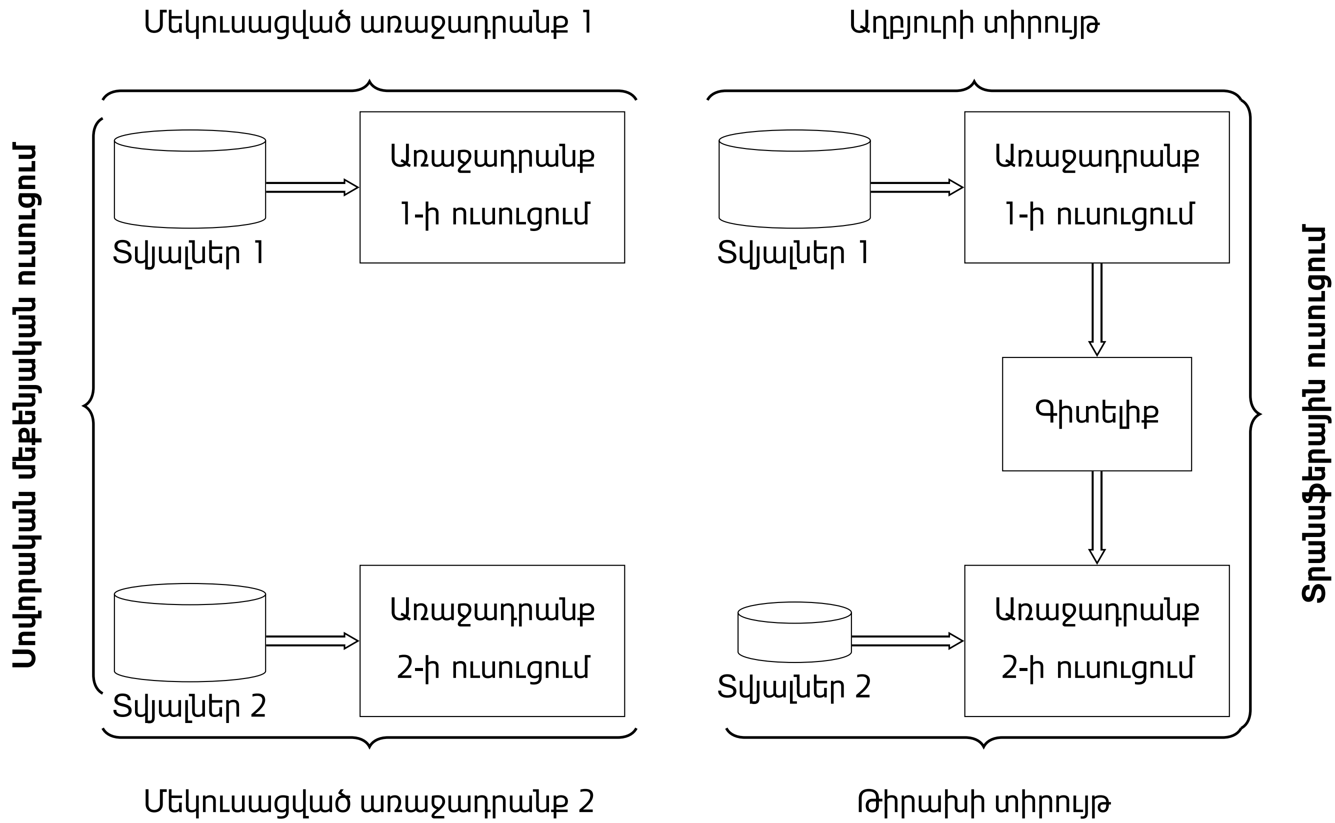
Վարժեցված նեյրոնային ցանցի ներկայացումների օգտագործումը այլ առաջադրանքներում տրանսֆերային ուսուցման մեթոդ է: Սույն մագիստրոսական աշխատանքում տեսականորեն կհիմնավորենք, թե ինչու է փորձնական ճանապարհով ստացվող ներկայացումները բարելավվում նոր դասակարգման առաջադրանքների արդյունավետությունը:

Տրանսֆերային ուսուցում

Մեքենայական ուսուցման բազմաթիվ մեթոդներում կատարվում է մի ընդհանուր ենթադրություն, որ ուսուցման և փորձարկվող տվյալները ընտրվում են միևնույն առանձնահատկությունների տարածությունից, անկախ և միևնույն բաշխումից: Երբ բաշխումը փոխվում է, վիճակագրական և մեքենայական ուսուցման մոդելների մեծ մասը պետք է սկզբից վերակառուցել՝ օգտագործելով նոր հավաքագրված ուսուցման տվյալներ: Բազմաթիվ կիրառական խնդիրներում անհրաժեշտ ուսուցման տվյալների հավաքագրումը անհնար է կամ ծախսատար: Խնդիրների որոշ տիրույթներ ինչպիսիք են բիոինֆորմատիկան և ռոբոտաշինությունը, տվյալների ձեռքբերման և անոտացման ծախսատար լինելու պատճառով՝ մեծ քանակությամբ անոտացված տվյալներ հավաքագրելը շատ բարդ է և թանկարժեք, որն էլ սահմանափակում է այդ տիրույթների զարգացումը: Շատ կարևոր է կրճատել ուսուցման տվյալների հավաքագրման անհրաժեշտությունը խնդիրների այնպիսի տիրույթներում, որտեղ բավարար քանակությամբ ուսուցման տվյալներ չկան: Անբավարար տվյալների դեպքում մեքենայական ուսուցման հիմնական խնդիրները լուծելու համար տրանսֆերային ուսուցումը շատ կարևոր գործիք է:

Մարդիկ հատուկ կարողություն ունեն տարբեր առաջադրանքներում օգտագործած գիտելիքները և հմտությունները կիրառել նաև այլ առաջադրանքներում: Ինչ-որ առաջադրանքի կատարումը սովորելու ընթացքում ձեռք բերված գիտելիքները մենք օգտագործում ենք նմանատիպ առաջադրանքներ լուծելու ժամանակ: Որքան իրար հետ կապված և նման են առաջադրանքները, մեզ համար այնքան հեշտ է գիտելիքների փոխանցումը առաջադրանքների միջև: Նոր առաջադրանքներ կատարելու համար, մենք այն չենք սովորում ամենասկզբից, այլ անցյալում ունեցած մեր գիտելիքները և հմտությունները օգտագործում ենք այդ առաջադրանքում:

Բազմաթիվ մեքենայական ուսուցման և խորը ուսուցման ալգորիթմները նախատեսված են մեկուսացված առաջադրանքների մոդելավորման համար: Տրանսֆերային ուսուցումը առաջարկում է նոր մոտեցում ըստ որի մեկուսացված առաջադրաքների ուսուցումը փոխարինվում է հետևյալ կերպ՝ արդեն իսկ սովորած ինչ-որ առաջադրանքից ձեռք բերված գիտելիքները օգտագործել առնչվող այլ առաջադրանքներում: Այսպիսով տրասֆերային ուսուցումը փորձում է խնդրի կամ առաջադրանքի մի տիրույթի՝ այսպես կոչված աղբյուր տիրույթից, գիտելիքն ու փորձը փոխանցել դեպի այլ առաջադրանքի թիրախային տիրույթ: Սովորական մեքենայական ուսուցման և տրանսֆերային ուսուցման միջև տարբերությունները պատկերված են նկար 2-ում:



Նկ. 1: Տրանֆերային ուսուցման և սովորական մեքենայական ուսուցման միջև տարբերությունը

Տրանսֆերային ուսուցման տեսակները

Տրանսֆերային ուսուցման տեսակները և մեթոդները բազմաթիվ են (տես օրինակ [2–4]), որոնք կարելի կիրառել կախված առաջադրանքից և տվյալների հասանելիությունից: Տրանսֆերային ուսուցման մեթոդները կարելի է դասակարգել ըստ սովորական մեքենայական ուսուցման ալգորիթմի տեսակների՝

Ինդուկտիվ տրանսֆերային ուսուցում: Այս դեպքում աղբյուրի և թիրախի տիրույթները նույն են սակայն տարբեր են առաջադրանքները և թիրախ տիյույթի առաջադրանքները վերահսկվող են: Այս տեսակին պատկանող ալգորիթմները փորձում են աղբյուր տիրույթի համար կատարված ենթադրությունները օգտագործելով թիրախի առաջադրանքի լուծման արդյունավետությունը լավացնել: Կախված նրանից, թե աղբյուր տիրույթը պարունակում է պիտակավորված տվյալնել, թե ոչ կարելի ինդուկտիվ տրանսֆերային ուսուցումը համապատասխանաբար բաժանել ենթատեսակների՝ *բազմառաջադրանքային ուսուցում* և *ինքնուրույն ուսուցում*:

Ոչ վերահսկվող տրանսֆերային ուսուցում: Նման է ինդուկտիվ տրանսֆերային ուսուցմանը, սակայն թիրախ տիրույթի առաջադրանքները ոչ վերահսկվող են: Աղբյուրի և թիրախի տիրույթները նույնն են և երկու տիրույթներում էլ պիտակավորված տվյալներ հասանելի չեն:

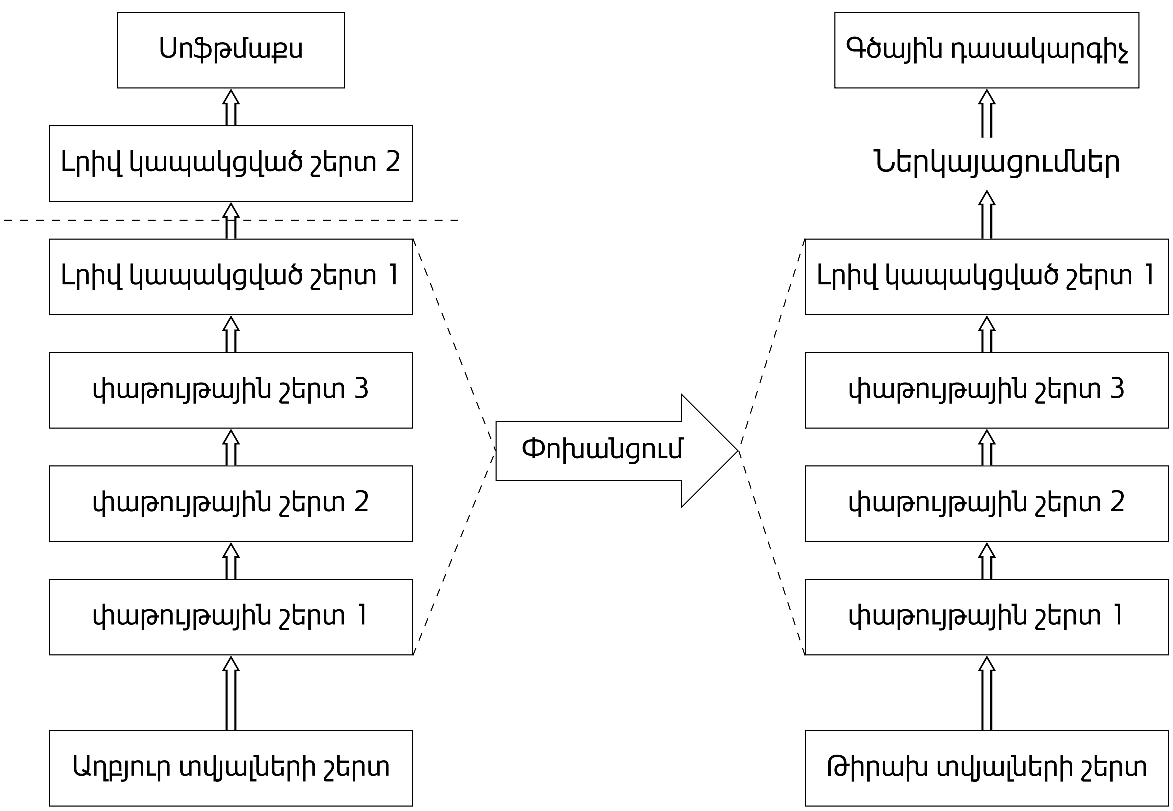
Շարունակական տրանսֆերային ուսուցում: Այս դեպքում աղբյուրի և թիրախի առաջադրանքները նմանություններ ունեն, բայց համապատասխան տիրույթները տարբեր են: Աղբյուրի տիրույթում կա մեծ քանակությամ անոտացված տվյալներ, մինչդեռ թիրախի տիրույթում անոտացված տվյալներ հասանելի չեն:

Տրանսֆերային ուսուցումը խորը ուսուցման համար

Վերջին տարիներին խորը ուսուցման միջոցով կիրառական բազմաթիվ խնդիրներ հնարավոր է բարձր արդյունավետությամբ լուծել: Խորը ուսուցման ալգորիթմները փորձում են զանգածային տվյալներից ավելի բարձր մակարդակի ներկայացումներ սովորել: Ոչ վերահսկվող կամ կիսավերահսվող խորը ուսուցման ալգորիթմները ավտոմատ կերպով տվյալներից նոր և բարձր մակարդակի ներկայացումներ է դուրս հանում: Ի տարբերություն ավանդական մեքենայական ուսուցման մեթոդների, որտեղ տվյալների առանձնահատկությունները ձեռքով են ընտրվում, խորը ուսուցման մեթոդներում տվյալների առաձնահատկությունները ավտոմատ կերպով կառուցվում են ուսուցման ընթացքում: Սակայն ուսուցման ժամանակը և անհրաժեշտ տվյալների քանակը խորը ուսուցման համակարգեր ստեղծելու համար շատ ավելին է՝ համեմատած ավանդական մեքենայական ուսուցման համակարգերի համար: Վերջին տարիներին բնական լեզվի մշակման(տես օրինակ [5, 6]) և պատկերների ճանաչման (տես օրինակ [7, 8]) տարատեսակ խնդիրների համար բարձր արդյունավետությամբ(երբեմն մարդուն գերազանցող) խորը ուսուցման նեյրոնային ցանցեր են մշակվել: Շատ դեպքերում նախապես սովորած նեյոնային ցանցերը օգտագործվում են այլ առաջադրանքներում: Խորը ուսուցման համատեքստում նախապես մշակված ցանցերը կամ մոդելները կազմում են հիմքը խորը փոխանցումային ուսուցման համար: Խորը փոխանցումային ուսուցման մեթոդները մտնում են ինդուկտիվ տրասֆերային ուսուցման մեթոդների մեջ: Այժմ նկարագրենք խորը փոխանցումային ուսուցման ամենահայտնի մեթոդներից մեկը:

Մշակված նեյրոնային ցանցը որպես տվյալների ներկայացման հիմք

Խորը ուսուցման համակարգերը և մոդելները ունեն շերտային կառուցվածքներ, տարբեր շերտեր տվյալների տարբեր ներկայացումներ են սովորում: Վերահսկվող ուսուցման դեպքում վերջին ելքային շերտը, որի նեյրոները քանակը համընկնում է դիտարկվող առաջադրանքում մասնակցող դասերի քանակին և յուրաքանչյուր դասի համապատասխանում է այդ շերտի մեկ նեյրոն, միանում է տվյաների ներկայացման շերտերից վերջինին: Այս շերտային կառուցվածքը հնարավոր է դարձնում է օգտագործել արդեն պատրաստի մշակված ցանցը առանց վերջին շերտի: Նախավերջին շերտը օգտագործվում է որպես տվյալների ներկայացման աղբյուր և օգտագործվում այլ առաջադրանքներում: Սա ամենատարածված մեթոդներից մեկն է տրանսֆերային ուսուցում իրականացնելու համար՝ օգտագործելով խորը նեյրոնային ցանցեր: Փորձնական եղանակով ցույց է տրվել որը պատրաստի խորը նեյրոնայից ցանցերի ներկայացումների միջոցով սահմանափակ տվյալների վրա այլ դասակարգման առաջադրանքներ հնարավոր է լուծել բարձր արդյունավետությամբ (տես օրինակ [9]):



Նկ. 2: Տրանսֆերային ուսուցումը պատրաստի նեյրոնային ցանցի միջոցով

Տրասֆերային ուսուցման որոշակի մեթոդի տեսական երաշխիքներ

$\mathcal{X}$-ով նշանակենք բոլոր հնարավոր տվյալների օրինակները, իսկ $\mathcal{C}$-ով նշանակենք բոլոր պիտակների կամ դասերի բազմությունը: Յուրաքանչյուր $c \in \mathcal{C}$ դասին համապատասխանում է $\mathcal{X}$ բազմության վրա որոշված ինչ-որ $\mathcal{D}_c(x)$ բաշխում, այն ցույց է տալիս, թե x օրինակը ինչքանով է c դասին համապատասխան: Ուսուցումը կատարվում է $\mathcal{F}$ ներկայացումների ֆունկցիաների դասի վրա: $\forall f \in \mathcal{F}$ ֆունկցիա $\mathcal{X}$ տվյաների բազմությունը արտապատկերում d-չափանի էվկլիդյան $\mathcal{R}^d$տարծություն՝ $f : \mathcal{X} \to \mathcal{R}^d$, բացի այդ կդիտարկենք միայն սահմանափակ ֆունկցիաները՝

$$||f(x)|| \leq R \, \forall x \in \mathcal{X} \text{ և } R > 0:$$

Վերահսկվող առաջադրանքներ

Այժմ կնկարագրենք այն առաջադրանքները, որոնց միջոցով փորձարկվելու է ներկայացումների f ֆուկցիան: $k+1$ դասերից բաղկացած $\mathcal{T}$ վերահսկվող առաջադրանքը, բաղկացած է

$$\{c_1, ..., c_{k+1}\} \subseteq \mathcal{C}$$

միմյանցից տարբեր դասերից: Կենթադրենք որ վերահսկվող առաջադրանքները ունեն $\mathcal{P}(\mathcal{T})$ բաշխում, որը բնութագրում է այդ առաջադրանքը դիտարկվելու հավանականությունը: $k+1$ դասերից բաղկացած վերահսկվող առաջադրանքների բաշխումը հետևյալն է՝

$$\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k+1)$$

Պիտակավորված տվյալների բազմությունը $\mathcal{T}$ առաջադրանքի համար բաղկացած է m հատ միմյանցից անկախ և միևնույն բաշխումից ընտրրված օրինակներից: Այդ օրինակները ընտրվում են ստորև նկարագրված պրոցեսով:

$c \in \{c_1, ..., c_{k+1}\}$ դասը ընտրվում է ըստ $\mathcal{D}_{\mathcal{T}}$ բաշխման, որից հետո x օրինակը ընտրվում է $\mathcal{D}_c$ բաշխումից: Դրանք միասին ձևավորում են պիտակավորված (x, c) զույգը, որը ունի հետևյալ բաշխումը՝

$$\mathcal{D}_{\mathcal{T}}(x, c) = \mathcal{D}_c(x)\mathcal{D}_{\mathcal{T}}(c) :$$

Վերահսկվող ներկայացումների գնահատման չափը

f ներկայացումների ֆունկցիաի որակի գնահատումը կատարվում է $\mathcal{T}$ բազմադաս դասակարգման առաջադրանքի միջոցով՝ օգտագործելով գծային դասակարգիչ: Ֆիքսենք $\mathcal{T} = \{c_1, ..., c_{k+1}\}$ առաջադրանքը: $\mathcal{T}$ առաջադրանքի բազմադաս դասակարգիչը ֆուկցիա է՝ $g : \mathcal{X} \to \mathcal{R}^{k+1}$, որի արժեքի կորդինատները ինդեքսավորված են $\mathcal{T}$ առաջադրանքի դասերով: $(x, y) \in \mathcal{X} \times \mathcal{T}$ կետում g դասակարգիչով պայմանավորված կորուստը սահմանենք հետևյալ կերպ՝

$$l(\{g(x)_y - g(x)_{y'}\}_{y \neq y'}),$$

որը ֆունկցիա կախված k չափանի վեկտորից, այն ստացվում է $k+1$ չափանի $g(x)$ վեկտորի կորդինատների տարբերությունից, բացի այդ $\{g(x)_y - g(x)_{y'}\}_{y \neq y'}$ վեկտորի կոմպոնենտները կամայական հերթականությամբ կարելի է համարակալել և l-ի արժեքը կախված չէ վեկտորի կոմպոնենտների համարակալման հերթականությունից: Պրակտիկայում մեծ կիրառություն ունեցող երկու կորուստի ֆունկցիաներ ենք դիտարկելու աշխատանքում՝ ստանդարտ հինջ կորստի ֆունկցիան որը սահմանվում է հետևյալ կերպ՝

$$l(v) = \max\{0, 1 + \max_i\{-v_i\}\}$$

և լոգիստիկ կորստի ֆունկցիան՝

$$l(v) = \log_2(1 + \sum_i e^{-v_i}),$$

որտեղ $v \in \mathcal{R}^k$: $\mathcal{T}$ առաջադրանքի համար g դասակարգիչի կորուստը հետևյալն է՝

$$L(\mathcal{T}, g) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{(x,c) \sim \mathcal{D}_{\mathcal{T}}} [l(\{g(x)_c - g(x)_{c'}\}_{c \neq c'})]$$

f ներկայացումների ֆունկցիան օգտագործելու նպատակով, $g(x) = Wf(x)$ տեսքի դասակարգիչներն ենք դիտարկելու, որտեղ $W \in \mathcal{R}^{(k+1)\times d}$, որը ունի սահմանափակ նորմ՝ $||W||_\infty \leq Q$ և $Q > 0$: $\mathcal{W}$-ով նշանակենք սահմանափակ նորմ ունեցող մատրիցաների բազմությունը՝

$$\mathcal{V} = \{W : ||W||_\infty \leq Q \text{ և } Q > 0\}$$

$\mathcal{T}$ առաջադրանքի համար $g(x) = Wf(x)$ ներկայացումից կախված գծային դասակարգչի կորստի ֆունկցիան հետևյալն է՝

$$L(\mathcal{T}, f, W) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{(x,c)\sim\mathcal{D}_\mathcal{T}} [l(\{Wf(x)_c - Wf(x)_{c'}\}_{c\neq c'})]$$

Ֆիքսելով որևէ f ներկայացում կարելի լավագույն W գտնել, այնպես որ f-ից կախված գծային դասակարգչի կորուստը լինի ամենափոքրը, ուստի f ներկայացման վերահսկիչ կորուստը $\mathcal{T}$ առաջադրանքի համար կսահմանենք, այն կորուստը, երբ լավագույն W ենք ընտրել f-ի համար՝

$$L(\mathcal{T}, f) \stackrel{\text{def}}{=} \inf_{W\in\mathcal{V}} L(\mathcal{T}, f, W)$$

Սահմանում 1 (վերահսկիչ միջին կորուստ). *$k+1$ դասերից բաղկացած առաջադրանքների վերահսկիչ միջին կորուստը f ներկայացման համար սահմանվում է որպես՝*

$$L(f) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{\mathcal{T}\sim\mathcal{P}} [L(\mathcal{T}, f) \mid |\mathcal{T}| = k+1]$$

Սահմանում 2 (էմպիրիկ վերահսկիչ միջին կորուստ). *Դիցուք ունենք միմյանցից անկախ $\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k+1)$ բաշխումից ընտրված N հատ առաջադրանքներ՝ $\mathcal{T}_1, ..., \mathcal{T}_N$: Էմպիրիկ վերահսկիչ միջին կորուստը f ներկայացման համար հետևյալն է՝*

$$\hat{L}(f) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} L(\mathcal{T}_i, f)$$

Ռադեմախերի բարդությունը

$\mathcal{H}$-ով նշանակենք այն ֆունկցիաների բազմությունը, որի վրա իրականացվելու է ուսուցումը՝

$$\forall h \in \mathcal{H}, h : \mathcal{X} \to \mathcal{C}$$

$\mathcal{H}$ ֆունկցիաների բազմությունը կոչվում է հիպոթեզների բազմություն կամ հիպոթեզների դաս: Ուսուցման ալգորիթմը $\mathcal{H}$ ֆունկցիաների բազմությունից, ընտրում է $h \in \mathcal{H}$ հիպոթեզ: Վերջավոր հիպոթեզների համար էմպիրիկ սխալանքի մինիմիզացիայի միջոցով ընտրված հիպոթեզը, ուսուցման էֆեկտիվ ալգորիթմ է (տես օրինակ [10, 11]): Սակայն մեքենայական ուսուցման մեջ օգտագործվող հիպոթեզների բազմությունների հզորությունները սովորաբար անվերջ են: Հիպոթեզների դասերի բարդությունը գնահատող տարբեր մեծություններ կան: Պարզվում է անվերջ հիպոթեզների համար էֆեկտիվ ուսուցման ալգորիթմի գոյությունը կապված է հիպոթեզների բարդությունը գնահատող մեծությունների հետ: Ստորև կներկայացնենք հիպոթեզների բարդությունը գնահատող մեծություններից մեկը՝ *Ռադեմախերի բարդությունը* (տես օրինակ [10, 11]):

Դիցուք $l : \mathcal{H} \times \mathcal{Z} \to \mathbb{R}$ արտապատկերում է, որտեղ $\mathcal{Z} = \mathcal{X} \times \mathcal{C}$: $l(h, z)$-ը ցույց է տալիս h հիպոթեզի կորուստը $z = (x, c)$ պիտակավորված օրինակի համար: Ներմուծենք $\mathcal{G}$ կորստի ֆունկցիաների ընտանիքը $\mathcal{H}$ հիպոթեզների համար՝

$$\mathcal{G} = \{z \mapsto l(h, z) | h \in \mathcal{H}\}$$

Սակայն Ռադեմախերի բարդության սահմանումները կտանք ավելի ընդհանուր ֆունկցիաների $\mathcal{G}$ դասի համար, որի ֆունկցիաները $\mathcal{Z}$-ը արտապատկերում են դեպի $\mathbb{R}$:

Ռադեմախերի բարդությունը ցույց է տալիս թե ինչքանով է «հարուստ» ֆունկցիաների ընտանիքը՝ չափելով պատահական աղմուկի հետ կորելյացիան: Ստորև ֆորմալ կտանք էմպիրիկ և միջին Ռադեմախերի բարդության սահմանումները:

Սահմանում 3 (Էմպիրիկ Ռադեմախերի բարդություն)**.** *Դիցուք $\mathcal{G}$-ն $\mathcal{Z}$-ից դեպի $[a, b]$ հատված արտապատկերող ֆունկցիաների ընտանիք է և*

$$S = \{z_i | z_i \in \mathcal{Z}, \forall i \in [m]\}$$

m հատ ֆիքսված օրինակների բազմություն է: Այդ դեպքում $\mathcal{G}$ ֆունկցիաների ընտանիքի Ռադեմախերի բարդությունը կախված օրինակների S բազմությունից տրվում է հետևյալ կերպ՝

$$\hat{\mathcal{R}}_S(\mathcal{G}) = \frac{1}{m} \mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^m} \left[\sup_{g\in\mathcal{G}} \sum_{i=1}^{m} \sigma_i g(z_i) \right]$$

որտեղ $\sigma = (\sigma_1, ..., \sigma_m)^T$ և σ_i-երը պատահական մեծություններ են, հավասար հավանականությամբ արժեքներ են ընդունում $\{-1, +1\}$-ից: σ_i պատահական մեծությունները կոչվում են Ռադեմախերի փոփոխականներ:

Դիտարկենք g_S-ով նշանակենք այն m չափանի վեկտորը, որի կոմպոնենտները g ֆունկցիայի S բազմության օրինակների վրա ընդունած արժեքներն են՝ $g_S = (g(z_1), ..., g(z_m))^T$: Այժմ էմպիրիկ Ռադեմախերի բարդությունը կարող ենք գրել հետևալ ձևով՝

$$\hat{\mathcal{R}}_S(\mathcal{G}) = \mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^m} \left[\sup_{g\in\mathcal{G}} \frac{\langle\sigma, g_S\rangle}{m} \right]$$

$\langle\sigma, g_S\rangle$ սկալյար արտադրյալը ցույց է տալիս g_S վեկտորի և σ պատահական աղմուկի միջև կորելացիայի չափը: $\sup_{g\in\mathcal{G}} \frac{\langle\sigma, g_S\rangle}{m}$ սուպրեմումը ցույց է տալիս թե ինչքան է $\mathcal{G}$ ֆունկցիաների ընտանիքը S բազմության վրա կորելացված σ-ի հետ: Այսպիսով էմպիրիկ Ռադեմախերի բարդությունը ցույց է տալիս այն միջին չափը, թե ինչքանով է $\mathcal{G}$ ֆունկցիաների ընտանիքը S բազմության վրա կորելացված պատահական աղմուկի հետ: Այս մեծությունը բնութագրում է, թե ինչքան «հարուստ» է $\mathcal{G}$ ֆունկցիաների ընտանիքը, եթե ավելի «հարուստ» կամ բարդ է $\mathcal{G}$ ֆունկցիաների ընտանիքը, ապա ավելի g_S վեկտորներ կստեղծվեն $\mathcal{G}$-ի միջոցով և ավելի մեծ կլինի պատահական աղմուլի հետ միջին կորելացիան:

Սահմանում 4 (Ռադեմախերի բարդություն). *$\mathcal{D}$-ով նշանակենք այն բաշխումը որտեղից S օրինակները գեներացվում են: Կամայական m բնական թվի համար $\mathcal{G}$ ֆունկցիաների ընտանիքի Ռադեմախերի բարդությունը սահմանվում է որպես էմպիրիկ Ռադեմախերի բարդության մաթսպասում ըստ բլոր հնարավոր m օրինակների $\mathcal{D}^m$ բաշխման՝*

$$\mathcal{R}_m(\mathcal{G}) = \mathop{\mathbb{E}}_{S\sim\mathcal{D}^m} \left[\hat{\mathcal{R}}_S(\mathcal{G}) \right]$$

Այժմ ձևակերպենք Ռադեմախերի բարդության վրա հիմնված ընդհանրական սխալանքի գնահատականը, որը կկիրառենք աշխատանքում:

Թեորեմ 1 ([10]). *Դիցուք $\mathcal{G}$ ֆունկցիաների բազմությունը, որի յուրաքանչյուր ֆունկցիա Z-ը արտապատկերում է $[0,1]$ և $S = \{z_i\}_{i=1}^m$ m հզորությամբ միմյանցից անկախ և միևնույն բաշխումից ընտրված օրինակների բազմություն է: Այդ դեպքում ցանկացած δ դրական թվի համար առնվազն $1-\delta$ հավանականությամբ բոլոր $g \in \mathcal{G}$ ֆունկցիաների համար տեղի ունի հետևյալ անհավասարությունները՝*

$$\mathbb{E}[g(z)] \leq \frac{1}{m}\sum_{i=1}^{m} g(z_i) + 2\mathcal{R}_m(\mathcal{G}) + \sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{2m}} \tag{1}$$

և

$$\mathbb{E}[g(z)] \leq \frac{1}{m}\sum_{i=1}^{m} g(z_i) + 2\hat{\mathcal{R}}_S(\mathcal{G}) + 3\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{2m}} \tag{2}$$

Ներկայացումների $\mathcal{F}$ ֆունկցիաների ընտանիքի բարդության գնահատականի համար կօգտվենք Արորայի և այլոց աշխատանքում տրված սահմանումից: Յուրաքանչյուր f ներկայացման ֆունկցիա արտապատկերում է տվյալների բոլոր հնարավոր օրինակների $\mathcal{X}$ բազմությունը d չափանի էվկլիդյան տարածության մեջ՝ $\mathbb{R}^d$:

Սահմանում 5 (Ներկայացումների էմպիրիկ Ռադեմախերի բարդություն [1]). *Դիցուք $\mathcal{F}$ տվյալների ներկայացումների ֆունկցիաների ընտանիք է՝*

$$\forall f \in \mathcal{F}, f : \mathcal{X} :\to \mathbb{R}^d$$

և S-ը m հզորությամբ տվյալների ֆիքսված օրինակների բազմություն՝

$$S = \{x_i | x_i \in \mathcal{X}, \forall i \in [m]\}$$

Այդ դեպքում ներկայացումների $\mathcal{F}$ ընտանիքի էմպիրիկ Ռադեմախերի բարդությունը ֆիքսված օրինակների S բազմության համար սահմանվում է հետևյալ կերպ՝

$$\hat{\mathcal{R}}_S(\mathcal{F}) = \frac{1}{m} \mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^{md}} \sup_{f \in \mathcal{F}} \sum_{i=1}^{m} \sum_{j=1}^{d} \sigma_{ij} f_j(x_i)$$

Դիտողություն. *Ներկայացումների էմպիրիկ $\hat{\mathcal{R}}_S(\mathcal{F})$ Ռադեմախերի բարդությունը սերտորեն կապված է դասակարգման առաջադրանքներում պիտակավորված օրինակների համար սահմանված Ռադեմախերի բարդության հետ:*

$$\mathcal{G} = \{w^T f(\cdot) | f \in \mathcal{F}, ||w|| \leq 1\}$$

ֆունկցիաների դասը հնարավոր է օգտագործել բինար դասակարգման առաջադրանքի լուծման համար՝ օգտագործելով պիտակավորված տվյալներ: Հեշտությամբ կարելի է ցույց տալ, որ

$$\hat{\mathcal{R}}_S(\mathcal{F}) \leq d\hat{\mathcal{R}}_S(\mathcal{G})$$

որտեղ $\hat{\mathcal{R}}_S(\mathcal{G})$-ը $\mathcal{G}$-ի էմպիրիկ սովորական Ռադեմախերի բարդությունն է S բազմության վրա:

Որոշ անհավասարություններ Ռադեմախերի բարդությունների համար

Լիպշից հակությամբ օժտված ֆունկցիաների համար Ռադեմախերի բարդությունների միջև [12] աշխատանքում ցույց է տրվել մի անհավասարություն, որը կընդհանրացնենք և կօգտագործենք սույն մագիստրոսական աշխատանքում: Այժմ տանք L հաստատունով Լիպշից հատկությամբ օժտված ֆունկցիայի սահմանումը:

Սահմանում 6. *Կասենք $f : \mathbb{R}^n \to \mathbb{R}^m$ արտապատկերումը $L > 0$ հաստատունով Լիպշից հատկությամբ օժտված ֆունկցիա է $S \subseteq \mathbb{R}^n$ բազմության վրա եթե՝*

$$||f(x) - f(y)|| \leq L||x - y|| \, \forall x, y \in S$$

Ստորև ձևակերպենք այսպես կոչված վեկտորի կրճատման անհավասարությունը Ռադեմախերի բարդությունների համար՝

Թեորեմ 2 ([12]). *Դիցուք $\mathcal{X}$-ը որևէ բազմություն է և $(x_1, x_2, ..., x_n) \in X^N$: Տրված է նաև $\mathcal{F}$ ֆունկցիաների բազմություն, որի կամայական $f \in \mathcal{F}$ ֆունկցիա $\mathcal{X}$ բազմությունը արտապատկերում է $\mathbb{R}^d$ էվկլիդյան տարածություն՝ $f : \mathcal{X} \to \mathbb{R}^d$: Դիցուք h_i ֆունկցիաներ* *ունենք որոնք $\mathbb{R}^d$ էվկլիդյան տարածությունը արտապատկերում են իրական թվերի $\mathbb{R}$ տարածություն՝ $h_i : \mathbb{R}^d \to \mathbb{R}$, կամայական $i \in [n]$ համար: Կենթադրենք, որ բոլոր h_i ֆունկցիաները, ինչ-որ L դրական հաստատունով Լիպշից հատկությամբ օժտված ֆունկցիաներ են: Այդ դեպքում տեղի ունի հետևյալ անհավասարությունը՝*

$$\mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^n}\left[\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\sigma_i h_i(f(x_i))\right]\le\sqrt{2}L\mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^{nd}}\left[\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\sum_{j=1}^{d}\sigma_{ij}f_j(x_i)\right]\tag{3}$$

Այժմ ձևակերպենք մի պնդում, որն օգտագործելու ենք ընդհանրացված թեորեմի ապացույցի ընթացքում:

Պնդում 1 ([12]). *Կամայական $v\in\mathbb{R}^d$ վեկտորի համար տեղի ունի հետևյալը՝*

$$||v||\le\sqrt{2}\mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^d}\left|\sum_{i=1}^{d}\sigma_i v_i\right|$$

Թեորեմ 2-ը կարելի է ընդհանրացնել $h_i(v,y)\in\mathbb{R}$ ֆունկցիաների համար, որտեղ $v\in\mathbb{R}^d$, $y\in\mathcal{Y}$ և h_i ֆունկցիաները ըստ v փոփոխականի L հաստատունով Լիպշից հատկությամբ օժտված ֆունկցիաներ են կամայական $y\in\mathcal{Y}$ համար:

Թեորեմ 3. *Դիցուք $\mathcal{X}$-ը և $\mathcal{Y}$-ը որևէ բազմություններ են և $(x_1,x_2,...,x_n)\in X^N$: Տրված է նաև $\mathcal{F}$ ֆունկցիաների բազմություն, որի կամայական $f\in\mathcal{F}$ ֆունկցիա $\mathcal{X}$ բազմությունը արտապատկերում է $\mathbb{R}^d$ էվկլիդյան տարածություն՝ $f:\mathcal{X}\to\mathbb{R}^d$: Դիցուք h_i ֆունկցիաներ ունենք՝*

$$h_i:\mathbb{R}^d\times\mathcal{Y}\to\mathbb{R}$$

կամայական $i\in[n]$ համար: Կենթադրենք, որ բոլոր $h_i(v,y)$ ֆունկցիաները, ինչ-որ L դրական հաստատունով Լիպշից հատկությամբ օժտված ֆունկցիաներ են ըստ v-ի կամայական $y\in\mathcal{Y}$ համար: Այդ դեպքում տեղի ունի հետևյալ անհավասարությունը՝

$$\mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^n}\left[\sup_{\substack{f\in\mathcal{F}\\y\in\mathcal{Y}}}\sum_{i=1}^{n}\sigma_i h_i(f(x_i),y)\right]\le\sqrt{2}L\mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^{nd}}\left[\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\sum_{j=1}^{d}\sigma_{ij}f_j(x_i)\right]\tag{4}$$

Ապացույց. Սկզբում ցույց տանք, որ բոլոր $i\in[n]$-երի համար և կամայական $g:\mathcal{F}\times\mathcal{Y}\to\mathbb{R}$ ֆունկցիոնալի համար տեղի ունի հետևյալ անհավասարությունը՝

$$\mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}}\sup_{\substack{f\in\mathcal{F}\\y\in\mathcal{Y}}}\epsilon h_i(f(x_i),y)+g(f,y)\le\sqrt{2}L\mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d}\sup_{\substack{f\in\mathcal{F}\\y\in\mathcal{Y}}}\sum_{j=1}^{d}\epsilon_j f_j(x_i)+g(f,y)\tag{5}$$

Դիցուք $\delta > 0$ կամայական դրական թիվ է: Այդ դեպքում համաձայն Ռադեմախերի փոփոխականի սահմանաման կունենանք՝

$$2 \mathop{\mathbb{E}}_{\epsilon \sim \{\pm 1\}} \sup_{\substack{f \in \mathcal{F} \\ y \in \mathcal{Y}}} \epsilon h_i(f(x_i), y) - \delta = \sup_{\substack{f, \bar{f} \in \mathcal{F} \\ y \in \mathcal{Y}}} h_i(f(x_i), y) + g(\bar{f}, y) - h_i(\bar{f}(x_i), y) + g(\bar{f}, y) - \delta$$

Օգտվելով սուպրեմումի սահմանումից՝ գոյություն ունեն $f*, \bar{f}^* \in \mathcal{F}$ ֆունկցիաներ, որ տեղի ունի հետևյալը՝

$$\begin{aligned} &\sup_{\substack{f, \bar{f} \in \mathcal{F} \\ y \in \mathcal{Y}}} h_i(f(x_i), y) + g(\bar{f}, y) - h_i(\bar{f}(x_i), y) + g(\bar{f}, y) - \delta \leq \\ &\leq \sup_{y \in \mathcal{Y}} h_i(f^*(x_i), y) - h_i(\bar{f}^*(x_i), y) + g(f^*, y) + g(\bar{f}^*, y) \end{aligned}$$

Օգտագործելով h_i ֆունկցաի Լիպշիցի հատկությամբ օժտված լինելը կունենանք՝

$$\begin{aligned} &\sup_{y \in \mathcal{Y}} h_i(f^*(x_i), y) - h_i(\bar{f}^*(x_i), y) + g(f^*, y) + g(\bar{f}^*, y) \leq \\ &\leq L||f^*(x_i) - \bar{f}^*(x_i)|| + \sup_{y \in \mathcal{Y}} g(f^*, y) + g(\bar{f}^*, y) \end{aligned}$$

Պնդում 1-ը կիրառելով կստանանք՝

$$\begin{aligned} &L||f^*(x_i) - \bar{f}^*(x_i)|| + \sup_{y \in \mathcal{Y}} g(f^*, y) + g(\bar{f}^*, y) \leq \\ &\leq \sqrt{2} L \mathop{\mathbb{E}}_{\epsilon \sim \{\pm 1\}^d} \left| \sum_{j=1}^{d} \epsilon_j (f_j^*(x_i) - \bar{f}_j^*(x_i)) \right| + \sup_{y \in \mathcal{Y}} g(f^*, y) + g(\bar{f}^*, y) \leq \\ &\leq \sqrt{2} L \mathop{\mathbb{E}}_{\epsilon \sim \{\pm 1\}^d} \sup_{f, \bar{f} \in \mathcal{F}} \left| \sum_{j=1}^{d} \epsilon_j f_j(x_i) - \sum_{j=1}^{d} \epsilon_j \bar{f}_j(x_i) \right| + \sup_{y \in \mathcal{Y}} g(f, y) + g(\bar{f}, y) \end{aligned}$$

Դեշտ է նկատել, որ կամայական ֆիքսված ϵ-ի դեպքում

$$\sup_{f, \bar{f} \in \mathcal{F}} \left| \sum_{j=1}^{d} \epsilon_j f_j(x_i) - \sum_{j=1}^{d} \epsilon_j \bar{f}_j(x_i) \right| = \sup_{f, \bar{f} \in \mathcal{F}} \sum_{j=1}^{d} \epsilon_j f_j(x_i) - \sum_{j=1}^{d} \epsilon_j \bar{f}_j(x_i)$$

և քանի որ $\sup_{y \in \mathcal{Y}} g(f, y) + g(\bar{f}, y)$ ինվարիանտ է $f, \bar{f}$ ֆունկցիաների փոփոխման նկատմամբ, կունենանք՝

$$\sqrt{2}L \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \left|\sum_{j=1}^{d} \epsilon_j (f_j^*(x_i) - \bar{f}_j^*(x_i))\right| + \sup_{y\in\mathcal{Y}} g(f^*, y) + g(\bar{f}^*, y) \leq$$

$$\leq \sqrt{2}L \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \sup_{f,\bar{f}\in\mathcal{F}} \sum_{j=1}^{d} \epsilon_j f_j(x_i) - \sum_{j=1}^{d} \epsilon_j \bar{f}_j(x_i) + \sup_{y\in\mathcal{Y}} g(f,y) + g(\bar{f}, y) =$$

$$= \sqrt{2}L \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \sup_{f\in\mathcal{F}} \sum_{j=1}^{d} \epsilon_j f_j(x_i) + \sup_{y\in\mathcal{Y}} g(f,y) + \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \sup_{\bar{f}\in\mathcal{F}} - \sum_{j=1}^{d} \epsilon_j \bar{f}_j(x_i) + \sup_{y\in\mathcal{Y}} g(\bar{f}, y)$$

Հաշվի առնելով Ռադեմախերի ϵ_j փոփոխականների սիմետրիկությունը կստանանք՝

$$\sqrt{2}L \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \sup_{f\in\mathcal{F}} \sum_{j=1}^{d} \epsilon_j f_j(x_i) + \sup_{y\in\mathcal{Y}} g(f,y) + \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \sup_{\bar{f}\in\mathcal{F}} - \sum_{j=1}^{d} \epsilon_j \bar{f}_j(x_i) + \sup_{y\in\mathcal{Y}} g(\bar{f}, y) =$$

$$= 2\left(\sqrt{2}L \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \sup_{f\in\mathcal{F}} \sum_{j=1}^{d} \epsilon_j f_j(x_i) + \sup_{y\in\mathcal{Y}} g(f,y)\right) =$$

$$= 2\left(\sqrt{2}L \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}} \sum_{j=1}^{d} \epsilon_j f_j(x_i) + g(f,y)\right)$$

Այսպիսով կամայական $\delta > 0$ դրական թվի համար՝

$$\mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}} \sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}} \epsilon h_i(f(x_i), y) - \delta \leq \sqrt{2}L \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^d} \sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}} \sum_{j=1}^{d} \epsilon_j f_j(x_i) + g(f,y)$$

Քանի որ վերջինս տեղի ունի ցանկացած δ-ի համար, այստեղից անմիջապես հետևում է 5 անհավասարությունը:

Այժմ ինդուկցիայի միջոցով ցույց տանք, որ ցանկացած $m \in \{0, ..., n\}$ համար տեղի ունի հետևյալ անհավասարությունը:

$$\mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^n} \left[\sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}} \sum_{i=1}^{n} \epsilon_i h_i(f(x_i), y)\right] \leq \sqrt{2}L \mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^{md}} \left[\sup_{f\in\mathcal{F}} \sum_{i=1}^{m}\sum_{j=1}^{d} \sigma_{ij} f_j(x_i)\right] +$$

$$+ \mathop{\mathbb{E}}_{\epsilon\sim\{\pm1\}^{n-m}} \left[\sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}} \sum_{i=m+1}^{n} \epsilon_i h_i(f(x_i), y)\right]$$

4 անհավասարությունը անմիջապես հետևում է՝ վերցնելով $m = n$: Երբ $m = 0$ անհավասարության երկու կողմերում նույն արտահայտությունն է գրված և հետևաբար

տեղի ունի անհավասարությունը: Կատարենք ինդուկցիոն ենթադրություն և համարենք անհավասարությունը տեղի ունի $(m-1)$-ի համար, որտեղ $m \leq n$:

$$\underset{\epsilon\sim\{\pm1\}^n}{\mathbb{E}}\left[\sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}}\sum_{i=1}^{n}\epsilon_i h_i(f(x_i),y)\right] \leq \sqrt{2}L\underset{\sigma\sim\{\pm1\}^{(m-1)d}}{\mathbb{E}}\left[\sup_{f\in\mathcal{F}}\sum_{i=1}^{m-1}\sum_{j=1}^{d}\sigma_{ij}f_j(x_i)\right]+$$

$$+\underset{\epsilon\sim\{\pm1\}^{n-m+1}}{\mathbb{E}}\left[\sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}}\sum_{i=m}^{n}\epsilon_i h_i(f(x_i),y)\right]=$$

$$=\underset{\substack{\epsilon\sim\{\pm1\}^{n-m}\\ \sigma\sim\{\pm1\}^{(m-1)d}}}{\mathbb{E}}\ \underset{\epsilon_m\sim\{\pm1\}}{\mathbb{E}}\left[\sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}}\left(\epsilon_m h_m(f(x_m),y)+\sqrt{2}L\sum_{i=1}^{m-1}\sum_{j=1}^{d}\sigma_{ij}f_j(x_i)+\sum_{i=m+1}^{n}\epsilon_i h_i(f(x_i),y)\right)\right]$$

Սահմանենք

$$g(f,y)=\sqrt{2}L\sum_{i=1}^{m-1}\sum_{j=1}^{d}\sigma_{ij}f_j(x_i)+\sum_{i=m+1}^{n}\epsilon_i h_i(f(x_i),y)$$

և տեղադրելով այն վերջինիս մեջ և օգտագործելով 5 անհավասարությունը կստանանք՝

$$\underset{\substack{\epsilon\sim\{\pm1\}^{n-m}\\ \sigma\sim\{\pm1\}^{(m-1)d}}}{\mathbb{E}}\ \underset{\epsilon_m\sim\{\pm1\}}{\mathbb{E}}\left[\sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}}\left(\epsilon_m h_m(f(x_m),y)+g(f,y)\right)\right]\leq$$

$$\leq\underset{\substack{\epsilon\sim\{\pm1\}^{n-m}\\ \sigma\sim\{\pm1\}^{(m-1)d}}}{\mathbb{E}}\ \underset{\sigma_m\sim\{\pm1\}^d}{\mathbb{E}}\left[\sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}}\left(\sum_{j=1}^{d}\sigma_{mj}f(x_m)+g(f,y)\right)\right]=$$

$$=\sqrt{2}L\underset{\sigma\sim\{\pm1\}^{md}}{\mathbb{E}}\left[\sup_{f\in\mathcal{F}}\sum_{i=1}^{m}\sum_{j=1}^{d}\sigma_{ij}f_j(x_i)\right]+\underset{\epsilon\sim\{\pm1\}^{n-m}}{\mathbb{E}}\left[\sup_{\substack{f\in\mathcal{F}\\ y\in\mathcal{Y}}}\sum_{i=m+1}^{n}\epsilon_i h_i(f(x_i),y)\right]$$

□

Հոֆդինգի անհավասարությունը

Դիցուք ունենք $Z_1, ..., Z_m$ անկախ և միևնույն բաշխման պատահական մեծություններ: Հավանականությունների տեսությունից հայտնի *մեծ թվերի օրենքը* երաշխավորում է, որ եթե m-ը ձգտի անվերջության, ապա այդ պատահական մեծությունների էմպիրիկ միջինը զուգամիտելու է դրանց մաթսպասմանը: Սակայն մեծ թվերի օրենքը ընդամենը ասիմտոտիկ գնահատական է և տրված m օրինակների համար ինֆորմացիա չի տալիս էմպիրիկ միջինի և մաթսպասման միջև տարբերության մոդուլի մասին: Ստորև ձևակերպենք Հոֆդինգի անհավասարությունը, որը տրված m օրինակների համար գնահատական է տալիս էմպիրիկ միջինի և դրանց մաթսպասման միջև հեռավորության մասին: Այս անհավասարությունը կկիրառենք աշխատանքում:

Լեմմա 1 (Հոֆդինգի անհավասարություն [11]). *Դիցուք $Z_1, ..., Z_m$ անկախ և միևնույն բաշխման պատահական մեծություններ են և $\bar{Z} = \frac{1}{m}\sum_{i=1}^{m} Z_i$: Ենթադրենք $\mathbb{E}[\bar{Z}] = \mu$ և յուրաքանչյուր i-ի համար $\mathbb{P}[a \leq Z_i \leq b] = 1$: Այդ դեպքում ցանկացած $\epsilon > 0$ թվի համար տեղի ունի հետևյալը՝*

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^{m} Z_i - \mu > \epsilon\right] \leq e^{\frac{-2m\epsilon^2}{(b-a)^2}}$$

և

$$\mathbb{P}\left[\frac{1}{m}\sum_{i=1}^{m} Z_i - \mu < -\epsilon\right] \leq e^{\frac{-2m\epsilon^2}{(b-a)^2}}$$

Վերահսկիչ միջին կորուստի գնահատականը

Լեմմա 2. *Դիցուք ունենք միմյանցից անկախ $\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k+1)$ բաշխումից ընտրված N հատ առաջադրանքներ՝ $\mathcal{T}_1, ..., \mathcal{T}_N$ և ֆիքսենք կամայական $f \in \mathcal{F}$ ներկայացում: $\hat{L}(f)$ էմպիրիկ վերահսկիչ միջին կորուստն է f ներկայացման համար, իսկ $L(f)$-ը վերահսկիչ միջին կորուստը և դիցուք $|\cup_{i=1}^{N} \mathcal{T}_i| = n$: Այդ դեպքում առնվազն $1-\delta$ հավանականությամբ տեղի ունի հետևյալ անհավասարությունը:*

$$\hat{L}(f) \geq L(f) - B\sqrt{\frac{(k+1)\log\left(\frac{1}{\delta}\right)}{2n}} \tag{6}$$

որտեղ B ինչ-որ դրական հաստատուն է:

Ապացույց. Օգտվելով $L(\mathcal{T}_i, f)$ սահմանումից և օգտագործելով f-ի սահմանափակությունը հեշտ է համոզվել որ գոյություն ունի B դրական թիվ այնպես որ կամայական $i \in [N]$ տեղի ունի հետևալը՝

$$0 \leq L(T_i, f) \leq B$$

Այժմ նկատենք որ Հոֆդինգի լեմմայի պայմանները բավարարված են և օգտվելով այդ լեմմայի անհավասարությունից կունենաք՝

$$\mathbb{P}[\hat{L}(f) - L(f)] < -\epsilon] \leq e^{\frac{-2N\epsilon^2}{B^2}}$$

որտեղից և հավանականության $\mathbb{P}[A] = 1 - \mathbb{P}[\bar{A}]$ հատկությունը օգտագործելով՝

$$\mathbb{P}[\hat{L}(f) - L(f) \geq -\epsilon] \geq 1 - e^{\frac{-2N\epsilon^2}{B^2}}$$

$e^{\frac{-2N\epsilon^2}{B^2}}$ հավասարեցնենք δ-ի՝

$$\delta = e^{\frac{-2N\epsilon^2}{B^2}}$$

և լուծելով այն ϵ-ի նկատմաբ՝ կունենաք հետևյալը՝

$$\epsilon = B\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{2N}}$$

Այսպիսով առնվազն $1 - \delta$ հավանականությամբ տեղի ունի հետևյալ անահավսարությունը՝

$$\hat{L}(f) \geq L(f) - B\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{2N}}$$

Նկատենք որ $n \leq (k+1)N$, որտեղից անմիջապես հետևում է հետևյալ անհավասարությունը՝

$$\sqrt{\frac{k+1}{n}} \geq \sqrt{\frac{1}{N}}$$

Օգտագործելով վերջին անհավասարությունը կունենանք, որ առնվազն $1 - \delta$ հավանականությամբ տեղի ունի

$$\hat{L}(f) \geq L(f) - B\sqrt{\frac{(k+1)\log\left(\frac{1}{\delta}\right)}{2n}}$$

անհավասարությունը: $\square$

Նախապես ներմուծենք որոշակի նշանակումներ, որոնք կօգտագործենք հաջորդ լեմմայի ձևակերպման մեջ և ապացույցի ընթացքում: Դիցուք $\mathcal{T}_1, ..., \mathcal{T}_N$ առաջադրանքները միմյանցից անկախ ընտրված են $\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k+1)$ բաշխումից և $\mathcal{T} = \cup_{i=1}^{N} \mathcal{T}_i$, որի հզորությունը $n+1$ է՝ $|\mathcal{T}| = n+1$: Այժմ ρ_{min}-ով նշանակենք $\mathcal{T}$ առաջադրանքում պարունակող դասերից ամենափոքրի հավանականությունը՝

$$\rho_{min} \stackrel{\text{def}}{=} \min_{c \in \mathcal{T}} D_{\mathcal{T}}(c) \tag{7}$$

$m(c)$-ով նշանակենք այն T_i առաջադրանքների քանակը, որոնցում c դասը մասնակցում է, իսկ m_{max}-ով նշանակենք $m(c)$-ի առավելագույն արժեքը ըստ բոլոր $c \in \mathcal{T}$ դասերի՝

$$m_{max} \stackrel{\text{def}}{=} \max_{c \in T} m(c) \tag{8}$$

Դիտարկենք $(n+1) \times d$ չափանի մատրիցան մատրիցան, որի տողերը ինդեքսավորված են $\mathcal{T}$ առաջադրանքի $c_1, ..., c_n$ դասերով՝

$$W = \begin{bmatrix} — & w_{c_1}^T & — \\ — & w_{c_2}^T & — \\ & \vdots & \\ — & w_{c_{n+1}}^T & — \end{bmatrix}$$

$W_{\mathcal{T}_i}$-ով նշանակենք $(k+1) \times d$ չափանի մատրիցան, որի տողերը կազմած են W մատրիցայի այն դասերի տողերից, որոնք ձևավորում են $c_{i,1}, ..., c_{i,k+1}$ դասերից բաղկացած $\mathcal{T}_i$ առաջադրանքը՝

$$W_{T_i} = \begin{bmatrix} — & w_{c_{i,1}}^T & — \\ — & w_{c_{i,2}}^T & — \\ & \vdots & \\ — & w_{c_{i,k+1}}^T & — \end{bmatrix}$$

Լեմմա 3. *Դիցուք $\mathcal{T}_1, ..., \mathcal{T}_N$ առաջադրանքները միմյանցից անկախ ընտրված են $\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k+1)$ բաշխումից և $\mathcal{T} = \cup_{i=1}^{N} \mathcal{T}_i$, որի հզորությունը $n+1$ է: Այդ դեպքում* *կամայական* $f \in \mathcal{F}$ *ներկայացման և կամայական* $(n+1) \times d$ *չափանի* W *մատրիցայի համար տեղի ունի հետևյալ անհավասարությունը՝*

$$\frac{1}{N}\sum_{i=1}^{N} L(T_i, f, W_{T_i}) \leq \frac{m_{max}}{(n+1)\rho_{min}} L(T, f, W) \tag{9}$$

որտեղ ρ_{max} *և* m_{max} *մեծությունները սահմանված են համապատասխանաբար 7-ի և 8-ի միջոցով:*

Ապացույց. Առաջին հերթին կարելի է հեշտությամբ համոզվել որ դիտարկվող հինջ և լոգիստիկ կորստի ֆունկցիաները բավարարում են հետևյալ հատկությանը՝

$$\forall I \subseteq [t]\; l(\{v_i\}_{i \in I}) \leq l(\{v_i\}_{i \in [t]}) \tag{10}$$

Ենթադրենք որ T առաջադրանքը կազմող դասերը ինչ-որ կերպ համարակալված են և ֆիքսենք այդ $c_1, ..., c_{n+1}$ համարակալումը: Իսկ $\mathcal{T}_i$ առաջադրանքի դասերը համարակալված են հետևյալ կերպ՝ $c_{i,1}, ..., c_{i,k+1}$: $I(\mathcal{T}_i) \subseteq T$ նշանակենք այն ինդեքսների բազմությունը, որոնց համապատասխան դասերը կան նաև $\mathcal{T}_i$ առաջադրանքում: Այժմ $\mathcal{D}_{\mathcal{T}_i}(c_{ij})$ հավանականությունը գնահատենք $\mathcal{D}_{\mathcal{T}}(c_{ij})$ հավանականության միջոցով՝

$$D_{\mathcal{T}_i}(c_{ij}) = \frac{D_{\mathcal{T}}(c_{ij})}{\sum_{j \in I(\mathcal{T}_i)} D_{\mathcal{T}}(c_j)} \leq \frac{D_T(c_{ij})}{|I(\mathcal{T}_i)|\rho_{min}} = \frac{D_T(c_{ij})}{(k+1)\rho_{min}}$$

Համաձայն $L(T, f, W)$-ի կորստի ֆունկցիայի սահմանման՝

$$\begin{aligned} L(T, f, W) &= \mathop{\mathbb{E}}_{(x,c)\sim D_T}\left[l\left(\{(Wf(x))_c - (Wf(x))_{c'}\}_{\substack{c\neq c' \\ c' \in T_i}}\right)\right] = \\ &= \mathop{\mathbb{E}}_{c\sim D_{T_i}(c)}\mathop{\mathbb{E}}_{x\sim D_c(x)}\left[l\left(\{(Wf(x))_c - (Wf(x))_{c'}\}_{\substack{c\neq c' \\ c' \in T}}\right)\right] \end{aligned}$$

Կատարենք նոր նշանակում՝

$$\phi(c, T, f, W) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{x\sim D_c(x)}\left[l\left(\{(Wf(x))_c - (Wf(x))_{c'}\}_{\substack{c\neq c' \\ c' \in T}}\right)\right]$$

Տեղադրելով վերոնշյալ նոր նշանակումը $L(T, f, W)$-ի մաթսպասման մեջ կունենանք՝

$$L(T, f, W) = \mathop{\mathbb{E}}_{c\sim D_T(c)}[\phi(c, T, f, W)]$$

Համանման ձևով $\forall i \in [N]$ համար՝

$$L(T_i, f, W_{T_i}) = \underset{c\sim D_T(c)}{\mathbb{E}} [\phi(c, T_i, f, W_{T_i})]$$

Օգտվելով 10 անհավասարությունից՝ հեշտ է նկատել, որ $\forall i \in [N], \forall f \in \mathcal{F}, \forall W \in \mathbb{R}^{(n+1)\times d}$ և $\forall c \in T_i$ համար տեղի ունի՝

$$\phi(c, T_i, f, W_{T_i}) \leq \phi(c, T, f, W)$$

Միավորելով այս ամենը կունենանք՝

$$\frac{1}{N}\sum_{i=1}^{N} L(\mathcal{T}_i, f, W_{\mathcal{T}_i}) = \frac{1}{N}\sum_{i=1}^{N} \underset{c\sim D_{\mathcal{T}_i}(c)}{\mathbb{E}} [\phi(c, \mathcal{T}_i, f, W_{\mathcal{T}_i})] = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k+1} D_{\mathcal{T}_i}(c_{ij})\phi(c_{ij}, \mathcal{T}_i, f, W_{\mathcal{T}_i}) \leq$$
$$\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k+1} \frac{D_T(c_{ij})}{(k+1)\rho_{min}}\phi(c_{ij}, \mathcal{T}_i, f, W_{\mathcal{T}_i}) \leq \frac{1}{N(k+1)\rho_{min}}\sum_{i=1}^{N}\sum_{j=1}^{k+1} D_T(c_{ij})\phi(c_{ij}, \mathcal{T}, f, W) \leq$$
$$\leq \frac{1}{N(k+1)\rho_{min}}\sum_{i=1}^{n+1} m(c) D_T(c_i)\phi(c_i, \mathcal{T}, f, W) \leq \frac{m_{max}}{N(k+1)\rho_{min}} L(T, f, W) \leq$$
$$\leq \frac{m_{max}}{(n+1)\rho_{min}} L(T, f, W)$$

Հետևաբար կամայական $f \in \mathcal{F}$ և կամայական $W \in \mathbb{R}^{(n+1)\times d}$ մատրիցայի համար տեղի ունի 9 անահավասարությունը և լեմմայի ապացույցը ավարտված է:

□

Դիցուք ունենք միմյանցից անկախ $\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k+1)$ բաշխումից ընտրված N հատ առաջադրանքներ՝ $\mathcal{T}_1, ..., \mathcal{T}_N$ և $\mathcal{T} = \cup_{i=1}^{N} \mathcal{T}_i$: Միավորված առաջադրանքի հզորությունը n է՝ $|\mathcal{T}| = n$: Այժմ ենթադրենք միավորված $\mathcal{T}$ առաջադրանքի համար ունենք միմյանցին ակախ և $D_{\mathcal{T}}$ բաշխումից ընտրված M օրինակներ՝

$$S = \{(x_1, y_1), (x_2, y_2), ..., (x_M, y_M) | x_i \in \mathcal{X}, y_i \in \mathcal{T} \text{ և } i \in [M]\}$$

$\mathcal{T}$ առաջադրանքի համար դիցուք $g(x) = Wf(x)$ գծային դասակարգչն է ըստ $f \in \mathcal{F}$ ներկայացման, որտեղ W-ն $(n+1) \times d$ չափանի մատրիցա է և $W \in \mathcal{V}$: $g(x)$ դասակարգչի էմպիրիկ սխալանքը S բազմության վրա սահմանենք հետևյալ կերպ՝

$$\hat{L}(\mathcal{T}, f, W) = \frac{1}{M}\sum_{i=1}^{M} l(\{(Wf(x_i))_{y_i} - (Wf(x_i))_{y_j}\}_{y_i \neq y_j})$$

Ալգորիթմը որով սովորելու ենք ներկայցման ֆունկցիա $\mathcal{F}$ դասից հետևյալն է՝

$$(\hat{f}, \hat{W}) = \underset{\substack{f \in \mathcal{F} \\ W \in \mathcal{V}}}{\text{argmin}} \, \hat{L}(\mathcal{T}, f, W)$$

որտեղ $\hat{f}$ փնտրվող ներկայացումն է: Այսպիսով ալգորթմը ըստ f ներկայացման և գծային դասկարգիչի W մատրիցայի մինիմիզացնում է $\mathcal{T}$ առաջադրանքի վերահսկիչ էմպիրիկ սխալանքը S օրինակների բազմության վրա:

Մինչև հաջորդ լեմմային անցնելը ներկայցնենք $f : \mathbb{R}^n \to \mathbb{R}^m$ դիֆերենցելի ֆունկցիայի Լիպշիցի հատկությամբ օժտված լինելու համար բավարար պայմանը(ավելի մանրամասն տես օրինակ [13, էջ 60-61]):

$$||f(x) - f(y)|| \leq ||J||_F^* ||x - y||, \forall x, y \in \mathbb{R}^n$$

որտեղ

$$||J||_F^* = \max_{x \in \mathbb{R}^n} ||J||_F$$

և J-ն f(x) ֆունկցիայի Յակոբյանն է: Լիպշից ֆունկցիայի սահմանումը հաշվի առնելով՝

$$L \leq ||J||_F^*$$

Լեմմա 4. *Դիցուք δ-ն կամայական դրական թիվ է: Այդ դեպքում առնվազն $1 - \delta$ հավանականությամբ կամայական $f \in \mathcal{F}$ ներկայացման և կամայական $W \in \mathcal{V}$ մատրիցայի համար տեղի ունի հետևյալ անհավասարությունը՝*

$$L(\mathcal{T}, \hat{f}, \hat{W}) \leq L(\mathcal{T}, f, W) + Gen_M$$

,

Ապացույց. Սահմանենք G ֆունկցիաների բազմությունը հետևյալ կերպ՝

$$G = \left\{ (x, y) \mapsto g_{f,W}(x, y) = \frac{1}{B} l(\{[Wf(x)]_y - [Wf(x)]_{y'}\}_{y \neq y'}) | f \in \mathcal{F}, W \in \mathcal{V} \right\}$$

Վերցնենք $Z = \mathcal{X} \times \mathcal{T}$ և $S = \{z_i = (x_i, y_i)\}_{i=1}^M$, կիրառելով 1 թեորեմը G ֆունկցիաների բազմության համար կունենանք՝

$$\mathbb{E}[g(z)] \leq \frac{1}{m} \sum_{i=1}^{m} g(z_i) + \frac{2}{M} \underset{\sigma \sim \{\pm 1\}^M}{\mathbb{E}} \sup_{\substack{f \in \mathcal{F} \\ W \in \mathcal{V}}} \sum_{i=1}^{M} \sigma_i g_{f,W}(z_i) + 3\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{2m}} \tag{11}$$

Այժմ ցույց տանք, որ ցանկացած $W \in \mathcal{V}$ և $i \in [M]$ համար $h_i(f(x_i), W) = g_{f,W}(z_i)$ ֆունկցիան ըստ $f(x_i)$-ի ինչ-որ L հաստատունով օժտված է Լիպշիցի հատկությամբ: Ներմուծենք $\Phi_y(f(x), W)$ ֆունկցիան, այնպես որ $h_i = \frac{1}{B} l \circ \Phi_{y_i}$: Ֆիքսենք որևէ $y \in \mathcal{T}$ դաս և մնացած n դասերը համարակալենք $\mathcal{T} \setminus \{y\} = \{y'_1, y'_2, ..., y'_n\}$: $\Phi_y : \mathbb{R}^d \times \mathcal{V} \to \mathbb{R}^n$ որի տեսքը հետևյալն է՝

$$\Phi_y(x, W) = (w_y x - w_{y'_i} x)_{i \in [n]}$$

Ըստ x փոփոխականի Φ_y ֆունկցիայի Յակոբյանը նշանակենք J_{Φ_y}-ով:

$$J_{\Phi_y} = \begin{pmatrix} \frac{\partial \Phi_{y1}}{\partial x_1} & \frac{\partial \Phi_{y1}}{\partial x_2} & \cdots & \frac{\partial \Phi_{y1}}{\partial x_d} \\ \frac{\partial \Phi_{y2}}{\partial x_1} & \frac{\partial \Phi_{y2}}{\partial x_2} & \cdots & \frac{\partial \Phi_{y2}}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Phi_{yn}}{\partial x_1} & \frac{\partial \Phi_{yn}}{\partial x_2} & \cdots & \frac{\partial \Phi_{yn}}{\partial x_d} \end{pmatrix} = \begin{pmatrix} w_{y1} - w_{y'_1 1} & w_{y2} - w_{y'_1 2} & \cdots & w_{yd} - w_{y'_1 d} \\ w_{y1} - w_{y'_2 1} & w_{y2} - w_{y'_2 2} & \cdots & w_{yd} - w_{y'_2 d} \\ \vdots & \vdots & \ddots & \vdots \\ w_{y1} - w_{y'_n 1} & w_{y2} - w_{y'_n 2} & \cdots & w_{yd} - w_{y'_n d} \end{pmatrix}$$

$$\begin{aligned} ||J_{\Phi_y}||_F &= \sqrt{\sum_{i=1}^{n} \sum_{k=1}^{d} \left(w_{yk} - w_{y'_i k}\right)^2} = \sqrt{n \sum_{k=1}^{d} w_{yk}^2 - 2 \sum_{i=1}^{n} \sum_{k=1}^{d} w_{yk} w_{y'_i k} + \sum_{i=1}^{n} \sum_{k=1}^{d} w_{y'_i k}^2} \\ &\leq \sqrt{nQ^2 + 2nQ^2 + nQ^2} = 2Q\sqrt{n} \end{aligned}$$

Այսպիսով Φ_y ֆունկցիան ըստ x-ի փոփոխականի $2Q\sqrt{n}$ հաստատունով Լիպշիցի հատկությամբ օժտված ֆունկցիա է և քանի որ l-ը η հաստատունով Լիպշիցի հատկությամբ էր օժտված, ապա կունենաք որ h_i ֆունկցիաները բոլոր $i \in [M]$ համար $\frac{2\eta Q\sqrt{n}}{B}$ հաստատունով ըստ $f(x_i)$-ի Լիպշիցի հատկություն ունի ցանկացած $W \in \mathcal{V}$ մատրիցայի համար:

Նկատենք որ թեորեմ 3-ի պայմանները բավարարված են և կիրառելով այն կունենանք՝

$$\mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^M} \sup_{\substack{f \in \mathcal{F} \\ W \in \mathcal{V}}} \sum_{i=1}^{M} \sigma_i g_{f,W}(z_i) \leq \frac{2\sqrt{2}\eta Q\sqrt{n}}{B} \mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^{Md}} \sup_{f \in \mathcal{F}} \sum_{i=1}^{M} \sum_{j=1}^{d} \sigma_{ij} f(x_i)$$

Վերջինս տեղադրենք 11-ի մեջ և անհավասարության երկու կողմը բազմապատկենք B-ով, ցանկացած $g \in G$ համար կունենանք՝

$$\mathbb{E}[Bg(z)] \leq \frac{1}{M} \sum_{i=1}^{M} Bg(z_i) + \frac{4\sqrt{2}\eta Q\sqrt{n}}{M} \mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^{Md}} \sup_{f \in \mathcal{F}} \sum_{i=1}^{M} \sum_{j=1}^{d} \sigma_{ij} f(x_i) + 3B\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{2M}}$$

որտեղից էլ՝

$$L(\mathcal{T},f,W) \le \hat{L}(\mathcal{T},f,W) + \frac{4\sqrt{2}\eta Q\sqrt{n}}{M} \underset{\sigma\sim\{\pm 1\}^{Md}}{\mathbb{E}} \sup_{f\in\mathcal{F}} \sum_{i=1}^{M}\sum_{j=1}^{d} \sigma_{ij} f(x_i) + 3B\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{2M}} \quad (12)$$

որը տեղի ունի $\forall f \in \mathcal{F}$ և $\forall W \in \mathcal{V}$: Քանի որ 12-ը տեղի ունի $\forall f \in \mathcal{F}$ և $\forall W \in \mathcal{V}$, հետևաբար այն տեղի ունի նաև $\hat{f}$ և $\hat{W}$-ի համար՝

$$L(\mathcal{T},\hat{f},\hat{W}) \le \hat{L}(\mathcal{T},\hat{f},\hat{W}) + \frac{4\sqrt{2}\eta Q\sqrt{n}}{M} \underset{\sigma\sim\{\pm 1\}^{Md}}{\mathbb{E}} \sup_{f\in\mathcal{F}} \sum_{i=1}^{M}\sum_{j=1}^{d} \sigma_{ij} f(x_i) + 3B\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{2M}} \quad (13)$$

Դիցուք $f^*, W^* = \text{argmin}_{f\in\mathcal{F},W\in\mathcal{V}} L(\mathcal{T},f,W)$: Կիրառելով Հոֆդինգի անհավասարությունը առնվազն $1-\frac{\delta}{2}$ հավանականությամբ տեղի ունի հետևյալը՝

$$\hat{L}(\mathcal{T},f,^* W^*) \le L(\mathcal{T},f,^* W^*) + B\sqrt{\frac{\log\frac{2}{\delta}}{2M}}$$

Հաշվի առնելով որ $\hat{L}(\mathcal{T},\hat{f},\hat{W}) \le \hat{L}(\mathcal{T},f,^* W^*)$՝ 13 անհավասարությունը կարող ենք գրել հետևյալ կերպ՝

$$L(\mathcal{T},\hat{f},\hat{W}) \le L(\mathcal{T},f,^* W^*) + \frac{4\sqrt{2}\eta Q\sqrt{n}}{M} \underset{\sigma\sim\{\pm 1\}^{Md}}{\mathbb{E}} \sup_{f\in\mathcal{F}} \sum_{i=1}^{M}\sum_{j=1}^{d} \sigma_{ij} f(x_i) + 4B\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{2M}} \quad (14)$$

Հեշտ է նկատել որ 14 տեղի ունի $\forall f \in \mathcal{F}$ և $\forall W \in \mathcal{V}$ համար՝

$$L(\mathcal{T},\hat{f},\hat{W}) \le L(\mathcal{T},f,W) + \frac{4\sqrt{2}\eta Q\sqrt{n}}{M} \underset{\sigma\sim\{\pm 1\}^{Md}}{\mathbb{E}} \sup_{f\in\mathcal{F}} \sum_{i=1}^{M}\sum_{j=1}^{d} \sigma_{ij} f(x_i) + 4B\sqrt{\frac{\log\left(\frac{2}{\delta}\right)}{2M}} \quad (15)$$

Կիրառելով պատահույթների միավորման բանաձևը 15 տեղի ունի առնվազն $1-\delta$ հավանականությամբ և լեմման ապացուցված է:

□

Թեորեմ 4. *Դիցուք δ կամայական դրական թիվ է, այդ դեպքում առնվակն $1-\delta$ հավանականությամբ տեղի ունի հետևյալ անհավասարությունը՝*

$$L(\hat{f}) \le L(T,f,W) + Gen_{M,n}\ \forall f \in \mathcal{F} \text{ և } \forall W \in \mathcal{V}$$

Որտեղ M ուսուցման օրինակների քանակն է, իսկ N-ը առաջադրանքների քանակը:

Ապացույց. Դիցուք ունենք N հատ միմյանցից անկախ $\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_N$ առաջադրանքները ընտրված

$$\mathcal{P}(\mathcal{T} \mid |\mathcal{T}| = k+1)$$

բաշխումից և $\mathcal{T} = \cup_{i=1}^{N} \mathcal{T}_i$, որի հզորությունը հավասար է $n+1$-ի: Ինչպես նաև ունենք M հատ օրինակների ուսուցման բազմությունը ընտրված $\mathcal{D}_{\mathcal{T}}$ բաշխումից՝

$$S = \{(x_1, y_1), (x_2, y_2), ..., (x_M, y_M) | x_i \in \mathcal{X}, y_i \in \mathcal{T} \, \forall i \in [M]\}$$

Այժմ ֆիքսենք ցանկացած δ դրական թիվ: Օգտվելով լեմմա 4-ից առնվազն $1 - \frac{\delta}{2}$ հավանականությամբ տեղի ունի հետևյալ անհավասարությունը $f \in \mathcal{F}$-ի և $W \in \mathcal{V}$-ի համար՝

$$L(\mathcal{T}, \hat{f}, \hat{W}) \leq L(\mathcal{T}, f, W) + \frac{4\sqrt{2}\eta Q \sqrt{n}}{M} \mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^{Md}} \sup_{f \in \mathcal{F}} \sum_{i=1}^{M} \sum_{j=1}^{d} \sigma_{ij} f(x_i) + 4B \sqrt{\frac{\log\left(\frac{4}{\delta}\right)}{2M}} \quad (16)$$

Ըստ 10 հատկության $\forall i \in [N]$ համար, քանի որ $T_i \subseteq T$, տեղի ունի հետևյալը՝

$$L(\mathcal{T}_i, \hat{f}, \hat{W}_{\mathcal{T}_i}) \leq L(\mathcal{T}, \hat{f}, \hat{W}) \quad (17)$$

որտեղ $\hat{W}_{\mathcal{T}_i}$ կազմված է T_i առաջադրանքում մասնակցող դասերին համապատասխան $\hat{W}$ մատրիցայի տողերից: Համաձայն $L(\mathcal{T}_i, \hat{f})$ սահմանման ակնհայտ է, որ՝

$$L(\mathcal{T}_i, \hat{f}) \leq L(\mathcal{T}_i, \hat{f}, \hat{W}_{\mathcal{T}_i}) \quad (18)$$

Միավորելով 16, 17 և 18-ը $\forall i \in [N]$՝ կունենանք՝

$$L(\mathcal{T}_i, \hat{f}) \leq L(\mathcal{T}, f, W) + \frac{4\sqrt{2}\eta Q \sqrt{n}}{M} \mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^{Md}} \sup_{f \in \mathcal{F}} \sum_{i=1}^{M} \sum_{j=1}^{d} \sigma_{ij} f(x_i) + 4B \sqrt{\frac{\log\left(\frac{4}{\delta}\right)}{2M}} \quad (19)$$

Այժմ 19 անհավասարության աջ և ձախ մասերը ըստ բոլոր i-երի գումարենք և անհավասարության երկու մասերը բաժանենք N-ի կստանանք՝

$$\frac{1}{N} \sum_{i=1}^{N} L(\mathcal{T}_i, \hat{f}) \leq L(\mathcal{T}, f, W) + \frac{4\sqrt{2}\eta Q \sqrt{n}}{M} \mathop{\mathbb{E}}_{\sigma \sim \{\pm 1\}^{Md}} \sup_{f \in \mathcal{F}} \sum_{i=1}^{M} \sum_{j=1}^{d} \sigma_{ij} f(x_i) + 4B \sqrt{\frac{\log\left(\frac{4}{\delta}\right)}{2M}} \quad (20)$$

Նկատենք որ 20 անհավասարության ձախ մասը $\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_N$ առաջադրանքների էմպիրիկ վերահսկիչ միջին կորուստն է $\hat{f}$ ներկայացման համար՝

$$\hat{L}(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L(\mathcal{T}_i, \hat{f}) \tag{21}$$

21-ը տեղադրենք 20-ի մեջ՝

$$\hat{L}(\hat{f}) \leq L(\mathcal{T}, f, W) + \frac{4\sqrt{2}\eta Q\sqrt{n}}{M} \mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^{Md}} \sup_{f\in\mathcal{F}} \sum_{i=1}^{M}\sum_{j=1}^{d} \sigma_{ij} f(x_i) + 4B\sqrt{\frac{\log\left(\frac{4}{\delta}\right)}{2M}} \tag{22}$$

Համաձայն լեմմա 2-ի առնվազն $1 - \frac{\delta}{2}$ հավանականությամբ $\hat{f}$ ներկայացման համար տեղի ունի հետևյալը՝

$$\hat{L}(\hat{f}) \geq L(\hat{f}) - B\sqrt{\frac{(k+1)\log\left(\frac{2}{\delta}\right)}{2(n+1)}} \tag{23}$$

Կիրառելով պատահույթների միավորման բանաձևը՝ 22, 23-ից հետևում է, որ առնվազն $1 - \delta$ հավանականությամբ՝

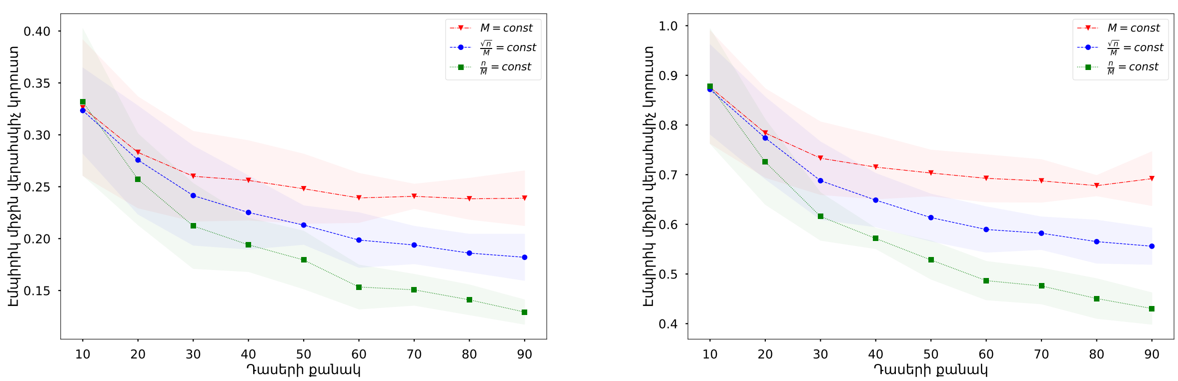
$$L(\hat{f}) \leq L(\mathcal{T}, f, W) + \frac{4\sqrt{2}\eta Q\sqrt{n}}{M} \mathop{\mathbb{E}}_{\sigma\sim\{\pm1\}^{Md}} \sup_{f\in\mathcal{F}} \sum_{i=1}^{M}\sum_{j=1}^{d} \sigma_{ij} f(x_i) + 4B\sqrt{\frac{\log\left(\frac{4}{\delta}\right)}{2M}} + B\sqrt{\frac{(k+1)\log\left(\frac{2}{\delta}\right)}{2(n+1)}}$$

$\square$

Փորձարարական արդյունքներ

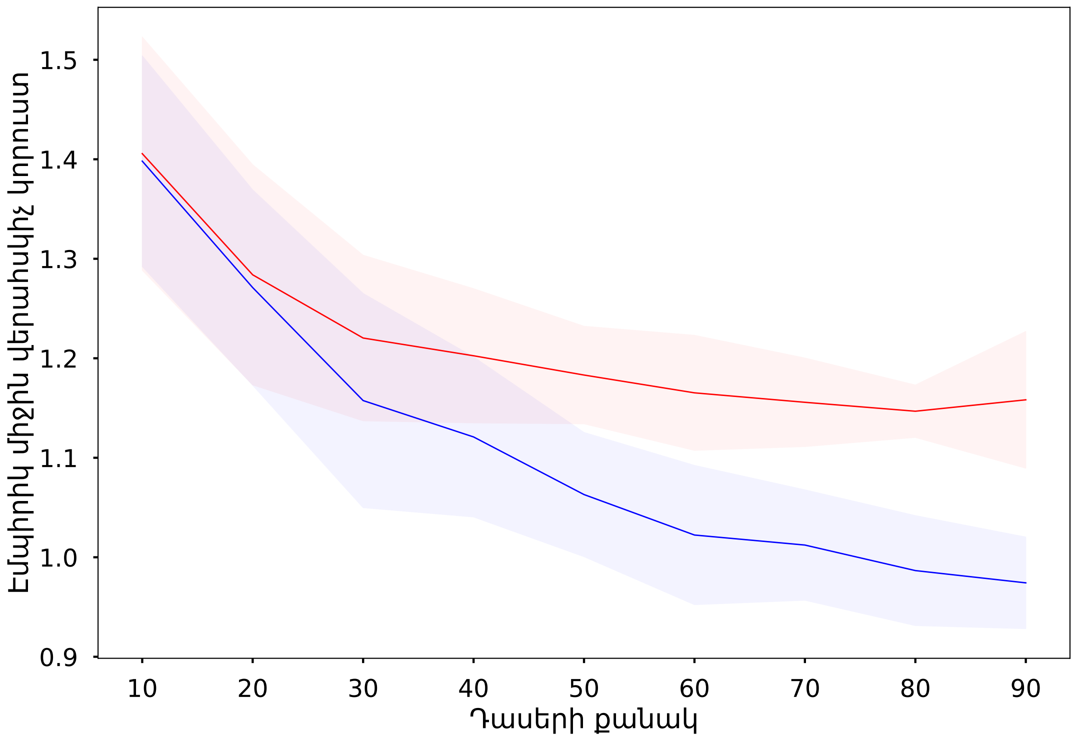
Նկարագրենք այն փորձարկումները, որոնք իրականցրել ենք պարզելու համար, թե արդյոք միջին վերահսկիչ կորստի ընդհանրական սխալանքի գնահատականում մասնակցող քանակական մեծությունների միջև կապը արտահայտվում է իրական տվյալներում: Մասնավորապես կարևոր է հասկանալ տվյալների ներկայցման առաջադրանքում մասնակցող դասերի տարբեր քանակների դեպքում ինչպես է փոխվելու էմպիրիկ միջին վերահսկից կորուստը ուսուցման մեջ չմասնակցող դասերից բաղկացած դասակարգման առաջդրանքների համար:

Փորձարկումների համար օգտագործել ենք $CIFAR-100$ տվյալների բազմությունը (տես [14]): Այն 100 դասերից բաղկացած 60000 հատ 32×32 չափսի փոքր նկարներ պարունակող տվյալների բազմություն է:



Նկ. 3: default

Նախապես պատահականորեն 10 դասեր ընտրվում են բոլոր հնարավոր 100 դասերից, որոնցից կազմվում են $k = 2$, $k = 5$ և $k = 10$ դասեր պարունակող առաջադրանքներ: Մնացած 90 դասերի նկարների օրինակները օգտագործվում են ներկայացումներ սովորելու համար: Ներկայացումների ֆունկցիաների ընտանիքի համար օգտագործվել է ֆիքսված ճարտարապետություն ունեցող փաթույթային ցանցը: Վարժեցումը կատարվում է 90 դասերից պատահական ընտրված առաջադրանքների համար: Հավասարահավանական ընտրված դասերով ձևավորված առաջադրանքները տարբեր քանակությամբ դասեր են պարունակում՝ $n = 10$-ից մինչև $n = 90$: Երբ դասերի քանակը 10 է ուսուցման նկարների քանակը՝ $M = 5000$ և յուրաքանչյուր դասի համար 500 հատ նկար կա տվյալներում: Վերժեցման տվյալների դասերի քանակի՝ n-ի, մեծացումից կախված ուսուցման մեջ մասնակցող նկարաների քանակը՝ M-ը փոփոխվում է երեք տարբեր կախվածությամբ՝ հաստատուն, գծային և $\frac{\sqrt{n}}{M}$-ը պահելով հաստատուն:



Նկ. 4: default

Գրականություն

[1] S. Arora, H. Khandeparkar, M. Khodak, O. Plevrakis and N. Saunshin. *A Theoretical Analysis of Contrastive Unsupervised Representation Learning,* https://arxiv.org/abs/1902.09229, 2019.

[2] S. Pan and Q. Yang. *A survey on transfer learning,* Knowledge and Data Engineering, IEEE Transactions on, 22(10):1345–1359, 2010.

[3] K. Weiss, T. M. Khoshgoftaar, and D. Wang. *A survey of transfer learning,* Journal of Big Data, vol. 9, no. 3, 2016.

[4] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu. *A survey on deep transfer learning,* https://arxiv.org/abs/1808.01974, 2018.

[5] Devlin, J., Chang, M.-W., Lee, K., and Toutanova. *K. Bert: Pretraining of deep bidirectional transformers for language understanding,* arXiv preprint arXiv:1810.04805, 2018.

[6] Logeswaran, L. and Lee, H. *An efficient framework for learning sentence representations,* In Proceedings of the International Conference on Learning Representations, 2018.

[7] K. Simonyan and A. Zisserman. *Very deep convolutional networks for large-scale image recognition,* In ICLR, 2015

[8] K. He, X. Zhang, S. Ren, and J. Sun. *Deep residual learning for image recognition,* In Proceedings of CVPR, pages 770–778, 2016. arxiv.org/abs/1512.03385

[9] A. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. *CNN Features off-the-shelf: an Astounding Baseline for Recognition,* CoRR, abs/1403.6382, 2014.

[10] M. Mohri, A. Rostamizadeh and A. Talwalkar. *Foundations of Machine Learning*, MIT press, 2018.

[11] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms,* Cambridge University Press, New York, NY, USA, 2014.

[12] A. Maurer. *A vector-contraction inequality for rademacher complexities*, In International Conference on Algorithmic Learning Theory, pp. 3–17. Springer, 2016.

[13] J.V. Burke. *Nonlinear Optimization.* University of Washington, 2004.

[14] A. Krizhevsky. *Learning multiple layers of features from tiny images.* Technical report, 2009.