

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Toomas Roosma

**ReLU närvivõrgu ennustusmääramatuse
hindamine regressioonülesannetes**

Bakalaureusetöö (9 EAP)

Juhendaja: Meelis Kull, PhD

Tartu 2023

ReLU närvivõrgu ennustusmääramatuse hindamine regressioonülesannetes

Lühikokkuvõte:

Käesolevas töös uuriti ReLU närvivõrgu võimekust enda ruutviga ennustada regressioonülesandel. Närvivõrkude ennustusmääramatuse teadmine on oluline, sest see võimaldab mõista mudeli ennustuste usaldusväärsust. Katsetati nelja erineva ennustatava funktsiooniga ning mõõdeti närvivõrkude ennustuste täpsust. Töös leiti, et nii piirkonnas leiduvate treeningandmete logaritmi pöördväärtus kui ka murdepunktide kogus on hinnangu andmisel olulised. Ühtlasi leiti, et lähemal asuvad murdepunktide ning treeningandmete kogus on olulisem kui kaugemal asuvate kogus.

Võtmesõnad:

masinõpe, ReLU närvivõrgud, ennustusmääramatus, regressioon

CERCS: P176 Tehisintellekt

Estimating ReLU deep neural networks uncertainty estimations

Abstract:

In this thesis, we investigated the ability of ReLU neural networks to predict its uncertainty estimations. Knowing the uncertainty of predictions is essential because it helps to understand the model's reliability. Four different experiments were conducted, and the accuracy of neural network uncertainty estimations was measured in each. The thesis found that both the reciprocal of the logarithm of the number of training points in a region and the number of breakpoints affect predictions. The closer the training data and breakpoint are located to the predicted value; the more importance they have.

Keywords:

machine learning, ReLU neural networks, predictive uncertainty, regression

CERCS: P176 Artificial intelligence

Sisukord

Sissejuhatus	4
1. Teoreetiline ülevaade	5
1.1 Regressioon	5
1.2 Lineaarne regressioon.....	6
1.3 ReLU närvivõrgud.....	8
1.4 Bayesi-optimaalne mudel	11
1.5 Dekompositsioon.....	11
1.6 Määramatuse modelleerimine	12
2. Metoodika	16
2.1 Treening- ja testandmestikud	16
2.2 Närvivõrgu valik.....	17
2.3 Mõõdetavad näitajad	18
2.4 Tunnuste ettevalmistamine	19
3. Tulemuste analüüs.....	20
3.1 Murdepunktidest ja trendidest	20
3.2 Tulemuste analüüs regressiooniga.....	23
3.2.1 Kõigi tunnustega lineaarregressioon	23
3.2.2 Punktidest konstrueeritud tunnuste kombineerimine	25
3.2.3 Murdepunktidest kombineeritud tunnus	25
3.3 Regressioon kombineeritud tulemustega.....	26
3.4 Mudeli tulemuslikkus mõõdetud vigade hindamisel	28
Kokkuvõte	29
Viidatud kirjandus	30
Lisad	32
I. Koodi repositoorium	32
II. Lisa joonised	33
III. Regressioonide kordajad	35
IV. Litsents	36

Sissejuhatus

Regressioonanalüüs on statistika ja masinõppe meetod, mille eesmärk on modelleerida seoseid kahe või enama muutuja vahel, et paremini mõista nende suhteid ja teha ennustusi. Selleks võib kasutada näiteks tehisnärvivõrke, lineaarset regressiooni või tugivektor-masinaid. Regressiooni näidetena võib tuua kinnisvara hindade ennustamise, kus regressioonimudelid kasutavad mitmeid sisendmuutujaid, nagu eluruumide pindala, asukoht ja muud näitajad, et prognoosida korteri turuväärtust. Teine näide on süsiniku imendumise arvutused metsades, kus regressiooni abil saab hinnata, kui palju süsinikku eri tüüpi puud ja ökosüsteemid aja jooksul õhust ära seovad, võttes arvesse tegureid nagu temperatuur, sademete hulk ning mulla omadused.

Lisaks ennustustele on oluline teada, kui suurt kõrvalekallet on oodata ennustusest. Pärilevivõrkudega on üks võimalus selle hindamiseks dispersiooni väljendava lisaväljundi loomine ja optimeeritava kaofunktsiooni muutmine. See tähendab, et mudel ennustab hoopis tõenäosusjaotust mitte ainult ühte väärtust. Kui närvivõrgud oskaksid seda kvaliteetselt väljastada, aitaks see mõista mudeli ennustuste usaldusväärsust, et paremini hinnata ja hallata riske, mis võivad mudeli ennustuste kasutamisega seotud olla.

Töö eesmärgiks on paremini mõista, millised faktorid mõjutavad regressiooniülesannete puhul ReLU närvivõrgu poolt väljastatud tõenäosusjaotuse dispersiooni vastavust tegelikele vigadele. Uuritav regressioon on ühe sisendtunnusega ning ühtlaselt muutuva müraga.

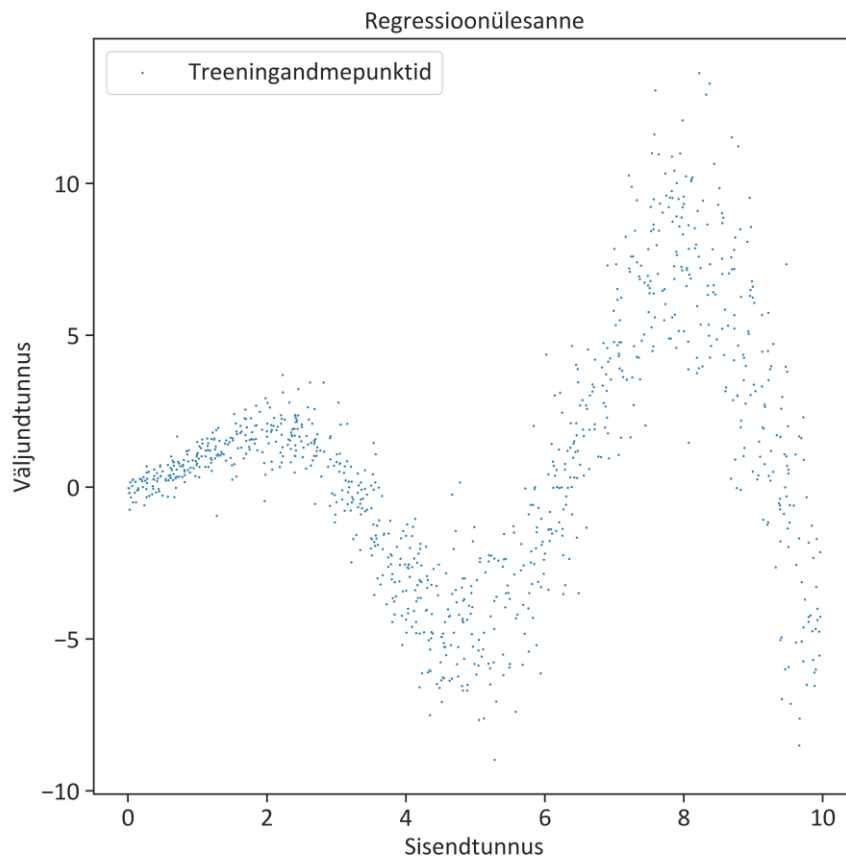
Töö esimeses osas antakse lugejale vajalikud taustateadmised regressioonist, närvivõrkudest ning määramatustest, et töö praktilise osana valminud eksperimente mõista. Töö teises osas kirjeldatakse, milline närvivõrk valiti ning millistel treeningandmetel neid treeniti. Lõpuks analüüsitakse katsete tulemusi lineaarse regressiooni abil ning tõlgendatakse analüüsi tulemusi.

1. Teoreetiline ülevaade

Käesolevas peatükis antakse ülevaade regressioonist ning ReLU närvivõrkudest ning tutvustatakse ReLU närvivõrkude murdepunktide mõistet, mida käesolev töö kasutab, et ennustada närvivõrkude määramatuse hinnanguid. Seejärel kirjeldatakse Bayesi-optimaalset mudelit ning sellega seonduvaid erinevaid määramatusi – aleatoorset ning episteemilist.

1.1 Regressioon

Regressioon on statistiline meetod, mida kasutatakse seoste uurimiseks kahe või enama muutuja vahel. Meetodi eesmärgiks on leida, kuidas üks või mitu sisendtunnust (seletavat muutujat) seletavad väljundtunnuse (sõltuva muutuja) muutumisi. Käesolevas töös kasutatakse regressiooni kahte erinevat tüüpi kahe erineva eesmärgi saavutamiseks: närvivõrgul põhinevat regressioonimudelit ReLU mudelite määramatuse uurimiseks ning lineaarset regressiooni eksperimentide tulemustest seoste leidmiseks. Mõlemal juhul on regressioon ülesanne, mille eesmärgiks on luua mudel, mis ennustab pidevat väljundtunnust, tuginedes ühele või mitmele sisendtunnusele. See erineb klassifikatsiooniülesannetest, mille eesmärk on ennustada kategoorilist tunnust.



Joonis 1. Regressioonülesanne seosega $y = x \sin(x) + \varepsilon$, $\varepsilon \sim N(0, 0,09x^2 + 0,09)$.

Joonisel 1 on kujutatud üks töös kasutusel olevat regressioonülesannet. Joonise x-teljel on sisendtunnus ning y-teljel on väljundtunnus. Sisendtunnuseks on 1000 andmepunkti ühtlasest jaotusest vahemikust $[0, 10)$. Seos on stohhastiline ehk antud sisendi puhul ei ole võimalik üheselt määrata vastavat väljundit. Kui on sisendtunnus X ja väljundtunnus Y , siis tähistatakse konkreetsele sisendile x tehtud ennustust $\hat{y} = \hat{f}(x)$.

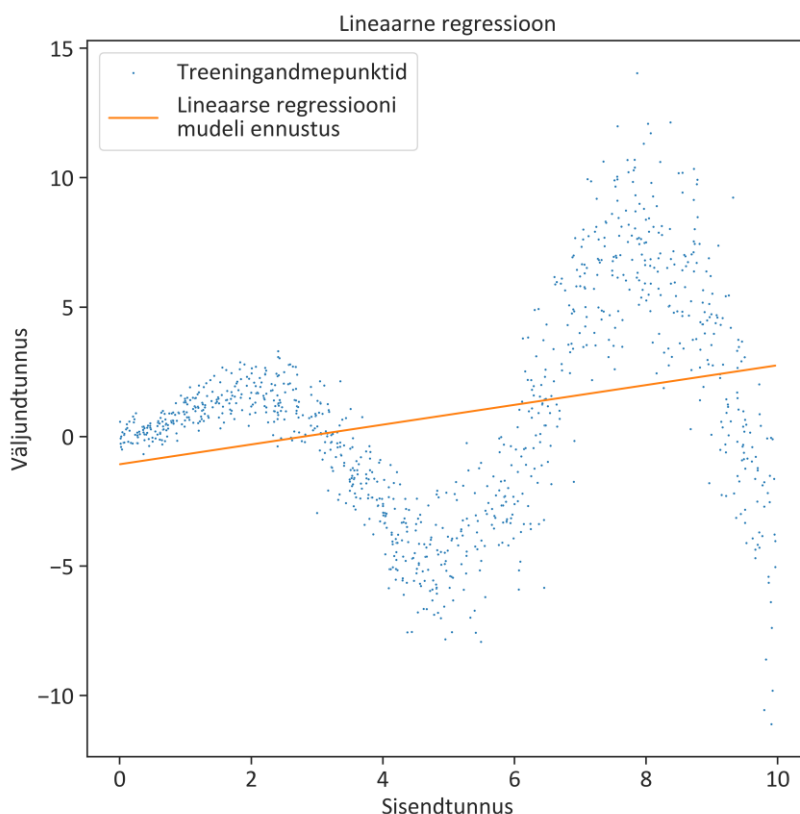
1.2 Lineaarne regressioon

Lineaarne regressioon eeldab, et sõltuv muutuja on sõltumatute muutujate lineaarne kombinatsioon, lisades juhusliku veakomponendi. Üldkujul on lineaarse regressiooni valem järgmine:

$$\hat{y} = \hat{f}(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p,$$

kus $x_1 \dots x_p$ on sisendtunnused ja \hat{y} sisendtunnusele vastav sõltuv muutuja. Lineaarne regressioonimudel minimeerib keskmist ruutviga (ingl *mean squared error* ehk MSE), mis on

defineeritud kui $MSE = \frac{1}{n} \sum_i (\hat{y}_i - y_i)^2$. Muutujad y_i tähistavad treeningandmestiku punkti ning \hat{y}_i ennustust, mida mudel konkreetsel andmepunktil teeb. Leitakse sirge, millel on keskmiselt kõige väikesem keskmine ruutviga treeningandmepunktidega (joonis 2). Tulemusena on parima sobivusega mudel, mis kirjeldab sõltuva ja sõltumatute muutujate vahelisi seoseid. Regressiooni poolt minimeeritavat funktsiooni nimetatakse ka kaofunktsiooniks (ingl *loss function*).



Joonis 2. Lineaarne regressioonimudel, mis ennustab alusfunktsiooni.

Mudeli ennustuse ning väljundtunnuse vahelise seose tugevuse mõõtmiseks kasutatakse determinatsioonikordajat R^2 , mis iseloomustab mudeli kirjeldusvõimet. R^2 on statistiline mõõdik, mille väärtus on vahemikus 0 kuni 1 ning on defineeritud kui:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2},$$

kus y_i tähistab treeningandmepunkti tegelikku väärtust, \hat{y}_i mudeli poolt ennustatud väärtust ning \bar{y} sõltuva muutuja tegelike väärtuste keskmist. Mida lähemal väärtus on ühele, seda

suuremas ulatuses seletatakse sõltuva muutuja variatsiooni sõltumatute muutujate kombinatsiooniga.

Regressioonanalüüsi kasutatakse laialdaselt erinevates valdkondades, näiteks majanduses, meditsiinis ning inseneriteaduses. See võimaldab mõista, kuidas sisendtunnused mõjutavad väljundtunnust ning võimaldab prognoosida väljundtunnust erinevates olukordades. Antud töös kasutatakse lineaarset regressioonimudelit eksperimentide tulemustest seoste leidmiseks.

1.3 ReLU närvivõrgud

Närvivõrgud on masinõppe algoritmid, mis imiteerivad inimajus leiduvate närvirakkude võrgustike tööd. Neid kasutatakse laialdaselt erinevates valdkondades, nagu pildituvastus ja loomuliku keele töötlus. Närvivõrke on eri tüüpi ja struktuuriga, olenevalt nende kasutusalast ja eesmärkidest. Üks närvivõrkude tüüpidest on pärilevivõrgud (ingl *feed-forward network*), mis on kõige lihtsamad ja tavalisemad närvivõrgud, kus info liigub ühesuunaliselt sisendkihist kuni väljundkihini. Pärilevivõrkudel ei ole tagasisidet ega tsüklilisi ühendusi.

Närvivõrkudel on igal neuronil oma aktivatsioonifunktsioon ehk funktsioon, mis määrab, kuidas sisenditest saadakse väljund. Siin töös kasutatakse neuronite aktivatsioonifunktsioonina lõigatud lineaarset funktsiooni¹ (ingl *Rectified Linear Unit* ehk ReLU). Lõigatud lineaarne funktsioon on defineeritud kui $f(x) = \max(0, x)$ (Fukushima, 1975). Nimetame ReLU funktsiooni f aktiivseks siis, kui tema kaalutud sisendite ning vabaliikme summa on positiivne ehk f väljund on samuti positiivne. Vastasel juhul nimetatakse ReLUt passiivseks (Serra, Tjandraatmadja, & Ramalingam, 2018).

Pärilevivõrkusid saab väljendada matemaatilise funktsioonina:

$$\hat{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m, \hat{f} = g \circ f_k \circ \dots \circ f_2 \circ f_1(x),$$

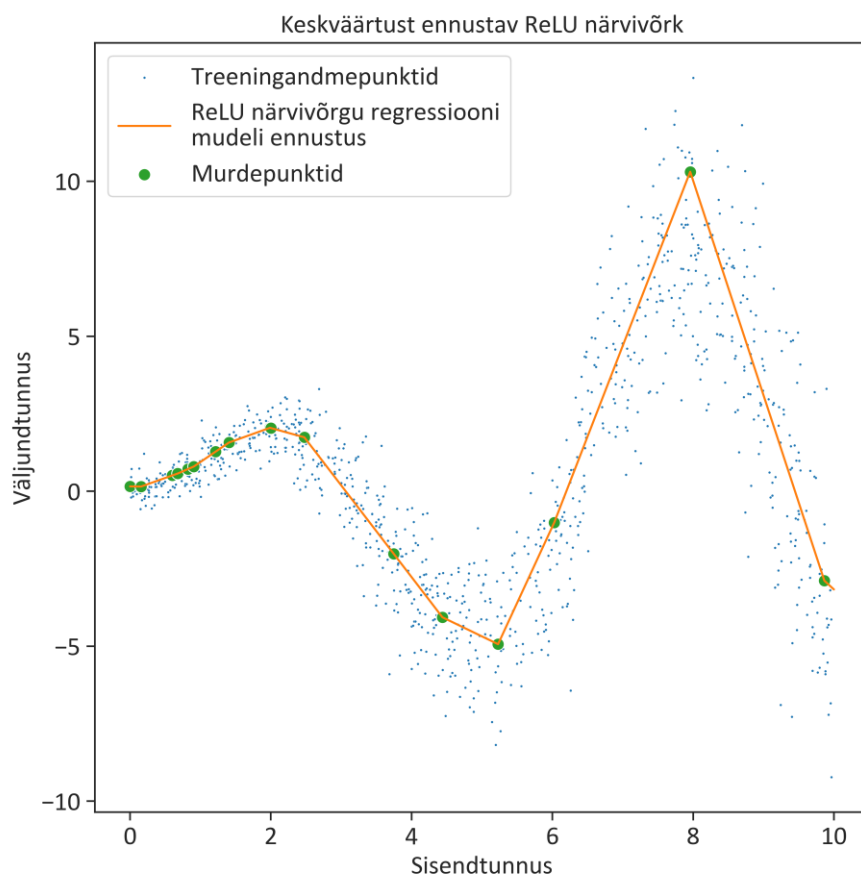
kus n on sisendmuutujate arv, m väljundmuutujate arv, g on väljundkihi funktsioon ning f_i vastava kihi funktsioon:

$$f_i(X) = a(W_i X + B_i),$$

¹ Eestikeelne tõlge on võetud leheküljelt <http://datasci.ee/masinoppe-sonastik/>. Konsensusena on välja toodud mittenegatiivne lineaarfunktsioon, kuid pärast arutelu juhendajaga kasutatakse siin töös edaspidi lühendit ReLU. Alternatiivid on ka parandatud lineaarfunktsioon, tasandatud lineaarfunktsioon.

kus $W_i X + B_i$ on kaalutud sisendi ja vabaliikme summa ning a on aktivatsioonifunktsioon (W_i ja B_i on närvivõrgu i kihi kaalud ja vabaliikmed vektorkujul).

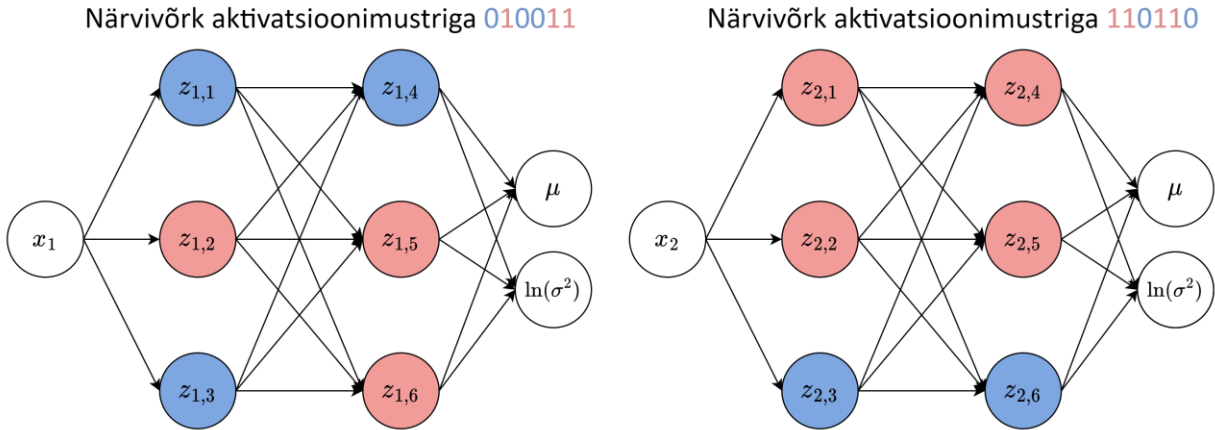
Kuna närvivõrgud on oma neuronite funktsioonide kompositsioon (funktsioonide järjestikune rakendamine), saab nende väljundit vaadelda kui ühelt neuronilt teisele edastatavate signaalide järjestikust teisendamist, kus iga järgmine neuron rakendab oma aktivatsioonifunktsiooni oma kaalutud sisendile. ReLU pärilevivõrkudes on iga neuroni väljund tükati lineaarne, kuna iga neuron rakendab lõigatud lineaarset funktsiooni, mis on tükati lineaarne funktsioon, teisendades positiivsed sisendväärtused lineaarselt ning muutes negatiivsed sisendväärtused nulliks. Kui kaks funktsiooni on lineaarsed, siis on ka nende kompositsioon lineaarne. Seega on ka ReLUdest koosneva närvivõrgu väljund tükati lineaarne. Joonisel 3 on näha ReLU närvivõrgu lahendust regressioonülesandele.



Joonis 3. Peatükis 2.2 kirjeldatav ReLU närvivõrk lahendamas regressiooni ülesannet.

Tükati lineaarsuse uurimiseks eristatakse, millised ReLUD on närvivõrgus aktiivsed ja millised mitte. Eristatakse kahte mõistet, aktivatsioonide muster ning aktivatsioonipiirkond ehk

lineaarne piirkond. Montúfar jt defineerivad need kaks mõistet järgnevalt: aktivatsioonide muster on hulk neuronitest, mis on aktiivsed; ning lineaarne piirkond on tunnusruumi piirkond, millel on sama aktivatsioonimuster (2014). Juhul, kui mõnel hilisemal kihil asuvate neuronite väljund on null, võivad varasemate neuronite aktivatsioonid irrelevantseks muutuda, liites mitu aktivatsioonimustrit üheks lineaarseks piirkonnaks.



Joonis 4 Närvivõrgu erinevad aktivatsioonimustrid

Montúfar (2017) defineerib aktivatsioonimustri järgnevalt. Olgu närvivõrk M , millel on kokku L neuronit. Olgu p aktivatsioonimuster, kus $p = (a_1, \dots, a_j, \dots, a_L)$, ning a_j tähistab j . neuronit. Muutuja $a_j = \text{sgn}(z_j)$ väärtus on 0 parajasti siis kui neuron on passiivne ning 1 kui on aktiivne (Montufar, 2017). Mudeli aktivatsioonimustrite hulk on seega hulk, mis moodustub kõigist saavutatavatest aktivatsioonimustritest. Saavutatavaks aktivatsioonimustriks nimetatakse mustrit, mille jaoks leidub mudelil M sisend x , mille mudelile sisendina andes on mudeli olekuks vastav aktivatsioonimuster. Joonisel 4 on näha närvivõrgu kahte erinevat olekut, millel on erinevad aktivatsioonimustrid. Joonisel tähistab z_{ij} neuroni j väljundit sisendi x_i puhul ning x_1 ja x_2 on kaks erinevat sisendit. Sinisega on tähistatud passiivsed neuronid ehk $z_{ij} = 0$ ning punasega aktiivseid neuroneid ehk $z_{ij} > 0$.

Serra jt defineerivad lineaarsed piirkonnad formaalselt: olgu $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ sügav närvivõrk, siis lineaarne piirkond on sisendite hulk, mis vastab samale aktivatsioonimustrile närvivõrgus (2018). Selle definitsiooni järgi võib üks lineaarne piirkond koosneda sisenditunnuse mittepidevast piirkonnast. Siin töös nimetatakse x -telje punkti, millel toimub üleminek ühest

lineaarsest piirkonnast teisele, murdepunktiks. Murdepunktide vahelised piirkonnad on lineaarsed, mittelineaarsete funktsioonide väljastamine saavutatakse mitmete murdepunktidega (joonis 3).

1.4 Bayesi-optimaalne mudel

Bayesi-optimaalne regressioonmudel on parim teoreetiliselt võimalik mudel f^* , mis on saavutatav konkreetse probleemi puhul. X ja Y on juhuslikud suurused, mis tähistavad vastavalt juhusliku andmepunkti sisendtunnuseid ja märgendit. $E[Y]$ tähistab suuruse Y keskvaärtust ning $E[Y|X]$ on tinglik keskvaärtus. Formaalselt saab seda defineerida kui:

$$f^*(x) = \arg \min_{\hat{y}} E[(Y - \hat{y})^2 | X = x] = E[Y | X = x]$$

Võrduse vasak pool kirjeldab parima võimaliku ennustuse valikut: valitakse punkthaaval, ehk iga x jaoks eraldi parim ennustus, minimeerides eeldatavat ruutviga üle kõigi jaotuste. Regressioonmudelid, mille sihtfunktsiooniks on MSE, on Bayesi-optimaalsed siis, kui nende ennustatud väärtus on võrdne tegeliku tõenäosusjaotuse keskvaärtusega (võrrandi parem pool). Bayesi-optimaalne mudel on kõigist mudelitest minimaalse kaofunktsiooni väärtusega, milleks on $E[(Y - f^*(X))^2]$. Seda taandamatut jääki, nimetatakse müraks ning tähistatakse on $\varepsilon = Y - f^*(X)$, kusjuures $E[\varepsilon | X = x] = 0$.

1.5 Dekompositsioon

Bayesi-optimaalne mudel saavutab regressiooniülesannetel minimaalse võimaliku oodatava MSE, siis mistahes muul mudelil saab oodatav MSE olla ainult suurem. Hastie jt näitavad, et mudeli oodatava MSE saab igas punktis dekomponeerida järgnevalt:

$$E[(\hat{Y} - Y)^2] = E[(\hat{f}(X) - Y)^2] = E[(\hat{f}(X) - f^*(X))^2] + E[(f^*(X) - Y)^2]$$

Kui eelnevas võrrandis fikseerida x , siis saab kirjutada:

$$\begin{aligned} E[(\hat{y} - Y)^2 | X = x] &= E[(\hat{f}(x) - Y)^2 | X = x] \\ &= E[(\hat{f}(x) - f^*(x))^2 | X = x] + E[(f^*(x) - Y)^2 | X = x] \\ &= (\hat{f}(x) - f^*(x))^2 + E[(f^*(x) - Y)^2 | X = x] \end{aligned}$$

Siit on näha, et mitmed ruuterinevused erinevate suuruste (\hat{y}, f^*, Y) vahel on omavahel seotud, näiteks teise võrrandi vasak pool on võrdne peatükis 1.2 defineeritud MSE keskväärtusena (Hastie, Tibshirani, & Friedman, 2009).

1.6 Määramatuse modelleerimine

Tavaliselt ennustab närvivõrk ainult keskväärtust ehk ennustusest ei tea, kui palju keskmiselt ta antud sisendi puhul eksib (joonis 3). Üks võimalus määramatuse modelleerimiseks on lisaks \hat{f} ennustamisele ka ennustada tõenäosusjaotust üle kõigi võimalike väärtuste. Tihti eeldatakse, et tõenäosusjaotuseks on normaaljaotus, ehk Gaussi jaotus. Seda jaotust iseloomustav tihedusfunktsioon pideva suuruse X jaoks on:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

kus μ on jaotuse keskväärtus ning σ standardhälve.

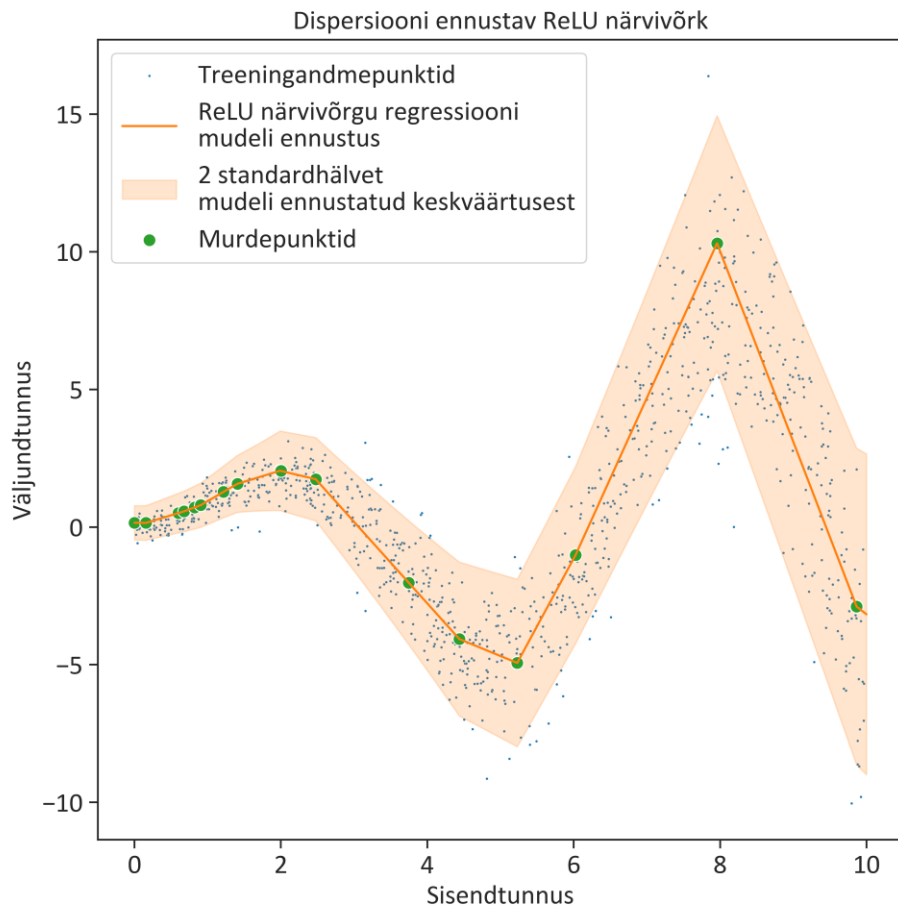
Kui dispersiooni õigesti hinnata, siis ongi dispersioon oodatav ruutviga:

$$E \left[(\hat{f}(x_i) - Y)^2 \mid X = x_i \right],$$

eeldusel, et \hat{f} on keskväärtus. Ehk närvivõrk ennustab igale sisendile väljundmuutuja tõenäosusjaotust, nii keskväärtust kui normaaljaotuse dispersiooni. Kui tavalisel regressioonülesandel saab kasutada keskmist ruutviga kaofunktsioonina, siis tõenäosusjaotuse ennustamiseks saab kasutada negatiivse log-tõepära kaofunktsiooni (ingl *negative log likelihood*). Kendall ja Gal toovad selle valemiks:

$$NLL = \frac{1}{N} \sum_i \log(\sigma(x_i)^2) + \frac{(y_i - \mu(x_i))^2}{\sigma(x_i)^2}$$

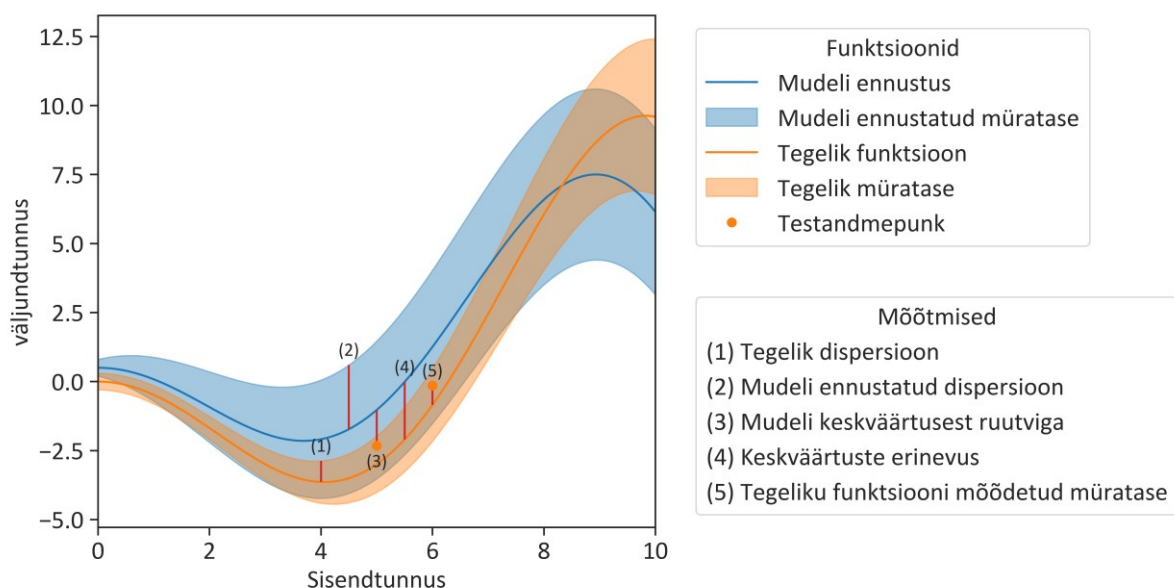
kus, $\sigma(x_i)^2$ on punktis x_i ennustatud dispersioon ning $\mu(x_i)$ on punktis x_i ennustatud keskväärtus. Andmepunkti tegelikku väärtust tähistab y_i (2017). Kaofunktsioon on tuletatud normaaljaotuse tihedusfunktsioonist. Jooniselt 5 on näha, kuidas juba varem kasutusel regressioonülesandes ReLU närvivõrk dispersiooni ennustab.



Joonis 5. ReLU närvivõrgu ennustatud keskväärtus ja määramatus.

Kasutades sünteetilist andmestikku, on võimalik mõõta vahesid sihttunnuse ehk tegeliku märgendi vahel (ehk koos müraga), tegeliku keskväärtuse (ehk ilma mürata) ja ennustatud keskväärtuse vahel. Töös mõõdetakse viit erinevat vahet, mida saab sünteetiliselt andmestikul ning mudeli ennustustel mõõta (joonis 6). Nendeks viieks mõõdetavaks suuruseks on:

1. tegeliku funktsiooni dispersioon.
2. mudeli ennustatav dispersioon.
3. mudeli ennustatava keskväärtuse ja tegeliku märgendi (ingl *ground truth*) vaheline ruutviga.
4. ennustatud keskväärtuse ning tegeliku keskväärtuse vahe ruut.
5. treeningandmetes mõõdetud müratase.



Joonis 6. Viie mõõdetava vahe illustreeriv joonis.

Neljas mõõdetav viga (joonis 6, mõõtmine (4)) on võrdne esimese liidetavaga $(\hat{f}(x) - f^*(x))^2$ ning seda nimetatakse episteemiliseks määramatuseks. Episteemilise määramatuse probleem seisneb selles, et f^* ei ole teada, seega tuleb seda lähendada. Seda viga on võimalik vaid sünteetilistel andmestikel mõõta ning pärismaailmas on see teadmata. Episteemiline määramatus on tingitud puudulikest teadmiste hulgast, sellepärast et see on mudeli ning Bayesi-optimaalse mudeli vahe. Episteemilist määramatust oleks võimalik vähendada, kui oleks rohkem treeningandmeid. Seega nimetatakse seda ka taandatavaks veaks.

Teine liidetav dekompositsioonis on ligikaudu võrdne viienda mõõdetavaga (joonis 6, mõõtmine (5)), ehk see on mürast tingitud viga, ehk taandamatu viga (ingl *irreducible loss*). Taandamatut viga nimetatakse ka aleatooriliseks määramatuseks. Aleatooriline (ehk statistiline) määramatus viitab juhuslikkusele, s.t eksperimendi tulemuste varieeruvusele, mis tuleneb juhuslike faktorite mõjudest. Aleatoorse määramatuse näitena saab tuua mündiviskamise: sellisel viisil andmete genereerimisel on alati juhuslikkuse komponent, mida ei ole võimalik vähendada rohkemate andmete abil (Hüllermeier, 2020: 458). Aleatoorset määramatust ei ole võimalik mõõta pärismaailmas. Üle kõigi Y , on see samaväärne y dispersiooniga, $\frac{1}{n} \sum (y_i^* - y_i)^2$.

Viienda mõõtmise väärtus tuleb otseselt müratasemest (ehk tegelikust dispersioonist). Kõige olulisem on sihttunnuse ja ennustatud keskväärtuse vaheline kaugus (joonis 6 mõõtmine 3), mida närvivõrk üritab ennustada (mõõtmine 2). Mõõtmine 5 tuleneb otseselt müratasemest (mõõtmine 1) ning seda mõõdetakse, et kontrollida, kas juhuslikkus oli ootuspärane.

Närvivõrkude episteemilise määramatuse ennustamine võrgu enda poolt on vähe kasutatud meetod, osaliselt selle tõttu, et on leitud, et tegu pole väga täpse meetodiga (Nix & Weigend, 1994) ning selle parandamiseks on loodud mitmeid meetodeid (Detlefsen, Jørgensen, & Hauberg, 2019). Seevastu põhjused, miks mudeli hinnangud sellele teadmatusele kehvad on, ei ole väga uuritud valdkond. Antud töö uurib seost mudeli määramatuse hinnangu täpsusel ning murdepunktide ja treeningandmete koguse vahel.

2. Metoodika

Töö praktilise osa tegemiseks kasutati Pythoni versiooni 3.9.1. Kasutatud teekideks olid Keras versioon 2.10.0 (Chollet, 2015), statsmodels versioon 0.13.5 (Seabold & Perktold, 2010), NumPy versiooni 1.23.4, pandas versioonil 1.5.1 ja seaborn versioonil 0.12.2. Koodirepositoorium on toodud töö lisas I.

Mudeli määramatuse hindamise uurimiseks on kasutusel kahe peidetud kihiga ReLUde närvivõrk ning viimaseks kihiks on kaks lineaarset neuronit, mis kombineerivad eelnevate kihide ReLUde väljundid kaheks muutujaks μ ja $\log(\sigma^2)$. Edaspidi töös tähistatakse mudeli väljundit μ ja σ^2 , sest eeldatakse, et väljundi teine osa alati eksponentsieeritakse. Närvivõrgu esimene väljundmuutuja on närvivõrgu ennustatav keskväärtus ning teine on ennustatav dispersioon. Närvivõrku treenitakse lähendama lihtsat matemaatilist funktsiooni ning väljastama oma määramatust. Närvivõrgu sisendiks on punkt arvteljel $x \in [0, 10)$ ning kaofunktsiooniks negatiivne log-tõepära. Närvivõrke treeniti mitmel erineva suurusega treeningandmestikel, kus mõnes piirkonnas oli andmeid tihedamalt ja mõnes teises hõredamalt. Kõik mudelid treeniti Tartu Ülikooli teadusarvutuste keskus (High Performance Computing centre).

Kasutatud koodiredaktoris Visual Studio Code oli laiedusena kasutusel GitHubi Github Copilot, mis soovitab suure keelemudeli põhjal konteksti sobituvaid koodiridasid. Kõik ennustused katsetati hoolega läbi enne kasutusele võtmist.

2.1 Treening- ja testandmestikud

Eksperimentideks valiti kaks erinevat f^* ning ε . Esimene f^* valiti konstantne selle jaoks, et kontrollida närvivõrgu võimekust valitud ε ära õppida. Teine f^* valiti sama kui Detlefsen jt oma sünteetiliseks andmestikuks valis (2019). Mõlemad funktsioonid on siin välja toodud:

- $f_1^*(x) = 0$
- $f_2^*(x) = x \cdot \sin(x)$

Sünteetilise müra lisati sihtfunktsioonidele järgnevatest normaalfaotustest:

- $\varepsilon_1 \sim N(0, 0,09x^2 + 0,09)$
- $\varepsilon_2 \sim N(0, 0,09(10 - x)^2 + 0,09)$

Treeningpiirkonnaks valiti piirkond $[0, 10)$, mis jagati viieks võrdseks osaks. Osadeks jagamine oli oluline, et oleks võimalik treenida närvivõrke ebaühtlase tihedusega andmetel,

et uurida, kuidas mõjutab erinev treeningandmepunktide tihedus närvivõrgu ennustatavat dispersiooni. Igas treeningandmestikus oli neljas osas võrdse tihedusega andmepunkte ning viiendas osas oli punktide tihedus kordajaga läbi korrutatud², mille tulemusena oli punktide tihedus selles osas kõrgem või madalam. Treeningandmepunktide kogused³ valiti selliselt, et vähemate andmepunktidega oleks rohkem mõõtmisi.

Igale treeningandmete punktide kogusele seati vastavusse kordaja ning alampiirkond, kus kordajat rakenduti. Seejärel genereeriti igasse piirkonda vajalik kogus treeningandmepunkte. Genereeritud punktid olid reaalarvud, mis olid pärit ühtlasest jaotusest ning kuulusid vahemikku [alampiirkonna algus, alampiirkonna lõpp). Selle tulemusena tekkis $5 \times 5 \times 19$ erinevat X hulka.

Y hulgad genereeriti kombineerides eelnevalt toodud f^* ning ε kokku. Eksperimendid olid:

- $Y_1 = f_1^* + \varepsilon_1$, ehk konstantne funktsioon kasvava müraga.
- $Y_2 = f_1^* + \varepsilon_2$, ehk konstantne funktsioon kahaneva (tagurpidise) müraga.
- $Y_3 = f_2^* + \varepsilon_1$ ehk kasvava müraga siinusfunktsioon (joonis 1).
- $Y_4 = f_2^* + \varepsilon_2$, ehk siinusfunktsioon kahaneva müraga.

Testandmete genereerimisel kasutati ühtlaste vahedega genereeritud punkte, vältimaks juhuslikkust, mis ühtlasest jaotusest valides tekib. Seda juhuslikkust siia töösse ei soovitud selle jaoks, et vähendada faktorite kogust, mis tulemusi mõjutavad. Testandmepunkte oli kokku 10^5 . Igale eksperimendile genereeriti vastavate valemitega Y komponendid juurde.

2.2 Närvivõrgu valik

Närvivõrguks valiti kolmekihiline närvivõrk, kus on kaks peidetud 20 ReLUst koosnevat kihti, ning väljundkihiks on lineaarne kiht, mis väljastab keskväärtust ning dispersiooni. Võrgu suurus valiti piisavalt väike, et hoida kokku arvutustele minevat ressursi, kuid siiski hoides võrgud piisavalt võimsad (ingl *representational power*), et lahendada ülesannet. Närvivõrk on sarnases suuruses kui Ji jt võrk, kes kasutavad samuti kahte peidetud ReLU kihti, kus on 32 neuronit (2022). Närvivõrgud treeniti 1000 epohhi funktsiooni $f_1^*(x)$ puhul ning 3000 epohi $f_2^*(x)$ jaoks, sest leiti, et selle järel on närvivõrgud koondunud.

² Valitud kordajateks olid 0,25, 0,5, 1, 2, 4.

³ Valitud väärtusteks oli 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 100, 150, 200 ning 400.

Detlefsen jt toovad välja mitmeid meetodeid (lokaalsuse proovivõtja (ingl *locality sampler* (*LS*)), keskmise ja dispersiooni jaotus (ingl *mean-variance split* (*MV*)), pöörd-gamma eeljaotus (ingl *inverse-gamma prior* (*IG*)), ekstrapoleeriv arhitektuur (ingl *extrapolating architecture* (*EX*)) ning kombinatsioon neist), kuidas mudeli hinnangulist määramatust paremaks muuta (2019). Selles töös uuritav närvivõrk ei kasuta neid meetodeid.

2.3 Mõõdetavad näitajad

Peatükis 1.5 välja toodud näitajaid mõõdeti iga treenitud närvivõrgu ennustustel ning tree-ningandmestikul, kasutades libiseva keskmise meetodit. Lisaks eelmainitud viiele veale, on olemas ka vead, mida arvutatakse välja vaid viies piirkonnas. Uuritav treeningpiirkond $[0, 10)$ on jagatud viieks võrdseks piirkonnaks $A = [0, 2)$, $B = [2, 4)$ jne. Lisaks eelmainitud näitajatele leiti treeningpiirkonnas ka iga närvivõrgu murdepunktid.

Igas alampiirkonnas on võetud alampiirkonna keskmise mudeli ennustatud määramatus d_{pred} (joonisel 6 on see tähistatud kui viga 2) ning mudeli keskvärtuse ennustuse ning testandmepunktide vahel olev keskmine ruutviga d_{mse} (joonisel 6 on tähistatud kui viga 3).

Nende kahe keskmistatud veast leiti ka nende vahe $A = d_{pred} - d_{mse}$ ning relatiivne viga $R = \frac{A}{d_{mse}}$. Relatiivne viga on see, mida tulemuste peatükis ennustatakse.

A negatiivseid väärtuseid tuleb tõlgendada kui mudeli müra alahindamist ehk mudel arvab, et treeningandmepunktid on lähemal tema ennustatud keskvärtusele kui tegelikult paiknevad.

Negatiivne suhteline viga näitab seda, et mudel alahindab müra ning väärtus $-0,5$ näitab, et mudeli poolt hinnatud müra on kaks korda väiksem kui tegelik müra. Kui suhteline viga on null, siis on mudeli ennustus täpne ning kui positiivne, siis on mudel liiga vähe enesekindel. Suhteline vahe on oluline, et neid vahesid saaks võrrelda kõigis piirkondades omavahel. Kuna kasutatakse heteroskedastiivset müra, siis on servades asuvates piirkondades ennustatavad mürad alati kas suuremad või väiksemad kui keskmistes piirkondades. Heteroskedastiivne müra on müra, mille dispersioon pole konstantne.

2.4 Tunnuste ettevalmistamine

Treeningandmed genereeriti 50 erineva juhusliku seemnega (ingl *random seed*) ning seejärel treenitud mudelitel tehtud mõõtmised keskmistati. Treenitud närvivõrkude uurimisel selgus, et mõned võrgud ei olnud koondunud peale vastavalt 1000 või 3000 epohhi, seega nende mõõtmised eemaldati. Eemaldamise kriteeriumiks valiti üle kolme standardhälbe mõõtmiste keskmisest (vastavas eksperimendis) erinemine.

Kõiki nelja eksperimenti vaadeldi enamasti koos, et tuvastada suurema üldistusvõimega mustreid. Eksperimente eraldi analüüsides võiks tekkida olukord, kus leitakse mõni muster, mis üldistub vaid kasvavale mürale või muster, mis kehtib vaid nulli lähedal.

Analüüsis kasutati lineaarset regressiooni, mille tarbeks on kõigilt treenitud mudelitelt mõõdetud tunnused normaliseeritud kasutades *minmax* meetodit ehk lahutati igast väärtusest miinimumi ning jagati maksimumi ja miinimumi vahega läbi. Seeläbi on kõik tunnused vahemikus nullist üheni. Lisati ka mõned konstrueeritud tunnused, et võimaldada lineaarsel regressioonil leida ka mittelineaarseid seoseid.

Selle jaoks, et lihtsamini analüüsida piirkondasid sõltumatult sellest, kas nad on treeningpiirkonna ääres või mitte, teisendatakse mõõdetud näidikud suhteliseks alampiirkonnaga, mis parasjagu uurimise all on. See tähendab, et igale alampiirkonnale leitakse tema naaberpiirkonnad ning kaugemad piirkonnad (naabrite naabrid). Kui huvialusel piirkonnal on kaks naabrit, võetakse naabrite keskmine väärtus mõõdetava tunnuse väärtuseks, kui üks, siis võetakse vaid see. Kui uuritav alampiirkond on piirkond B, siis tema naaberpiirkonnad on A ja C ning kauged piirkonnad D. Selle tulemusena on võimalik üldistatud kujul rääkida lähedal asuvate piirkondade mõõdikute olulisest ning võrrelda neid kaugemal asuvate piirkondade tulemustega.

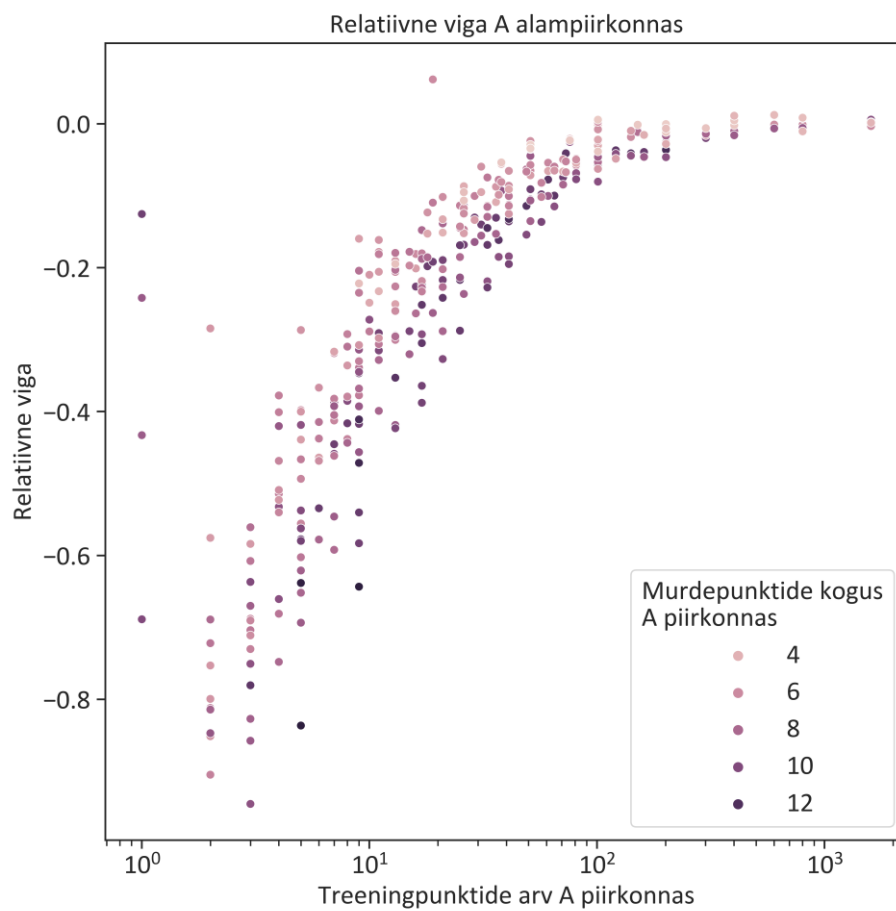
3. Tulemuste analüüs

Eelmises peatükis kirjeldati nelja regressioonülesannet, millel ReLU närvivõrkusid treeniti. Ülesanneteks oli erinevate ühtlaselt muutuvate müradega konstantse ning siinusfunktsiooni õppimine. ReLU närvivõrkusid treeniti väljastama tõenäosusjaotust ning seejärel mõõdeti närvivõrkude ennustatud dispersiooni ning tegelikku ruutviga. Närvivõrkude ennustatud dispersiooni relatiivset viga hakatakse analüüsima, kas ja kuidas on seda võimalik ennustada, lähtudes teadmistest närvivõrgu (murdepunktide asukohad) ning andmete kohta (treeningandmepunktide jaotus).

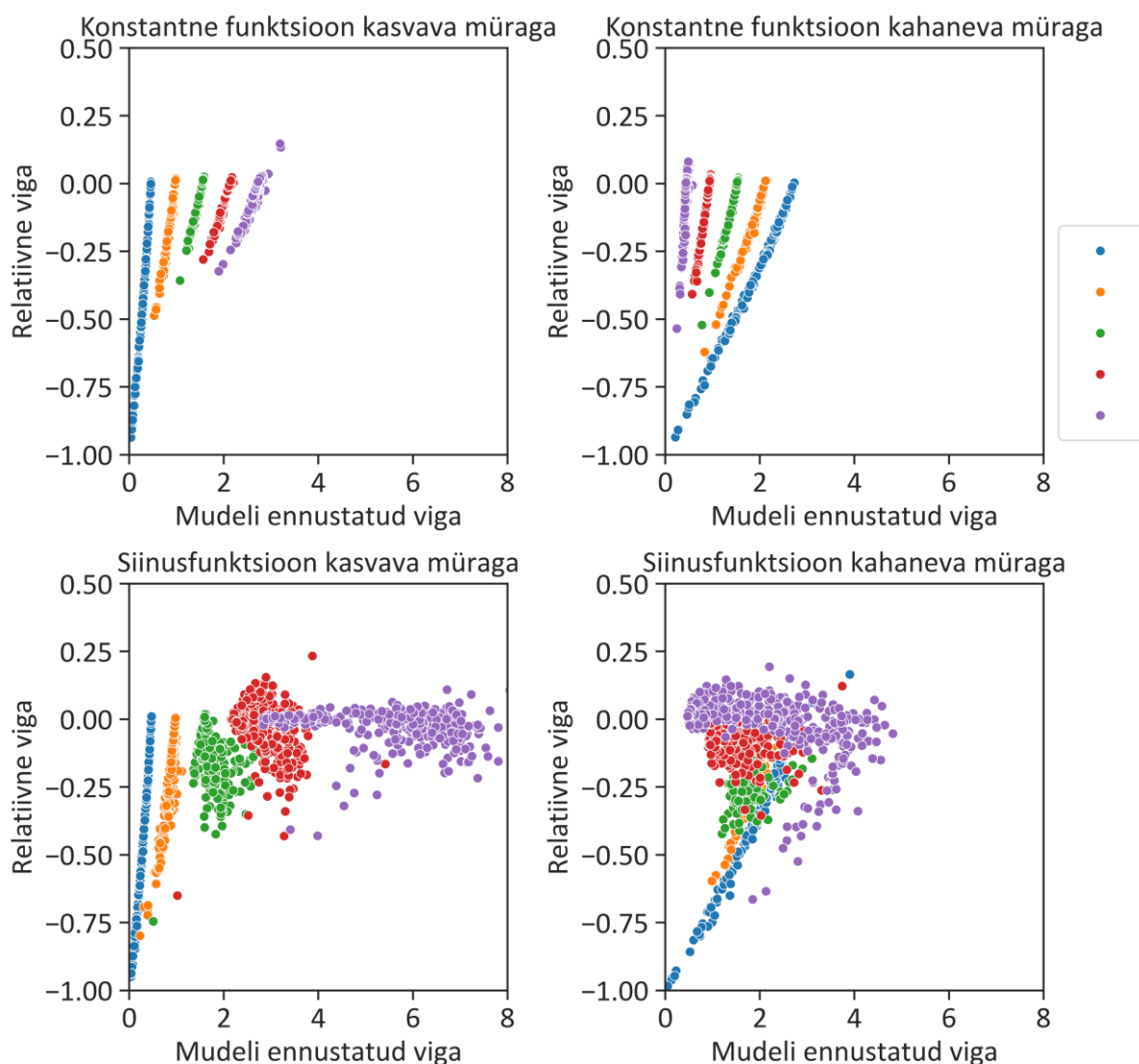
Tulemuste analüüsiks kasutati lineaarset regressiooni, täpsemalt statsmodelsi teegi vähimruutude meetodit (ingl *Ordinary Least Squares*) (Seabold & Perktold, 2010). Analüüsi tulemusena saadud lineaarse regressioonimudeli abil ennustatakse närvivõrkude igale alampiirkonnale kordaja, millega peaks dispersiooni hinnangu läbi korrutama, et kaofunktsioon väikesem oleks. Lõpetuseks uuritakse, kas närvivõrkude korrigeeritud dispersioonennustused vähendavad NLL kaofunktsiooni.

3.1 Murdepunktidest ja trendidest

Mudelite määramatuse ennustuste edukus on seotud treeningandmepunktide kogusega (piirkonna A relatiivsed vead on kujutatud joonisel 7, joonis 11 kujutab igat eksperimenti eraldi). Mida rohkem andmepunkte vastavas piirkonnas on, seda täpsemini ka mudel ennustab enda viga seal piirkonnas. Ühes piirkonnas on treeningpunktide arv vahemikus 0 kuni 1600, seega andmete paremaks visualiseerimiseks on x-teljele valitud logaritmiline skaala. Ühtlasi on märgata, et eksperimendid, millel ennustatav funktsioon oli lihtsam, on lihtsamini kirjeldatavate trendidega (joonis 8). Tähistades osades olevate treeningpunktide arvud N_A, \dots, N_E , siis j-ndas piirkonnas on konstrueeriti tunnused $1/N_j, \log(N_j), 1/\log(N_j)$, et lineaarne regressioon oleks võimeline ka tuvastama mittelineaarseid seosed.



Joonis 7. Relatiivne viga piirkonnas A seoses treeningpunktide arvuga samas piirkonnas.



Joonis 8. Mudeli ennustatud viga relatiivse vea suhtes sõltuvalt eksperimendist.

Murdepunktide seoseid on kergem näha joonisel 11, kust selgub, et mida vähem murdepunkte piirkonnas on, seda täpsemini mudel enda viga ennustab. See võib olla tingitud tree-ningandmete piirkonnast $[0,10)$, mis on ReLU jaoks liiga lai, sest ReLU aktivatsioonifunktsioon muutub nulli ümbruses. Teine võimalik seos on see, et mudel ennustab määramatust piirkonnas kõige paremini, kui piirkonna naaberpiirkondades on palju murdepunkte. Tulevikus võib korrata katseid valides uue treeningandmete piirkonnaks $[-1,1)$. Huvitava tähelepanekuna võib ka välja tuua, et kui piirkonnas on punktide tihedus suurem, siis sõltuvalt eksperimendist närvivõrk kas üle- või alahindab dispersiooni (joonis 12).

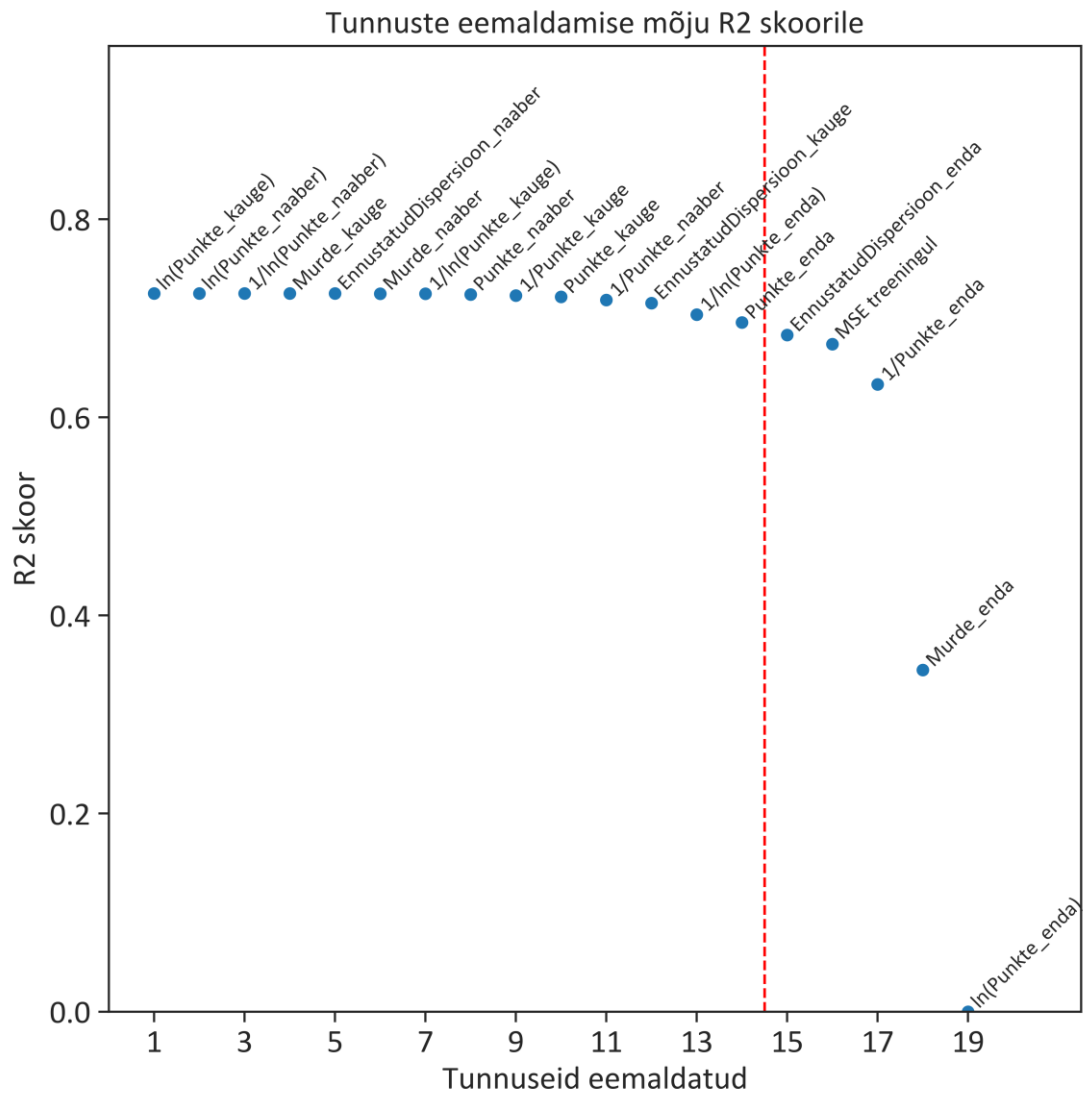
3.2 Tulemuste analüüs regressiooniga

Mustrid, mida üritatakse andmetest leida, on tõenäoliselt üsna nõrgad ning kuna tunnuseid on palju, on mõistlik rakendada tagurpidise sammregressiooni meetodit (ingl *backward stepwise regression*). Tagurpidine sammregressioon alustab täismudelist, mis sisaldab kõiki sisendtunnuseid ning igal iteratsioonil eemaldab tunnuse, mille eemaldamisel muutub valitud näitaja väärtus kõige vähem kehvemaks. Siin töös on selleks näitajaks valitud R^2 . Protsessi korratakse, kuni ainult üks tunnus on alles. Seejärel vaadatakse mudelite R^2 väärtust ning valitakse viimane mudel, enne kui R^2 oluliselt langema hakkas.

3.2.1 Kõigi tunnustega lineaarregressioon

Üks probleemidest lineaarse regressiooni sobitamisel on suur hulk kollineaarseid tunnuseid. Seetõttu uuritakse võimalusi, kuidas kollineaarsetest tunnustest valida vaid kõige olulisem või neid kombineerida üheks tunnuseks.

Sooritades tagurpidist sammregressiooni kõigi tunnustega andmestikul, on näha, et paljud tunnused on vähese mõjuga mudeli täpsusele (joonis 9). Kõige väikesema mõjuga on kaugete piirkondade punktidest konstrueeritud tunnused ning kõige suurema mõjuga on murdepunktide kogus ning enda piirkonna punktide kogus.



Joonis 9. Sammhaaval eemaldatud tunnused ning vastava mudeli R^2 väärtus.

Sobitatud mudeli R^2 väärtus on 0,696. Konstrueeritud tunnused annavad lineaarsele regressioonile häid tunnuseid, sest $\log(N_j)$ ning $1/N_j$ tunnused jäävad regressioonile sisse. Murdepunktide lähedus on samuti oluline relatiivse vea ennustamiseks, sest uuritava piirkonna murdepunktidel on regressioonimudeli täpsusele suurem mõju kui kaugemate piirkondade murdepunktide arvul. (täpsed kaalud on näha lisa III tabelis 4).

3.2.2 Punktidest konstrueeritud tunnuste kombineerimine

Tunnuste arvu vähendamiseks sobitati neli lineaarse regressiooni mudelit, kasutades algsete tunnuste alamhulka, mis ei sisalda treeningandmepunktide tunnuseid⁴. Neile tunnustele lisati üks treeningandmepunktide tunnus⁵. Nelja regressiooni põhjal otsustati, et kõige suurema olulisusega tunnuseks on treeningandmepunktide logaritmi pöördarv $1/\log(N_j)$ (tabel 1).

Tabel 1 Lineaarse regressioonimudelite, mis sisaldasid vaid ühte konstrueeritud tunnust, R^2 väärtused

Tunnus	R^2 väärtus
N_j	0,417
$1/N_j$	0,679
$\log(N_j)$	0,658
$1/\log(N_j)$	0,703

Uus mudel on suurema determinatsioonikordajaga kui mudel, mis oli sobitatud kõigi tunnustega (eelmise mudeli R^2 oli 0,696), mis viitab sellele, et tunnuseid oli algses mudelis liiga palju. Edaspidi kasutatakse treeningandmepunktidest tuletatud tunnuseks vaid $1/\log(N_j)$.

3.2.3 Murdepunktidest kombineeritud tunnus

Leidmaks, milliste piirkondade murdepunktid kõige rohkem ennustuse täpsust mõjutavad, eemaldati murdepunktide tunnuseid ükshaaval (tabel 2).

⁴ Kõikideks tunnusteks olid: Murde_enda, Murde_naaber, Murde_kauge, EnnustatudDispersioon_enda, EnnustatudDispersioon_naaber, EnnustatudDispersioon_kauge ja MSE treeningul.

⁵ Konstrueeritud tunnusteks oli: $1/N_j$, $\log(N_j)$, $1/\log(N_j)$, kus N_j tähistab punktide arvu osas.

Tabel 2 Murdepunktide tunnuste eemaldamisel saavutatavad R^2 väärtused

Eemaldatud tunnus(ed)	R^2 väärtus
Murde_kauge	0,703
Murde_naaber	0,703
Murde_naaber ja Murde_kauge	0,703
Murde_enda	0,700
Murde_enda, Murde_naaber ja Murde_kauge	0,544

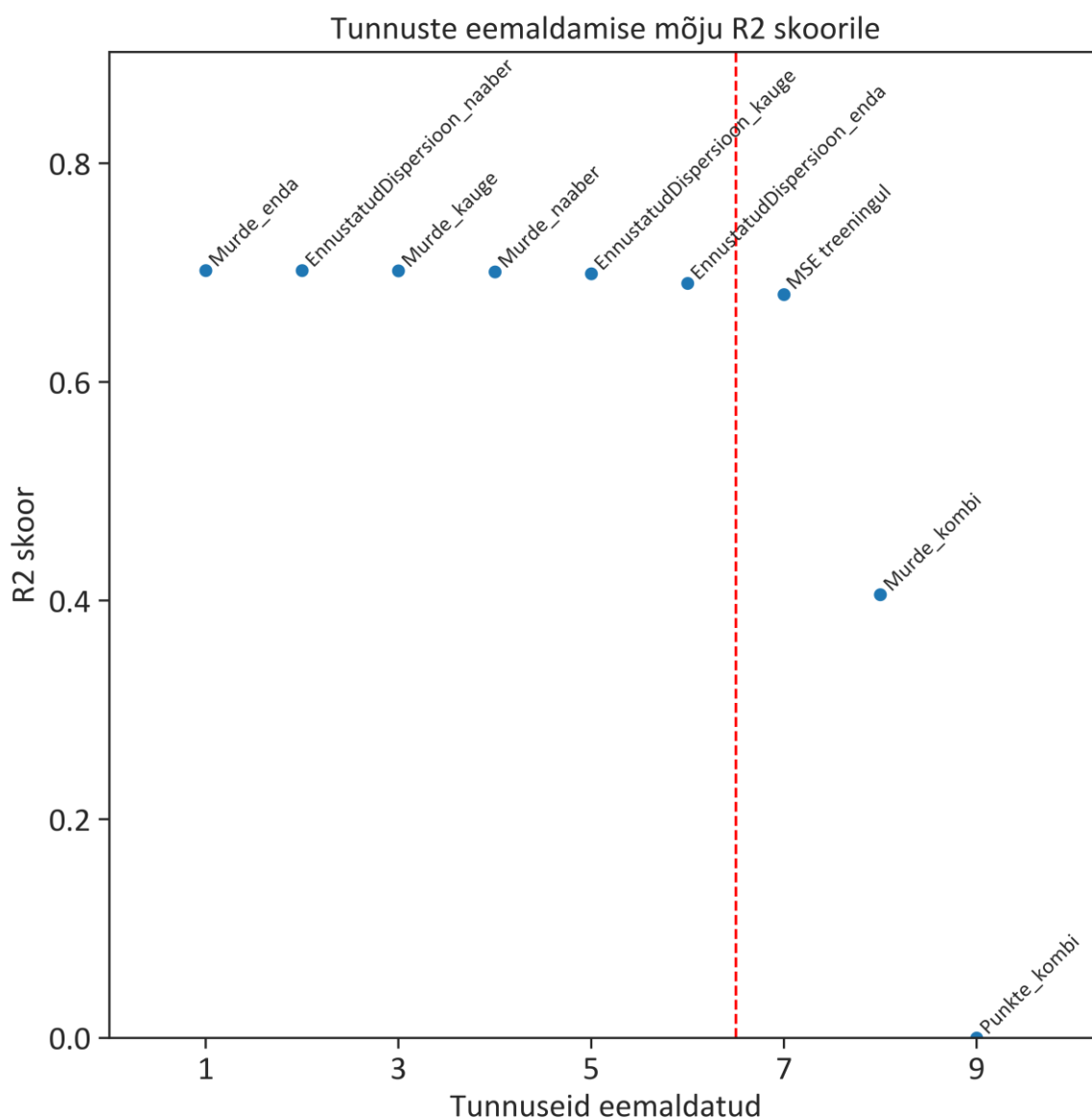
Kuna eelmises peatükis sobitatud mudeli R^2 väärtus oli samuti 0,703, on murdepunktide tunnustes leiduv info vajalik, et paremini relatiivset viga ennustada. Murdepunktide kogused erinevates piirkondades pole omavahel lineaarselt lineaarselt korreleeritud, kuid nad sisaldavad olulist infot relatiivse vea ennustamiseks. Põhjus, miks mudelite R^2 väärtused tulevad samad, eemaldades nii ainult naabermurdepunktide tunnuse kui ka naabrite ja kaugemate murdepunktide tunnused, on see, et kaugete murdepunktide tunnuste p väärtus osutus statistiliselt mitteoluliseks ja seega tunnus eemaldati, ehk teine ja kolmas regressioon on samadel tunnustel sobitatud.

3.3 Regressioon kombineeritud tulemustega

Lõpliku regressioonmudeli sobitamiseks kombineeriti nii eri osade $1/\log(N_j)$ tunnused kui ka eri osade murdepunktide tunnused üheks tunnuseks⁶. Nende uute tunnuste loomiseks vajalikud kaalud on leitavad tabelitest 5 ja 6.

Analoogselt kõigi tunnustega lineaarregressiooniga, mida sobitati peatükis 3.2.1, tehti tagurpidist sammregressiooni, et valida sobilik tunnuste alamhulk (joonis 10). Saadud mudeli R^2 väärtus on 0,690 (tabel 3).

⁶ Kui jätta treeningandmepunktide pöördväärtused ning murdepunktide väärtused kuue tunnusega mudelisse, on saadava mudeli R^2 väärtus 0,703.



Joonis 10. Kombineeritud tunnustega regressioonimudeli tunnuste valikukõver.

Tabel 3 Kombineeritud tunnustega regressioonimudeli kaalud

	coef	std err	t	P> t	[0.025	0.975]
MSE treeningul	-0.0725	0.004	-16.276	0	-0.081	-0.064
Punkte_kombi	0.6786	0.007	102.712	0	0.666	0.692
Murde_kombi	0.4059	0.005	84.548	0	0.396	0.415
Konstant	-0.9142	0.007	-135.215	0	-0.927	-0.901

Kõik lineaarregressiooni sisendtunnused olid normaliseeritud vahemikku 0 kuni 1, seega saab kaalude absoluutväärtuste järgi hinnata, kui palju milline tunnus mõjutab ennustust.

Kõige olulisemaks tunnuseks osutus treeningandmepunktide $\log(N_j)$ kogusest kombineeritud tunnus mitte murdepunktidest kombineeritud tunnus.

3.4 Mudeli tulemuslikkus mõõdetud vigade hindamisel

Eelmises punktis sobitatud lineaarse regressiooni mudelit kasutati, et ennustada relatiivset viga, mida närvivõrk oma dispersiooni ennustusel teeb. Relatiivse vea valem oli $R = \frac{d_{pred} - d_{mse}}{d_{mse}}$, seega selle jaoks, et närvivõrkude ennustatud dispersiooni korrigeerida, tuleb ennustatud dispersioon läbi korrutada koefitsiendiga $\frac{1}{R+1}$.

Korrigeerides kõigi eelnevalt treenitud närvivõrkude dispersiooni ennustusi vastava koefitsiendiga, kahanes NLL vaid 8,6% juhtudest. See näitab, et kuigi varasemalt jõuti järeldusele, et treeningandmepunktide kogus on olulisem kui murdepunktide kogus uuritavas piirkonnas, pole lineaarsete seostega võimalik närvivõrgu üle- ning alahindamisi ära seletada. Kõige rohkematel juhtudel õnnestus parendada kolmanda eksperimendi määramatusi ehk eksperimenti, kus närvivõrk ennustas siinusfunktsiooni kasvava müraga. Konstantsete funktsioonidega ei õnnestunud peaaegu üldse hinnangut parendada. Siinusfunktsiooni ennustamisel õnnestus ennustusi parendada 17,3% juhtudest.

Korrigeerides sama meetodiga testandmehulgal, mis koosnes vaid siinusfunktsioonide ennustustest, oli meetod võimeline parendama NLL-i 18,3% juhtudest. See näitab, et lineaarse regressiooni mudel ei ülesobitanud andmeid, kuid tulemus on tagasihoidlik. Problemaatiline on ka meetodi rakendamiseks vajalik info murdepunktide paiknemise kohta, mida on keeruline leida rohkemate sisendmuutujatega regressioonülesannetel.

Kokkuvõte

Töö eesmärgiks oli paremini mõista, mis mõjutab ReLU närvivõrgu määramatuse hinnangut. Määramatuse hinnangute analüüsiks treeniti kahekihilist ReLU närvivõrku erinevatel treeningülesannetel ning mõõdeti nende dispersiooni hinnanguid ning murdepunktide asukohtasid. Seejärel sobitati lineaarse regressiooni mudelit ennustamaks, kui palju närvivõrgu dispersiooni hinnang eksis.

Analüüsis olid kõige olulisemad faktorid määramatuse ennustamisel piirkonnas asuvate punktide pöördväärtus ning murdepunktide arv, kus mõlema tunnuse puhul uuritavale punktile lähemal asuvate treeningandmepunktide ning murdepunktide kogus on olulisem kui kaugemal asuvate kogus. Vähemal määral on ka oluline naaberpiirkondadesse ennustatud dispersioon, kuid see võib tuleneda katsetes kasutatud müra valemi lineaarsetest olemustest. Saadud tulemused kinnitasid, et närvivõrgu ennustuste kvaliteet on tugevalt seotud treeningandmete jaotusega sisendruumis. Korrigeerides saadud mudeli abil treenitud närvivõrkude ennustusi, paranes kaofunktsioon vaid juhtudel, kui mudel ei olnud koondunud.

Tulevikus tasuks uurida, kas samad tähelepanekud kehtivad, kui uuritav piirkond pole $[0, 10)$ vaid $[-1, 1)$, sest see on aktivatsioonipunkti ümber tsentreeritud. Võimalik on ka uurida erinevaid mürafunktsioone, et kontrollida, kas naaberpiirkondade dispersiooni olulisus tuleneb lineaarsest mürafunktsioonist või on see midagi selles töös seletamatut. Piirkondade kitsendamine ning arvu suurendamine tekitaks samuti rohkem andmepunkte, mida analüüsida.

Viidatud kirjandus

- Chollet, F. (2015). Keras. Retrieved from <https://keras.io>
- Detlefsen, N. S., Jørgensen, M., & Hauberg, S. (2019). Reliable training and estimation of variance networks.
- Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological cybernetics* 20, no. 3, 121-136.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). New York: Springer.
- Hüllermeier, E. (2020, September 18). Uncertainty in Machine Learning. 8-16. Paderborn.
- Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. In *Mach Learn 110* (pp. 467-506). Springer.
- Ji, X., Pascanu, R., Hjelm, D., Lakshminarayanan, B., & Vedaldi, A. (2022). Test Sample Accuracy Scales with Training Sample Density in Neural Networks. *Lifelong Learning Agents*. doi:arXiv:2106.08365
- Kendal, A., & Gal, Y. (2017). Advances in Neural Information Processing Systems. *Advances in Neural Information Processing Systems* (p. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?). Curran Associates, Inc.
- Kull, M., & Flach, P. (2015). Novel Decompositions of Proper Scoring Rules for Classification: Score Adjustment as Precursor to Calibration. *Machine Learning and Knowledge Discovery in Databases* (pp. 68-85). Porto: Springer.
- Montufar, G. (2017). *Notes on the number of linear regions of deep neural networks*. eScholarship.
- Montúfar, G., Pascanu, R., Cho, K., & Bengio, Y. (2014). On the Number of Linear Regions of Deep Neural Networks. *Advances in neural information processing systems* 27.
- Nix, D., & Weigend, A. (1994). Estimating the mean and variance of the target probability distribution. *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, (pp. 55-60). Orlando.
- Novak, R., Bahri, Y., Abolafia, D. A., Pennington, J., & Sohl-Dickstein, J. (2018). Sensitivity and Generalization in Neural Networks: an Empirical Study. *International Conference on Learning Representations*. ICLR.

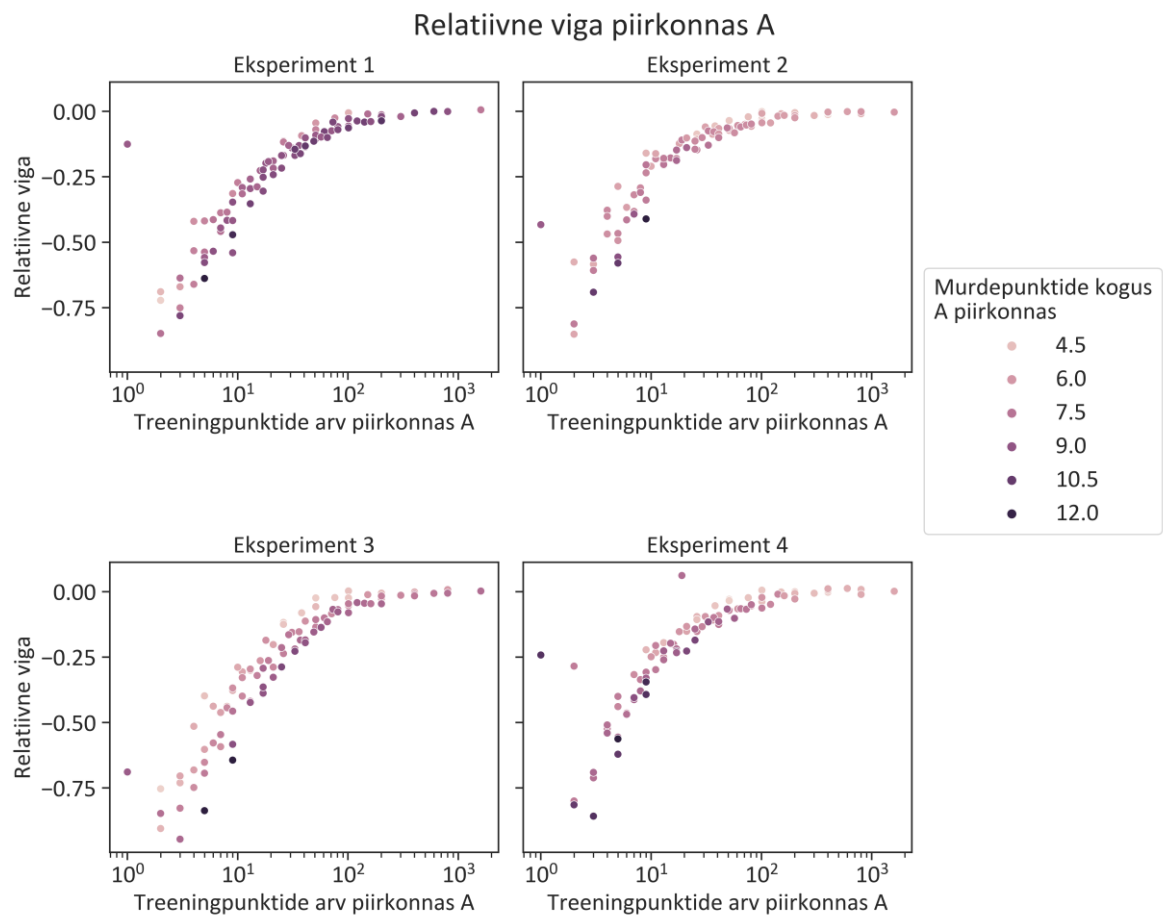
- Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*.
- Serra, T., Tjandraatmadja, C., & Ramalingam, S. (2018). Bounding and counting linear regions of deep neural networks. *International Conference on Machine Learning* (pp. 4558-4566). PMLR.

Lisad

I. Koodi repositoorium

Töös jooksutatud katsete ning loodud visualiseeringute kood on kättesaadav aadressil <https://github.com/ToomasRo/BAtoo>.

II. Lisa joonised



Joonis 11. Relatiivne viga A piirkonnas erinevatel andmestikel.

III. Regressioonide kordajad

Tabel 4 Kõigi tunnustega tagurpidise sammregressiooni kordajad

	coef	std err	t	P> t	[0.025	0.975]
Konstant	-0.1141	0.005	-21.968	0	-0.124	-0.104
Murde_enda	-0.3563	0.005	-67.264	0	-0.367	-0.346
EnnustatudDispersioon_enda	0.2138	0.012	18.243	0	0.191	0.237
1/Punkte_enda	-0.4783	0.014	-34.313	0	-0.506	-0.451
ln(Punkte_enda)	0.2866	0.008	35.553	0	0.271	0.302
MSE treeningul	-0.0982	0.005	-20.707	0	-0.108	-0.089

Tabel 5 Punktide pöördväärtuste kombineerimiseks vajalikud kaalud

	coef	std err	t	P> t	[0.025	0.975]
1/ln(Punkte_enda)	-0.4962	0.022	-22.271	0	-0.54	-0.453
1/ln(Punkte_naaber)	-0.2664	0.026	-10.405	0	-0.317	-0.216
1/ln(Punkte_kauge)	-0.183	0.023	-7.847	0	-0.229	-0.137
Konstant	0.0496	0.002	20.373	0	0.045	0.054

Tabel 6 Murdepunktide kombineerimiseks vajalikud kaalud

	coef	std err	t	P> t	[0.025	0.975]
Murde_enda	-0.3759	0.005	-71.718	0	-0.386	-0.366
Murde_naaber	-0.0598	0.006	-10.443	0	-0.071	-0.049
Murde_kauge	-0.0245	0.006	-4.343	0	-0.036	-0.013
Konstant	0.1122	0.003	32.522	0	0.105	0.119

IV. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Toomas Roosma,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose
ReLU närvivõrgu ennustusmääramatuse hindamine regressioonülesannetes
mille juhendaja on Meelis Kull,
reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi
DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks
Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative
Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost repro-
dutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada
teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega
isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Toomas Roosma

09.05.2023