

Masinõpe ja selle kvaliteet

Meelis Kull
TÜ tehisintellekti professor, uurija-professor

Tehisaru ja masinõpe

- Tehisaru on palju rohkem kui vaid masinõpe:
 - Otsingusüsteemid
 - Tõestussüsteemid
 - Simulatsioonisüsteemid
 - Planeerimissüsteemid
 - jpm
- Samas on üha vähem tehisarusüsteeme, mis on masinõppest puutumata



Masinõpe



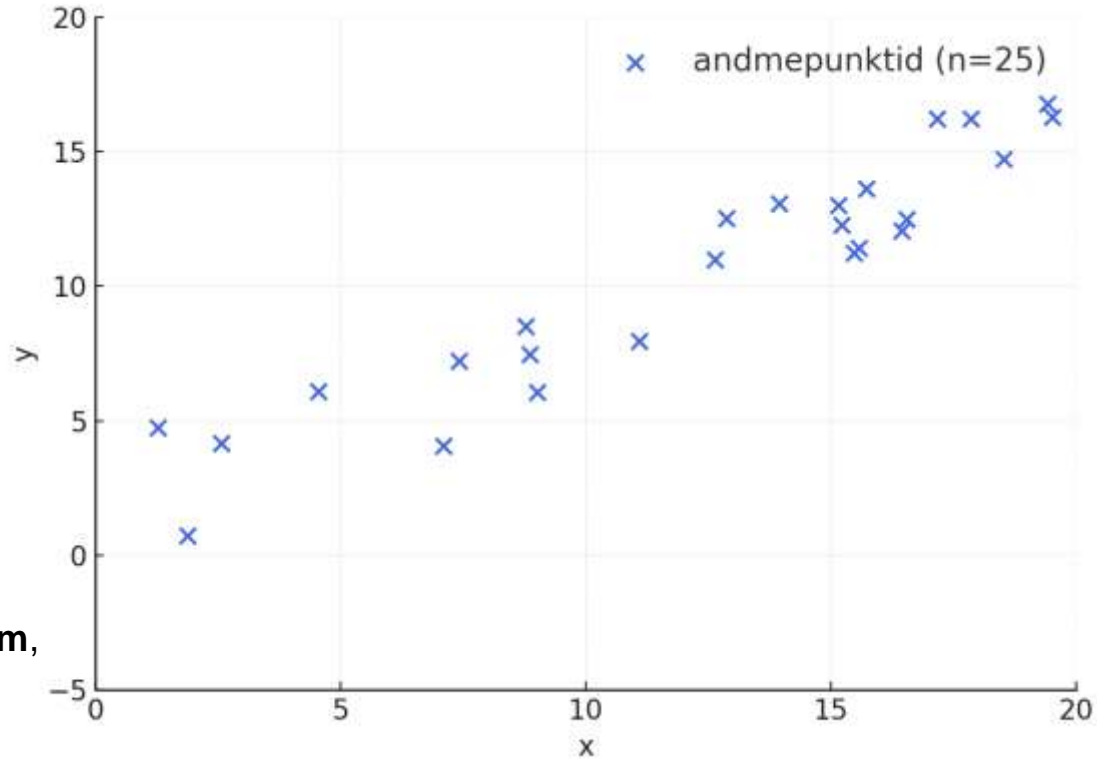
- Eesmärk:
Leida funktsioon, millega saada sisendist x soovitud omadustega väljund y
- Tüüpiline meetod – tehisnärvivõrk, mille puhul masinõppeinsener valib:
 - **Treeningandmed**
 - **Mudeli arhitektuuri** ehk parameetrilise funktsioonide pere, mis võiks sisaldada sobivat funktsiooni
 - **Sihtfunktsiooni**, mis mõõdab funktsiooni sobivust
 - **Optimeerimisalgoritmi**, mis sihtfunktsiooni optimeerib
- Mudeli optimeerimise tulemusena saadakse soovitud funktsioon
- Optimeerimisel ‘õpib’ mudel ära kasutama andmete korrapära, mustreid
- Tulevastel andmetel rakendatakse seda ‘õpitud’ funktsiooni

Treeningandmed

Mudeli arhitektuur
ehk parameetriline
funktsioonide pere,
mis võiks sisaldada
sobivat funktsiooni

Sihtfunktsioon,
mis mõõdab
funktsiooni sobivust

Optimeerimisalgoritm,
mis sihtfunktsiooni
optimeerib

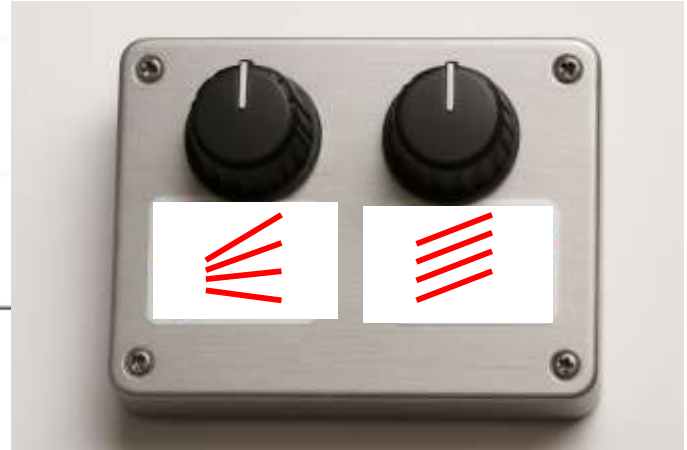
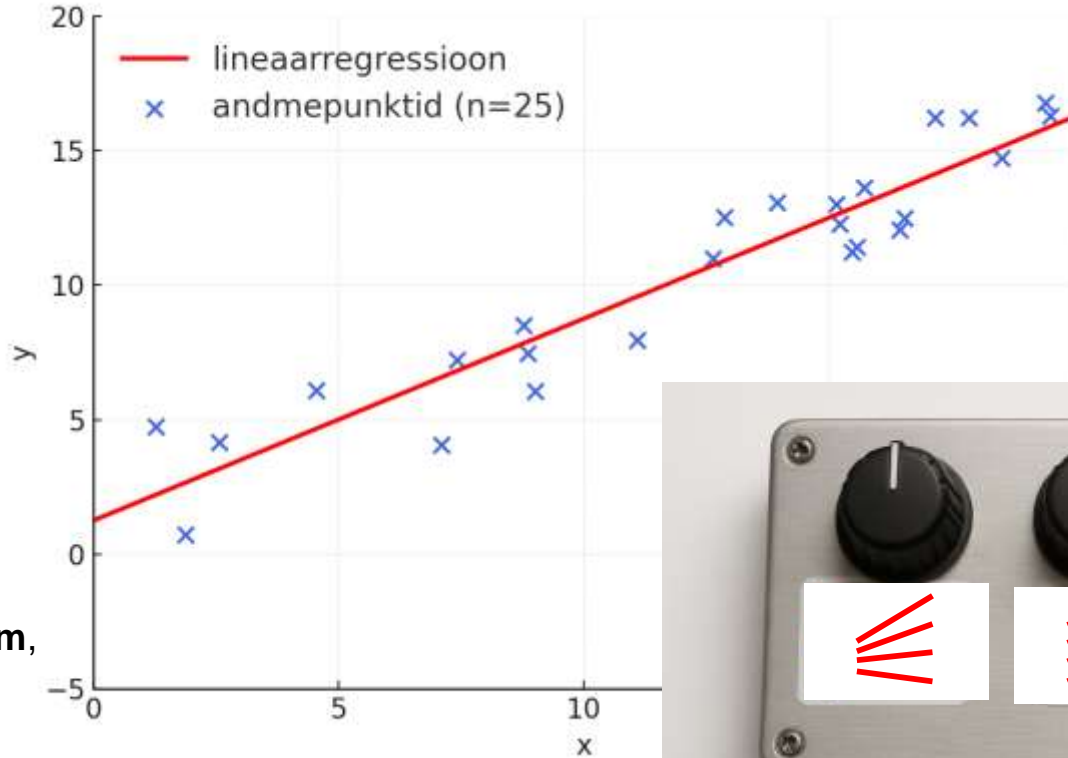


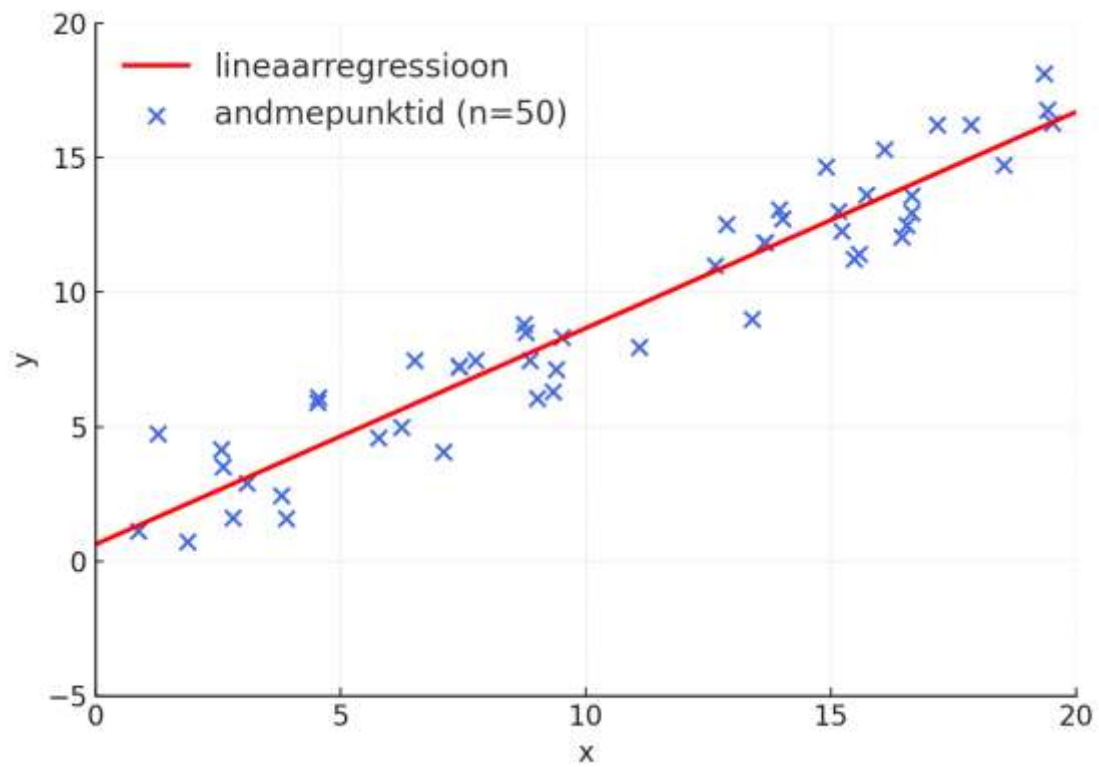
Treeningandmed

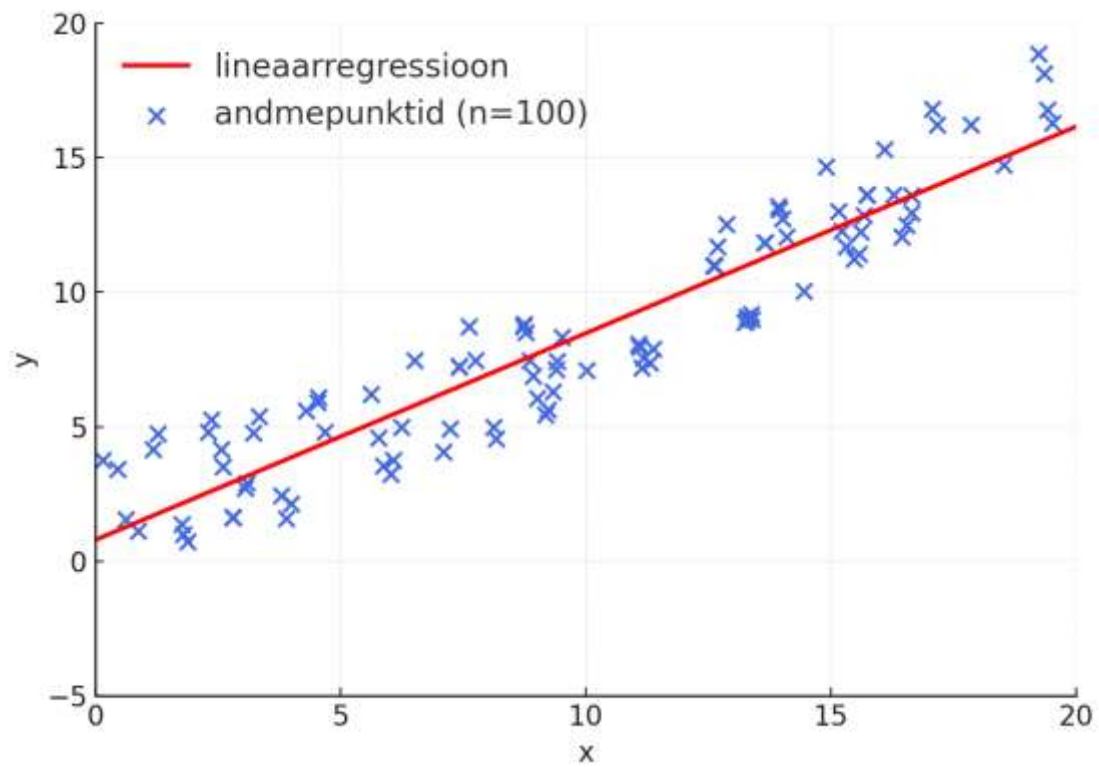
Mudeli arhitektuur
ehk parameetriline
funktsioonide pere,
mis võiks sisaldada
sobivat funktsiooni

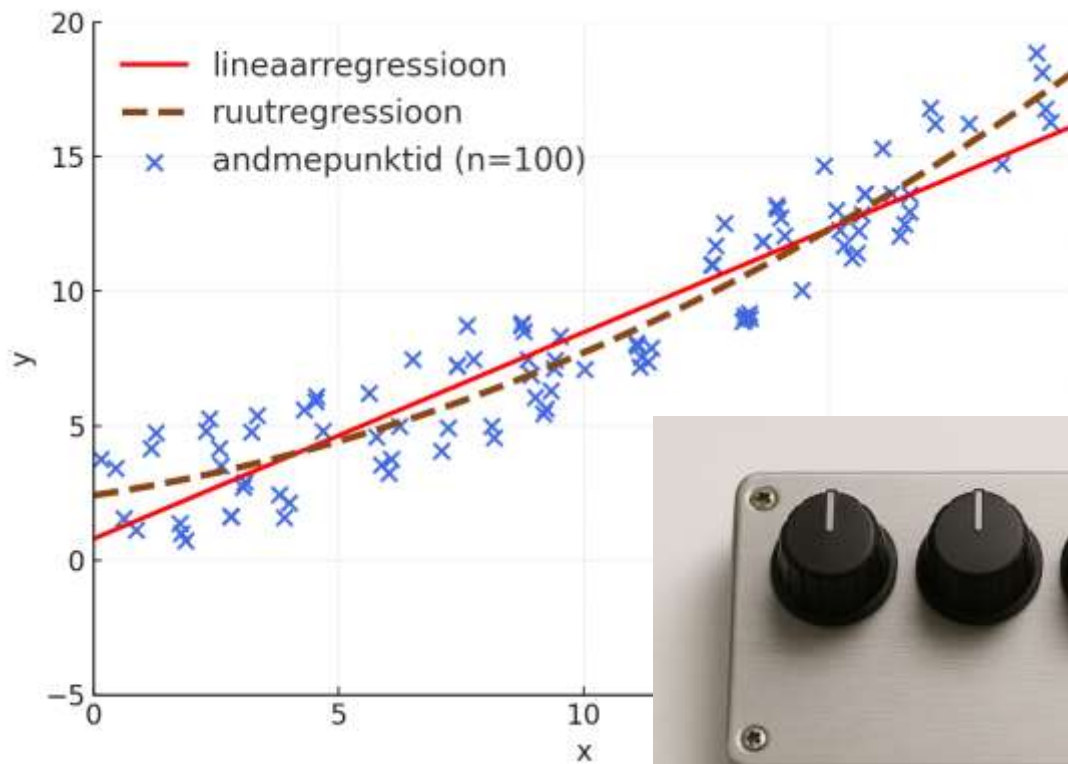
Sihtfunktsioon,
mis mõõdab
funktsiooni sobivust

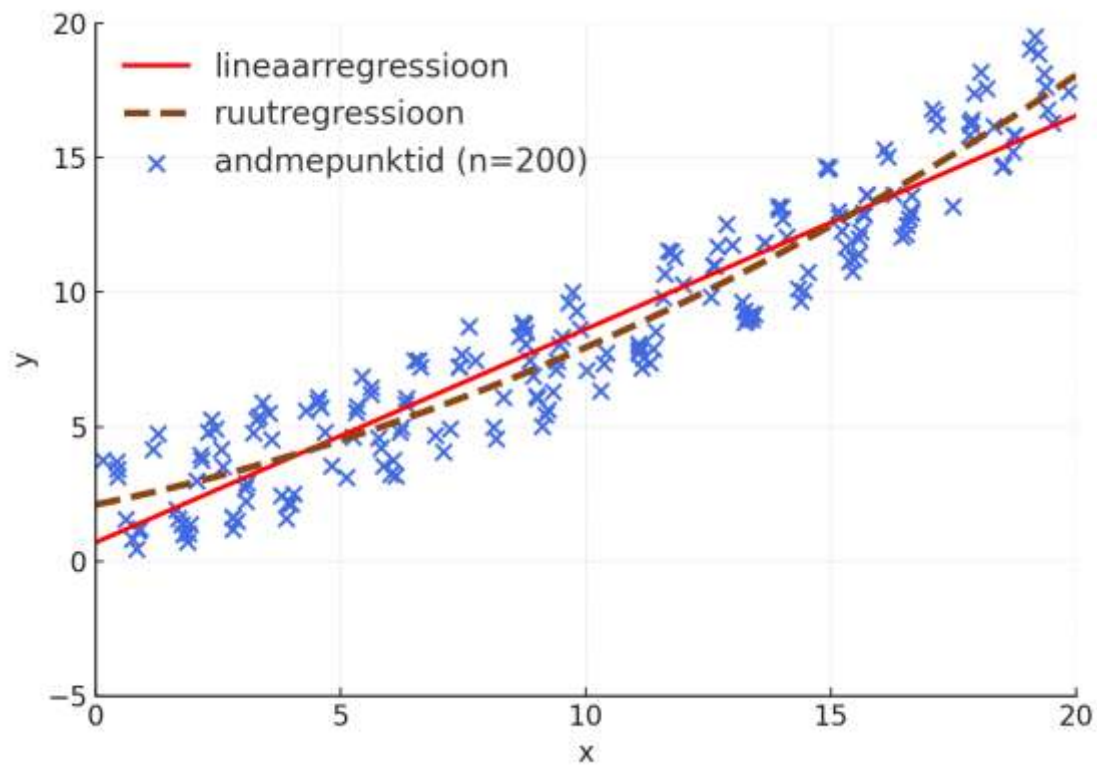
Optimeerimisalgoritm,
mis sihtfunktsiooni
optimeerib

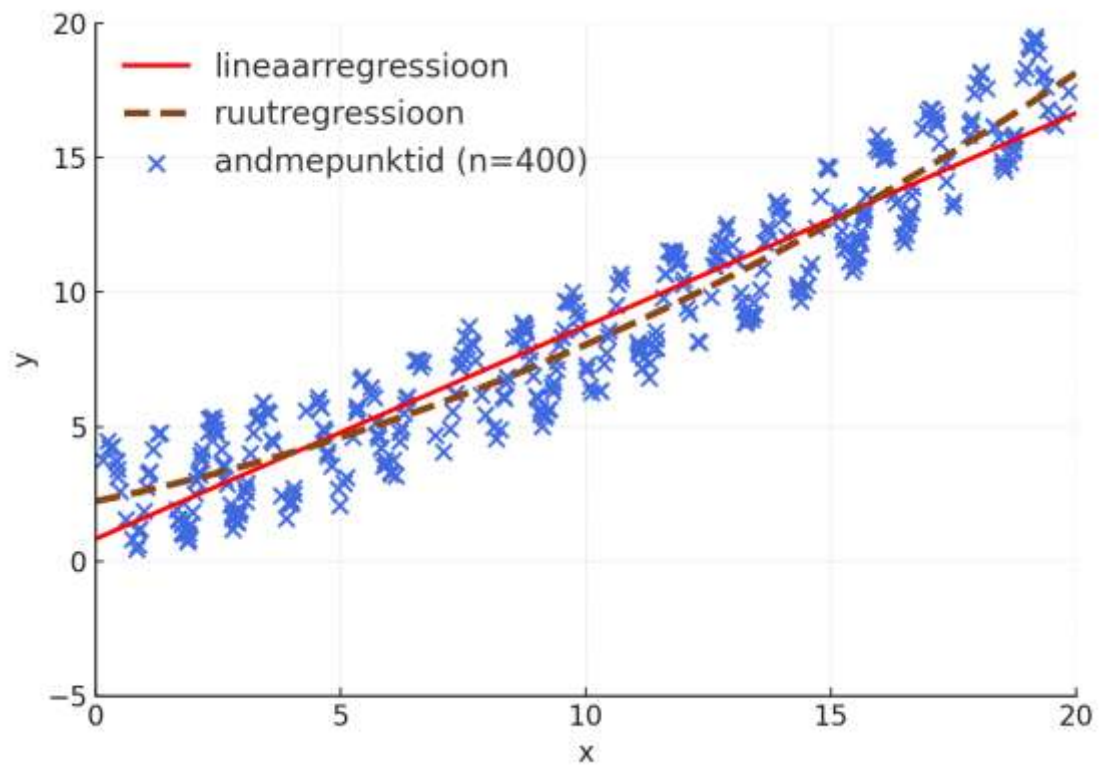


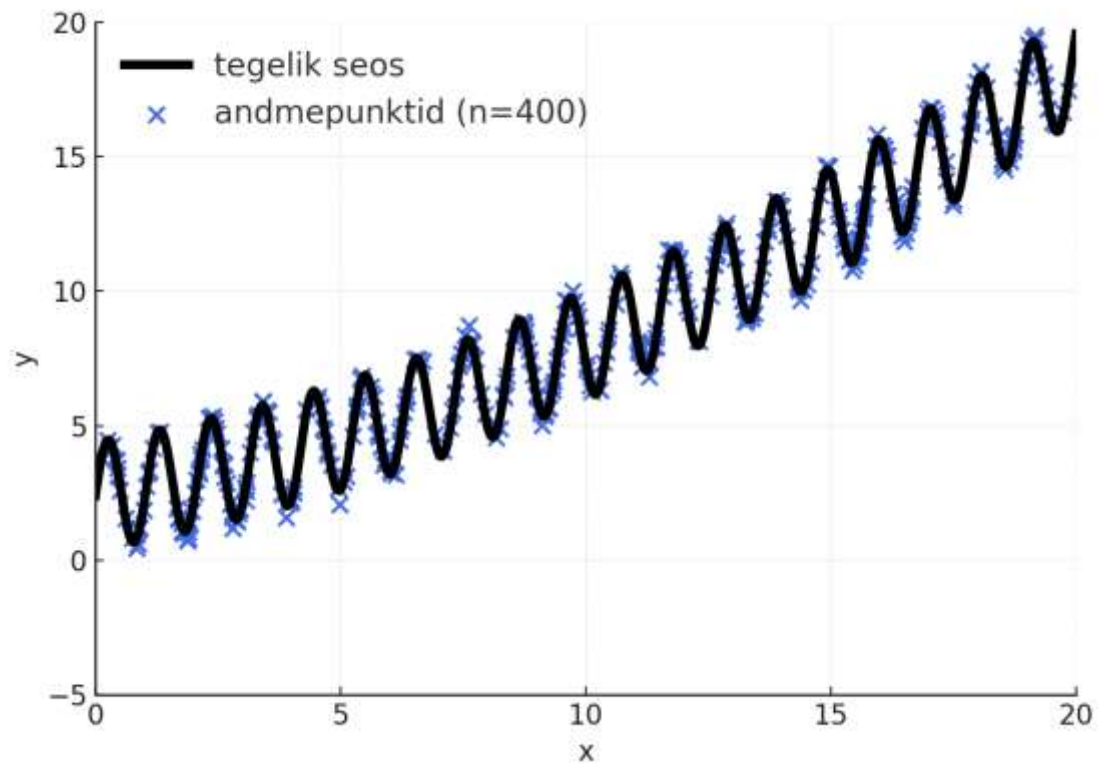


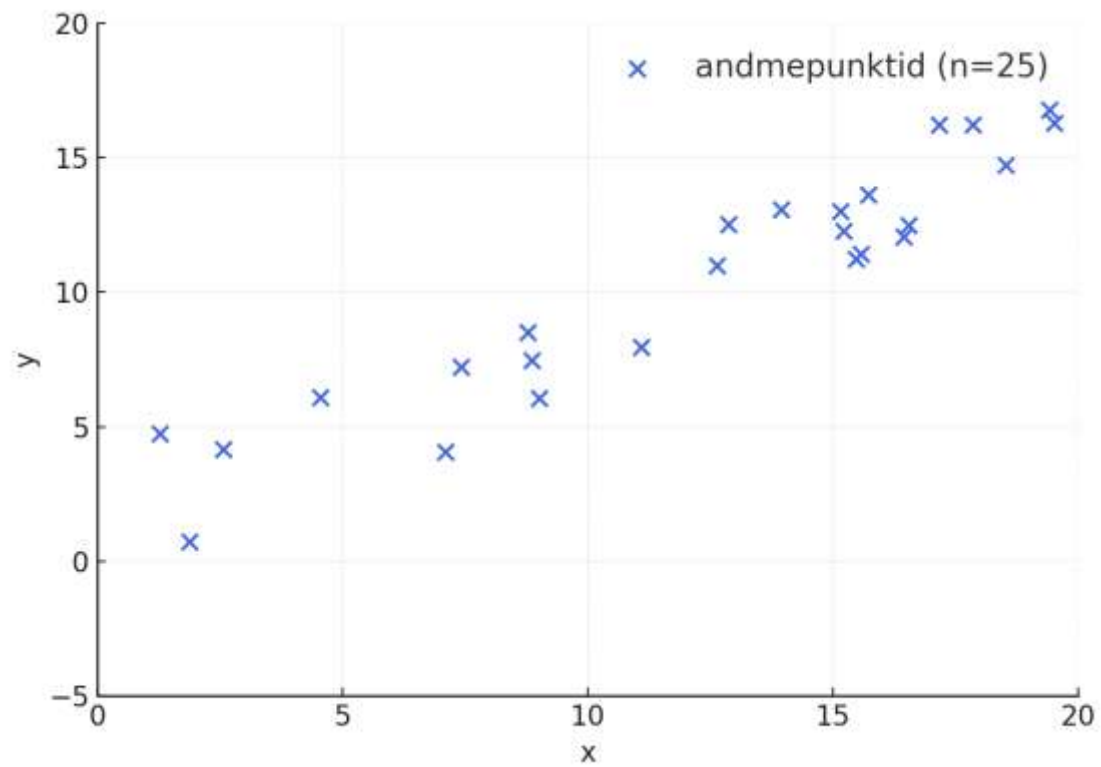










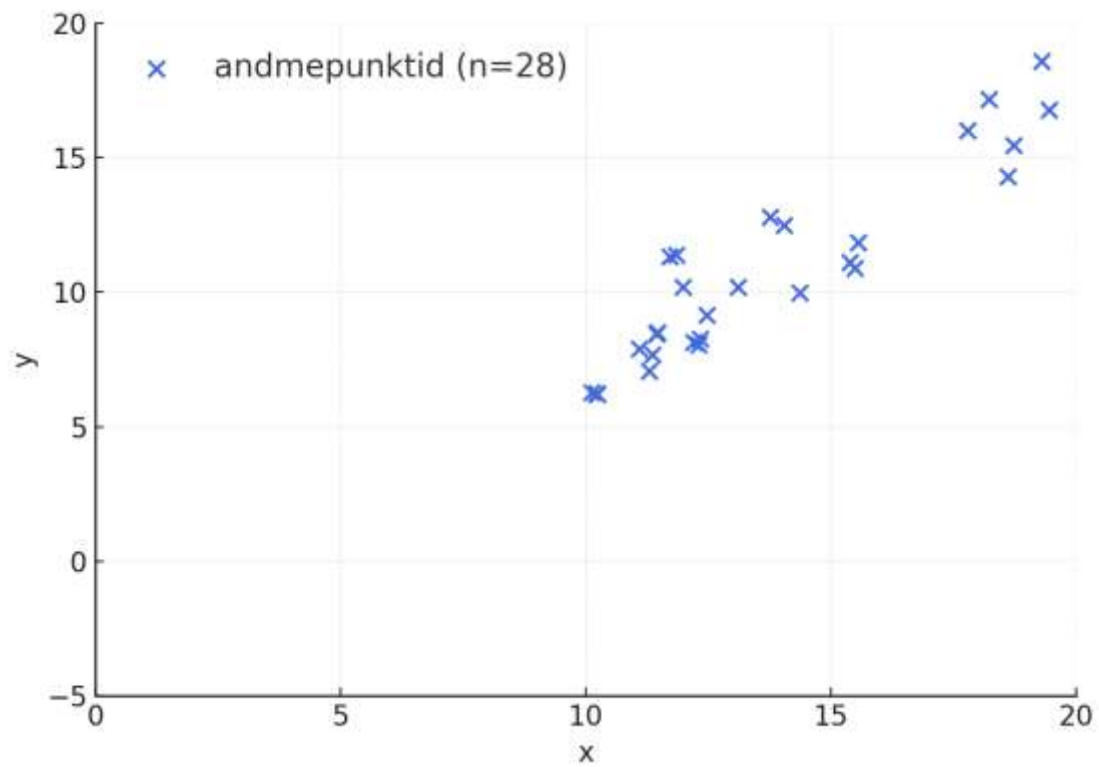


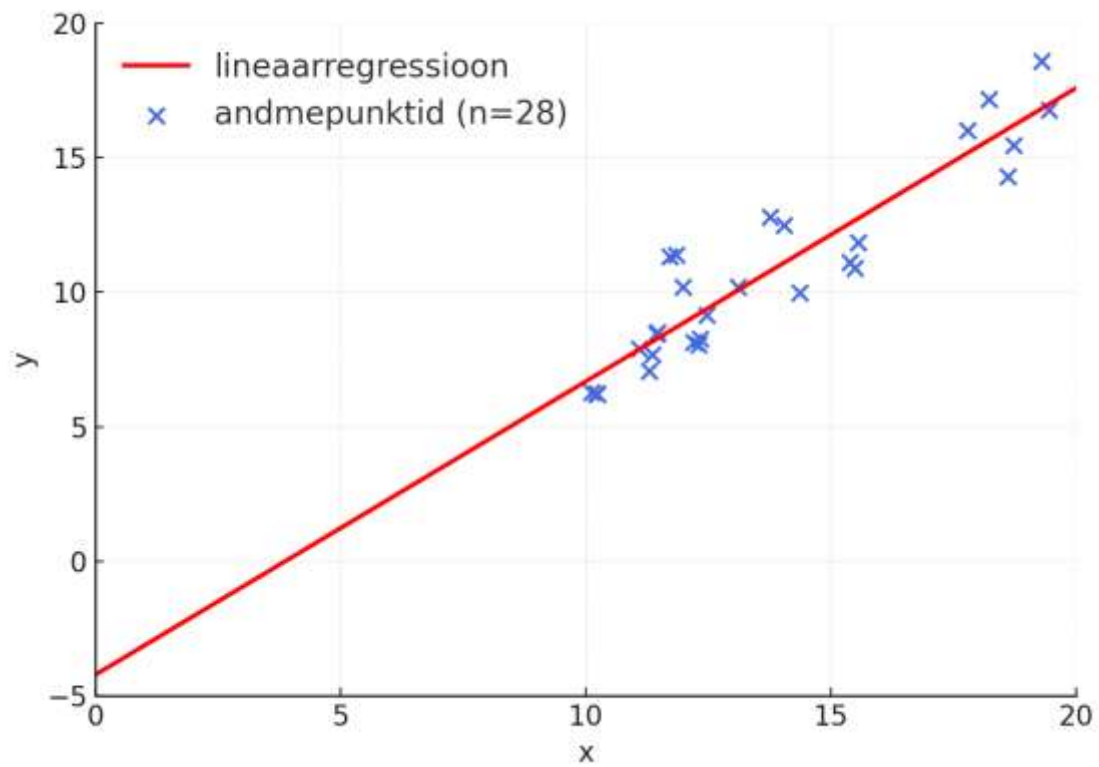
Tehisnärvivõrgud

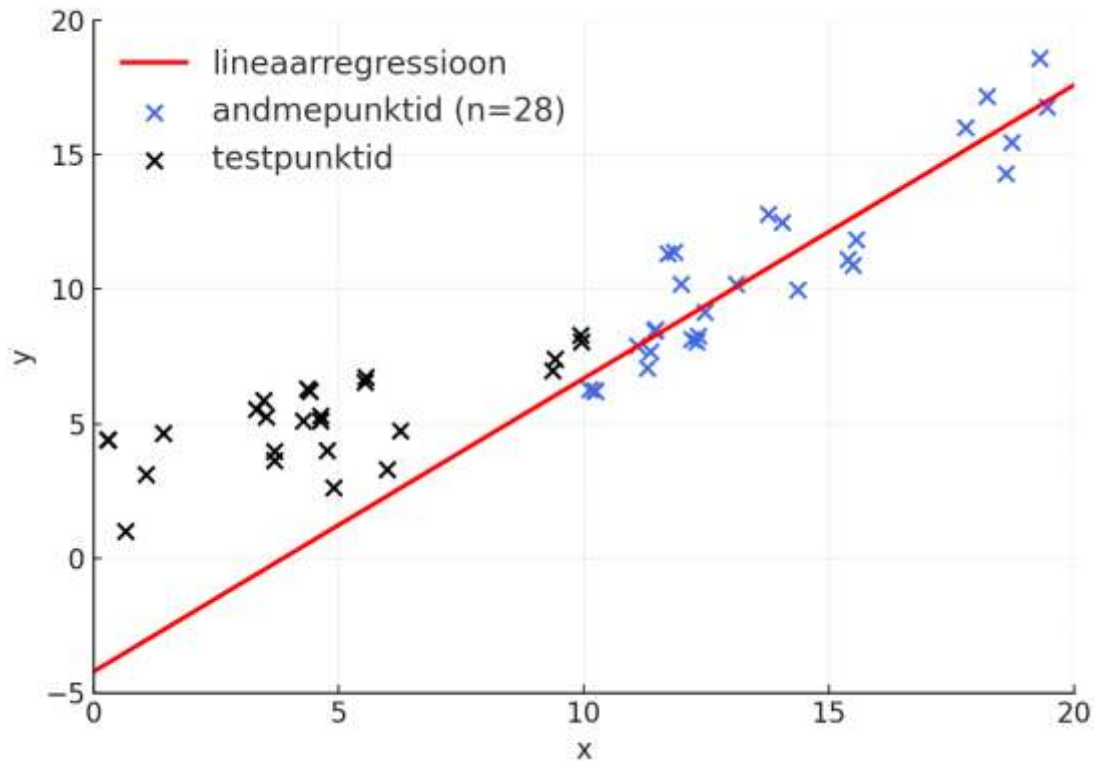
- Parameetrilised funktsioonide pered
- Palju erinevaid arhitektuure
- Mida suurem mudel, seda rohkem mustreid saab ära õppida
- Aga selleks kulub ka rohkem treeningandmeid



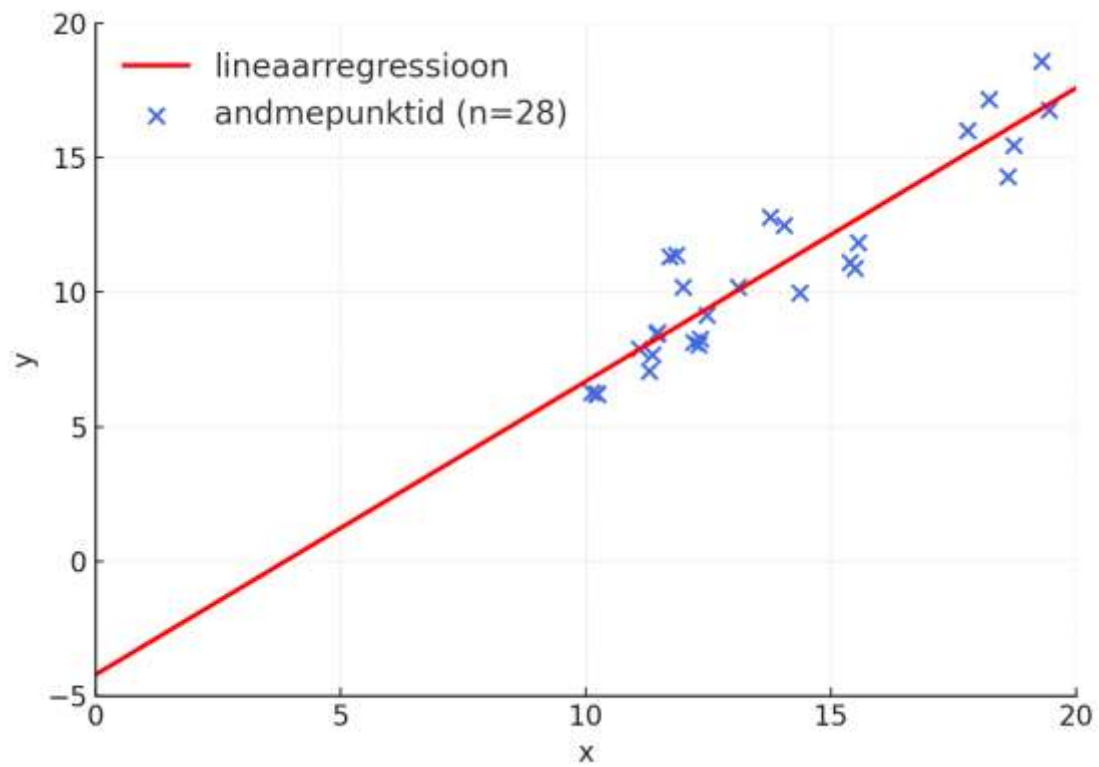
Kvaliteediprobleemid masinõppes

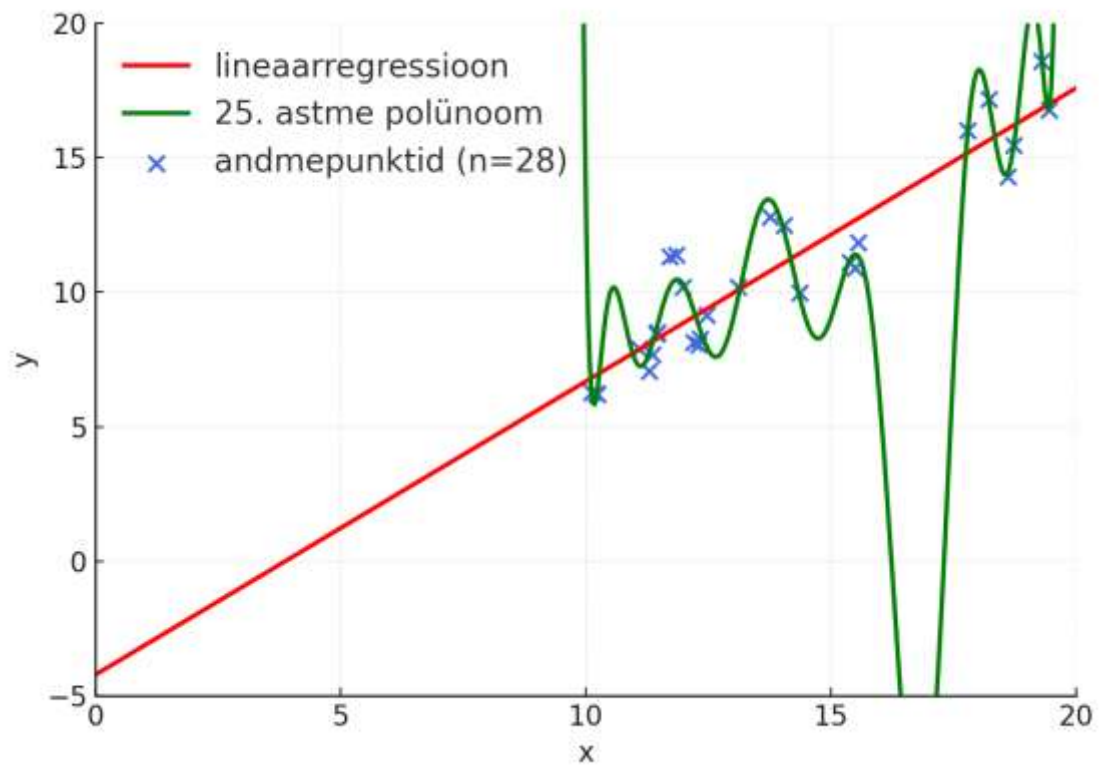


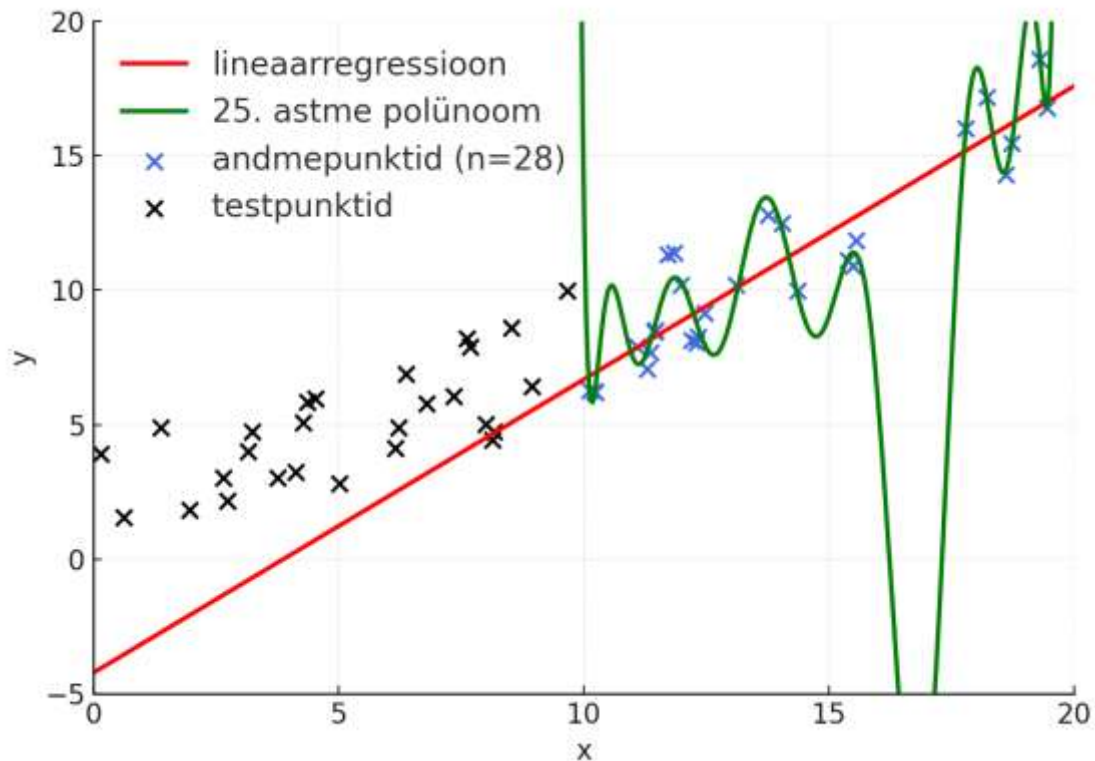




Probleem: kehv kvaliteediga ennustused ekstrapoleerimisel
Lahendus: mitmekesisemad treeningandmed

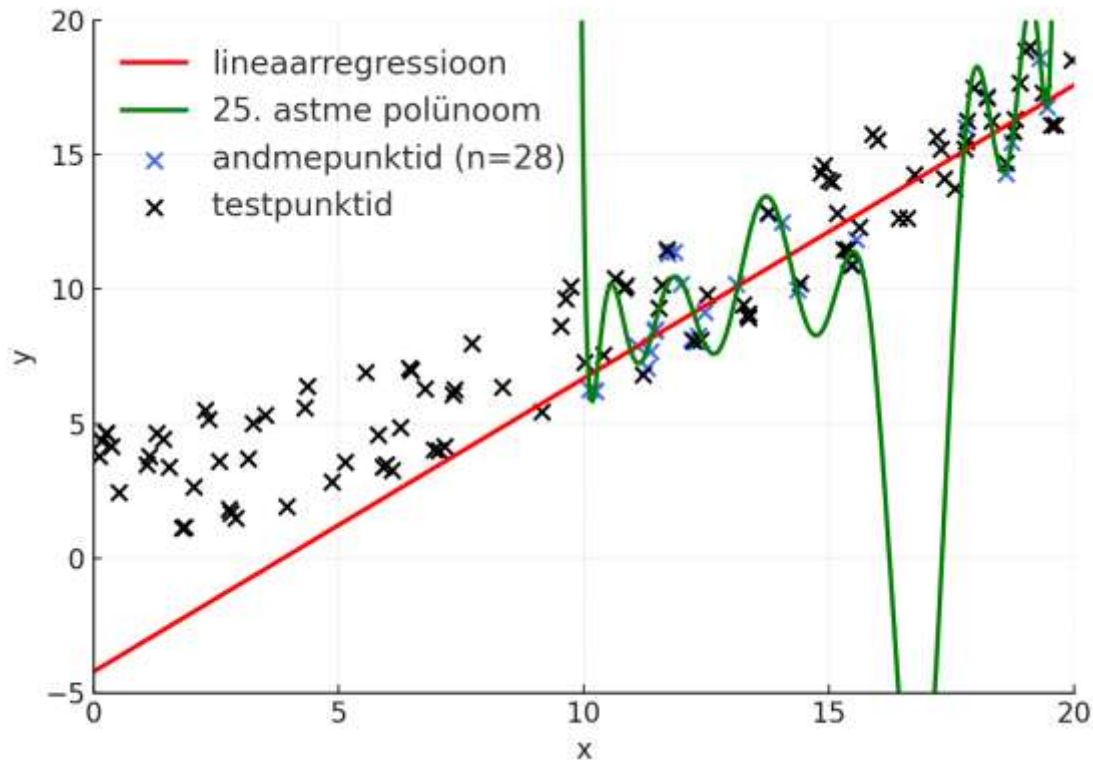




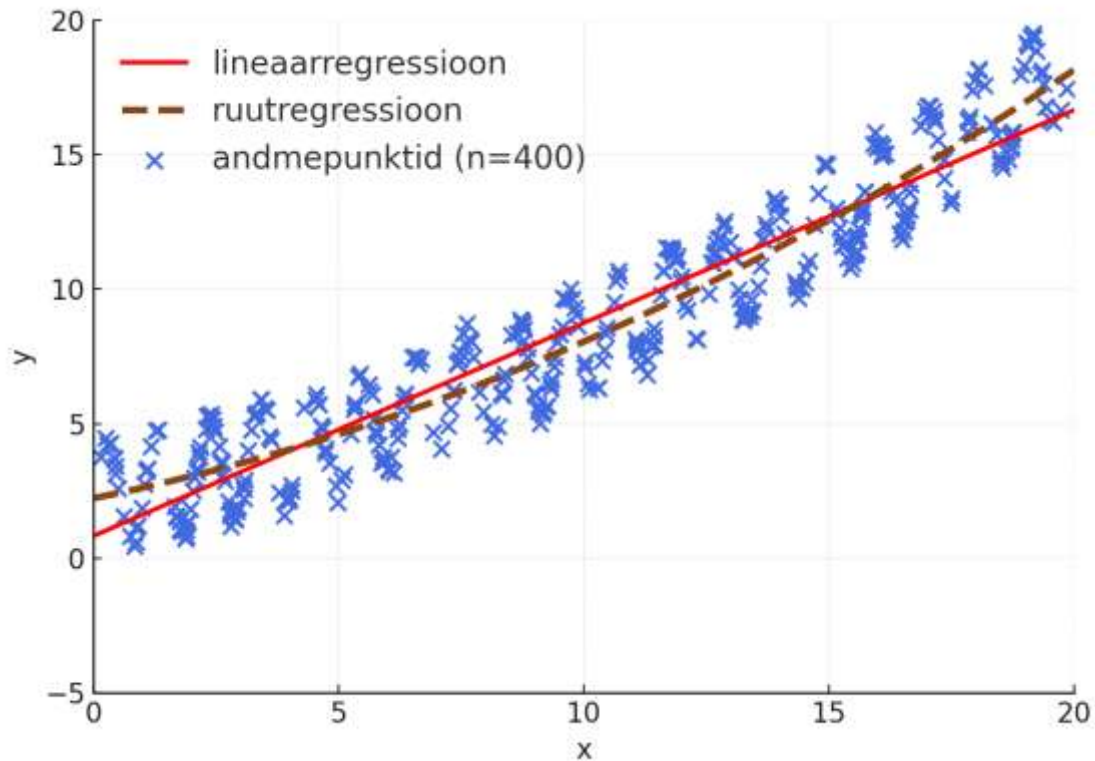


Probleem: ülesobitamine, kehv üldistusvõime

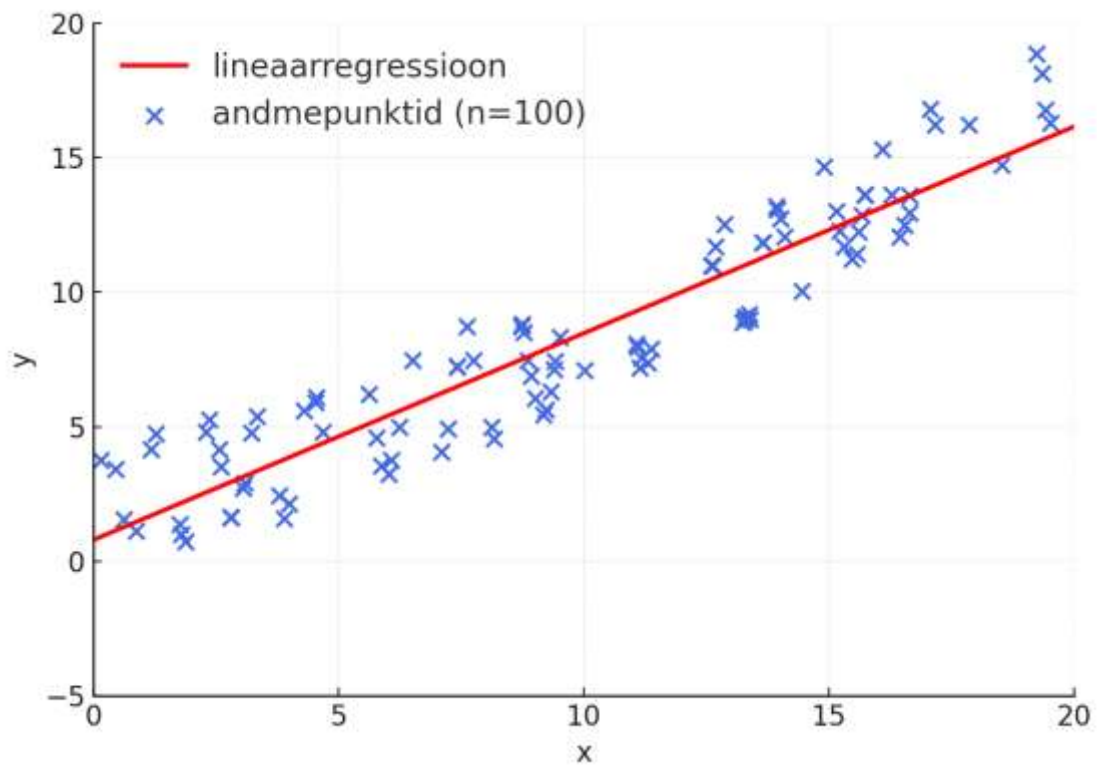
Lahendus: rohkem ja mitmekesisemad andmed või lihtsam mudel

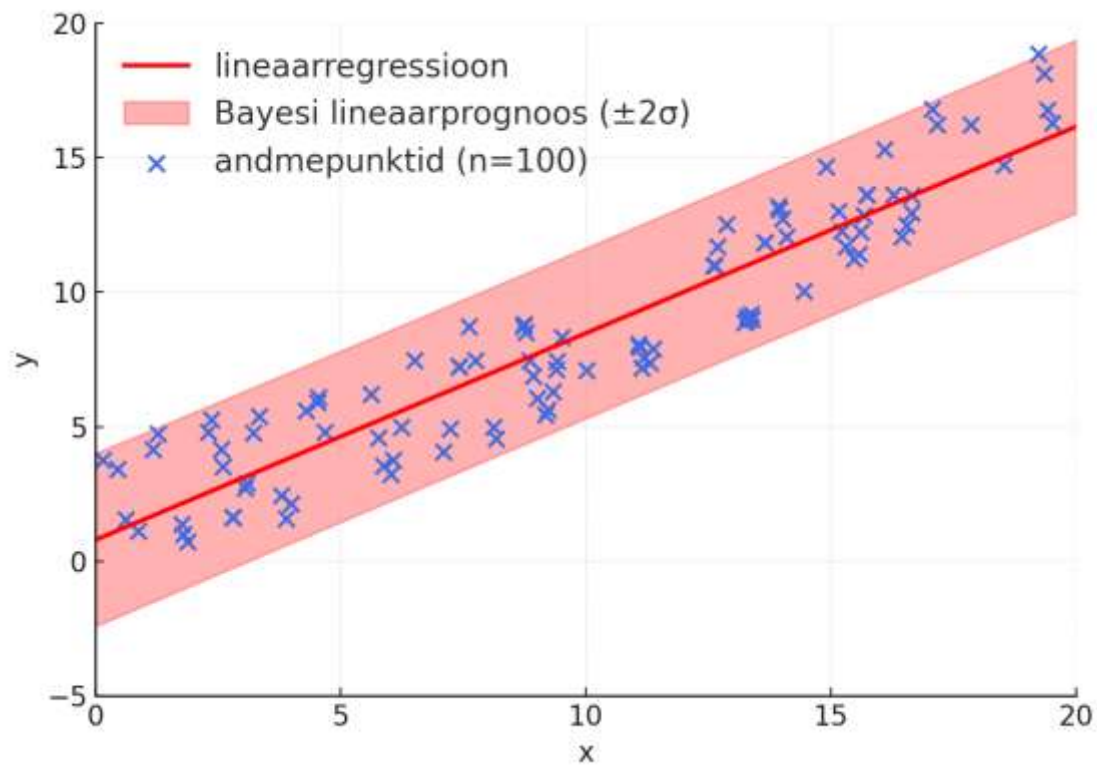


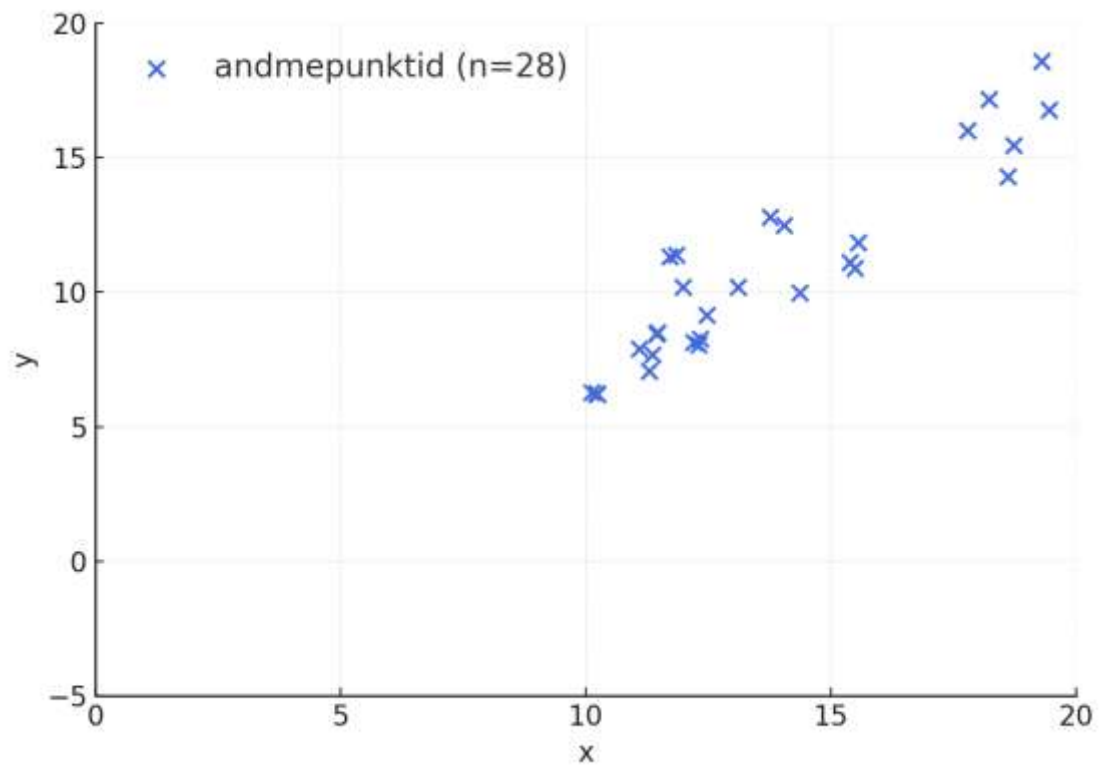
Probleem: ülesobitamine, kehv üldistusvõime (ka interpoleerides)
Lahendus: rohkem ja mitmekesisemad andmed või lihtsam mudel

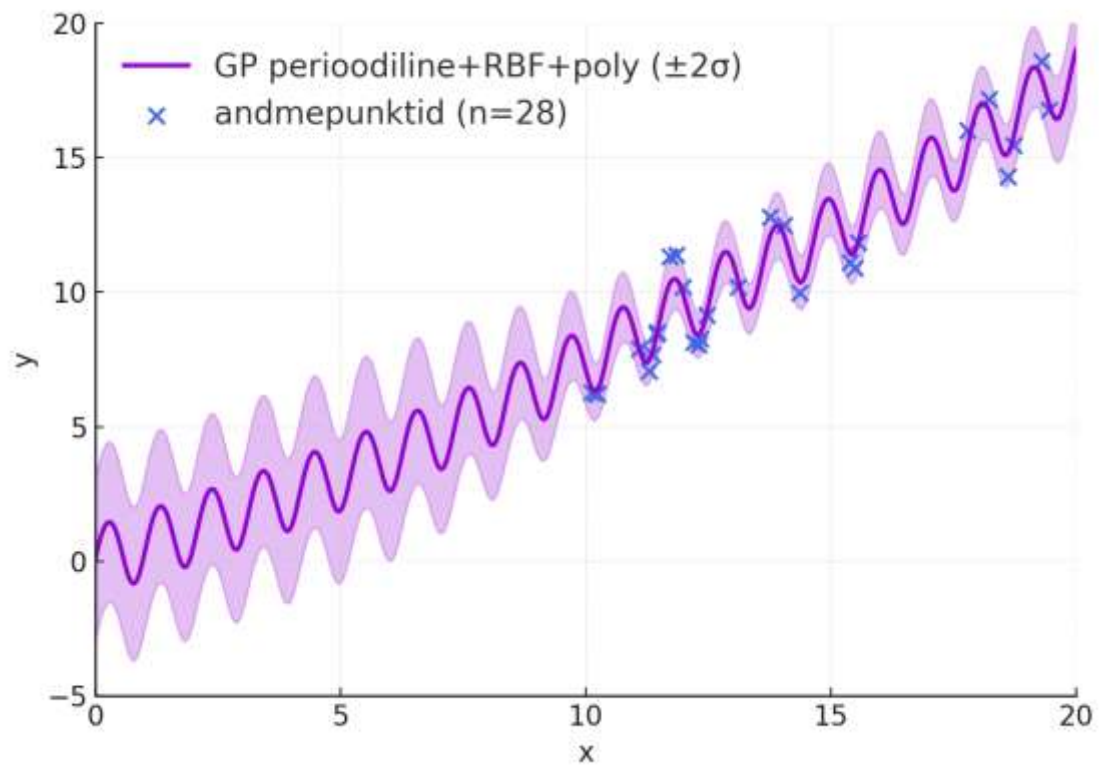


Probleem: alasobitamine, vähene sobivus treeningandmetega
Lahendus: suurem / keerulisem mudel









Ennustuste kvaliteedi probleemid masinõppes



- Kehva kvaliteediga ennustused ekstrapoleerimisel
Vaja mitmekesisemaid treeningandmeid
- Ülesobitamine, piiratud üldistusvõime
Rohkem ja mitmekesisemad andmed või lihtsam mudel
- Alasobitamine, vähene ennustusvõime ka treeningandmetel
Suurem / keerulisem mudel, arutlevam mudel
- Liigne enesekindlus nii ekstrapoleerides kui interpoleerides
Enesekindluse kalibreerimise meetodid (lahtine probleem)
- Info allikate teadvustamise probleemid, näiteks hallutsinatsioonid
Sisendi, mälu ja väljamõeldu sisemine eristamine (lahtine probleem)



Lahendused

- Rohkem ja mitmekesisemad treeningandmeid
- Suuremad mudelid
- Arutlevad mudelid
- Enesekindluse määramise meetodid
- Sisendi, mälu ja väljamõeldu sisemine eristamine

- Samas vaja keskkonna jalajälge vähendada

Millal tekstirobot eksib?

Millal tekstirobot eksib?



Millal tekstirobot eksib?

Siis kui ta suuda hästi järgmist tähte ennustada, sest kas





Millal tekstirobot eksib?

Siis kui ta suuda hästi järgmist tähte ennustada, sest kas

- (A) Mudelis puudub vajalik info



Millal tekstirobot eksib?

Siis kui ta suuda hästi järgmist tähte ennustada, sest kas

- (A) Mudelis puudub vajalik info

või

- (B) Piiratud mõttesammude arvu tõttu ei leia õiget infot



Millal tekstirobot eksib?

Siis kui ta suuda hästi järgmist tähte ennustada, sest kas

- (A) Mudelis puudub vajalik info
 - Puudub väga värsket info (peale mudeli treenimist)

või

- (B) Piiratud mõttesammude arvu tõttu ei leia õiget infot



Millal tekstirobot eksib?

Siis kui ta suuda hästi järgmist tähte ennustada, sest kas

- (A) Mudelis puudub vajalik info
 - Puudub väga värsket info (peale mudeli treenimist)
 - Mida vähem on faktid esinemisi internetis, seda tõenäolisemalt on mudelis see info osaline või puudu

või

- (B) Piiratud mõttesammude arvu tõttu ei leia õiget infot



Millal tekstirobot eksib?

Siis kui ta suuda hästi järgmist tähte ennustada, sest kas

- (A) Mudelis puudub vajalik info
 - Puudub väga värske info (peale mudeli treenimist)
 - Mida vähem on faktil esinemisi internetis, seda tõenäolisemalt on mudelis see info osaline või puudu

või

- (B) Piiratud mõttesammude arvu tõttu ei leia õiget infot

või

- (C) Mudelis on valeinfo (kas valesti õpitud või olid algallikad vigased)

Vajalikud tegevused Eestis edukaks tehisaru rakendamiseks



Vajalikud tegevused Eestis

- **Andmete kättesaadavaks tegemine tehisaru treenimiseks**
 - Välismaine treenimine
 - Eesti-ülene treenimine
 - Organisatsiooni-sisene peenhäälestus
- **Eesti-siseste arvutusressurside arendamine**
- **Välismaiste pilveteenuste ligipääsude saamine / hoidmine**
- **Teadustöö tehisaru alal ning tehisaru rakendamine teistes valdkondades**
- **Ekspertide koolitamine**
 - Doktorandid, magistrandid, bakalaureusetudengid, täiendõpe
- **Kasutajate harimine**

Eesti tehisintellekti tippkeskuse EXAI fookusteemad

Alusteaduslikud:

- Tehisintellekti teadmiste täiendamine
- Hü브리idsed süsteemid
- Ohutus ja usaldusväärsus
- Privaatsus ja turvalisus

Rakendusteaduslikud:

- Haridus
- Tervishoid
- Küberkaitse
- E-valitsemine
- Äriprotsessid



EXAI – Eesti tehisintellekti tippkesus



- Tartu Ülikool
 - 8 uurimisrühma
- Tallinna Tehnikaülikool
 - 4 uurimisrühma
- Cybernetica
 - 1 uurimisrühm

- 12 rühma arvutiteadusest
 - 3 keeletehnoloogia
 - 3 masinõpe, neuroteadus
 - 2 loogika ja hübriidne tehisaru
 - 2 turvalisus ja privaatsus
 - 2+mõned ülaltoodutest tegelevad rakendusvaldkondadega
- 1 rühm sotsiaalteadustest

13 uurimisrühma juhid:



Meelis
Kull



Tanel
Tammet



Liina
Kamm



Marlon
Dumas



Tanel
Alumäe



Mark
Fišel



Mihkel
Solvak



Jaak
Vilo



Rain
Ottis



Jaan
Aru



Paweł
Sobociński



Kairit
Sirts



Raul
Vicente



Kokkuvõte

- Tehisaru ja masinõpe
- Masinõpe ja tehisnärvivõrgud
- Kvaliteediprobleemid masinõppes
- Millal tekstirobot eksib
- Vajalikud tegevused Eestis

Aitäh!

- Teaduste Akadeemia tehisintellekti uurija-professor 2024-2025
- Eesti tehisintellekti tippkeskust EXAI rahastab Haridus- ja Teadusministeerium 2024-2030

Aitäh kuulamast!

