

Linær regresjoni

Eirik vetle Winnness

October 12, 2018

Abstract

I denne oppgaven eller eksprimentet skal jeg se på hvordan vi kan bruk linære reggresjon til å lage en model som kan tilsån at vi kan bruke tilgere data til å gjenskape en funksjon. Fungerer dette?. Slik resultane viser greier vi og oppnå noe lik form, men ikke eksakte like verdier som funksjonen selv har.

Introduction

I dette prosjektet skal jeg se på hvordan de tre forskjellige metodene OLS, ridge, og Lasso best greier å etterligne den ekte funksjonen (Frankfunksjon). Linær regresjon er en metode man kan bruke for å kunne spå hvordan ny data vil oppføre seg, basert på gammel data. Jeg skal studere MSE(), R², Bias, og. Jeg har skal studere random data ved å genere det i python programmet ankonda og dens pakker. Jeg skal så se på de fire faktorene og studere OLS, ridge og Lasso: Hvilken algoritme er det som greier å gi best resultat. Om noen av metodene er bedre til å etterligne datan. jeg kommer til å bruke trenings data for å lage en model, så ser jeg om denne modellen tilpasser seg godt opp mot den virkelige funksjonen.

I a,b og c bruker jeg k-fold for å splitte opp dataen i trening og test sett. Jeg splitter 1000 datapunkter inn i ti deler. Et test data sett, og 9 andre treningset. Når jeg analyser OLS ser jeg på de forskjellige treningsettene, mens i ridge og lasso tar jeg gjenommsnitte verdien (mse, r², bias, var), for se hvordan de oppfører seg i de forskjellige graden x^g .

Det blir litt anderledes når jeg ser på faktisk ekte data da velger jeg en del av et område i dataen. Jeg lager en model som produserer en β . jeg bruker en trenings verdi og en test verdi for \hat{X} . Nå som vi ikke har en funksjon då må vi bruke daten fra trenings sett som funksjonen. Hvor godt er det denne β -en greier å gjenskip den virkelige Z ($y = X\beta$). dette gjør jeg med 5 deler

fra dataen. Det er veldig tungt å gjøre dette det er derfor jeg ikke bruker så mange forskjellige treningsett. men velger heller å bruke det forskjellige z(områder) funksjonen. også varierer jeg gradene med(4-11)etter som jeg har set på verien og funnet ut av at dette er det området som gir greierst verdier(serlig r2) . håper ikke jeg har misforstått oppgaven. Det tok litt tid før jeg før jeg skjønnte at jeg skulle bruke den dataen vi lastet ned eller den dere ga oss verdiene til en funksjon.

Teori

MSE

MSE = Mean squared error

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2$$

$$MSE = Bias^2 + var(x)$$

R2

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1} (y_i - \bar{y})^2}$$

Gjennomsnittet av \hat{y} er definert som:

$$\bar{y} = \frac{1}{n} \sum_{i=0}^{n-1} (y_i)$$

$$\bar{y} = \frac{1}{n} \sum_{i=0}^{n-1} (y_i)$$

variancei

Varians kan bli sett på som hvordan et sett av random variabler er varierer i forhold til gjennomsnittet.

$$var(x) = (\bar{x^2}) - (\bar{x})^2$$

Bias

$$bias^2 = \left(\frac{1}{n} \sum_{i=1}^n (y - y_{\bar{s}_{pd}})^2 \right)^2$$

Ordinary Least squar

$$\hat{y} = \hat{X}\beta + e$$
$$\hat{\beta} = (\hat{X}^T \hat{X})^{-1} \hat{X}^T \bar{y}$$

Ridge Regression

Leas absolute shrinkage and select operator

$$h(b_0, b_1, \dots, b_k) = \sum_{i=1}^n [Y_i - (b_0 + b_1 x_{i1} + \dots + b_k x_{ik})]^2 + \lambda \sum_{j=1}^k |b_j|^2$$

$$\lambda = \text{tilpasningsverdi} = \text{konstant}$$

$$\hat{I} = \text{Identitetsmatrix}$$

$$\hat{\beta} = [\hat{X}^T \hat{X} + \lambda \hat{I}]^{-1} \hat{X}^T \hat{y}$$

Lasso Regression

Leas absolute shrinkage and select operator

$$h(b_0, b_1, \dots, b_k) = \sum_{i=1}^n [Y_i - (b_0 + b_1 x_{i1} + \dots + b_k x_{ik})]^2 + \lambda \sum_{j=1}^k |b_j|$$

Går ikke inn på denne teorien for den er ganske komplisert. I mine forsøk har jeg brukt `sklearn.linear model` linære model

K-fold

K-fold er en måte for å teste dataen model man har laget på data som man ikke har brukt for å lage modell. Da splitter man dataen man har opp i en hvis mengde deler. jeg splittet opp i 10 deler. Lot et sett være test data. Og lagde 9 andre forskjellige treningss data. Men fant ut senere at man skulle gjøre omvendt. At det nemlig var sån at man skulle lage en model og ha stor andel test data.

Ordinary Least squar

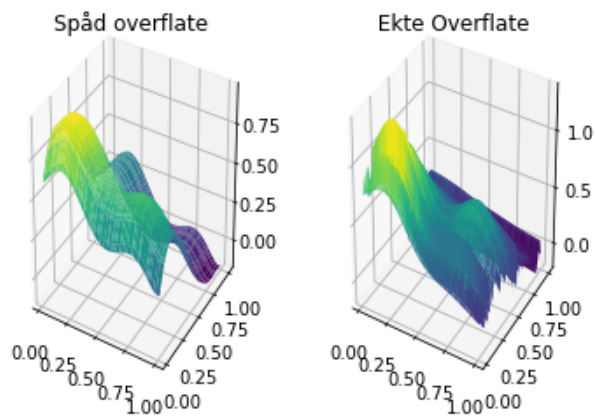


Figure 1: 3d plot. Det til vestre er spådde daen figur av grad 5. Figuren til høyre er ekte funksjonen

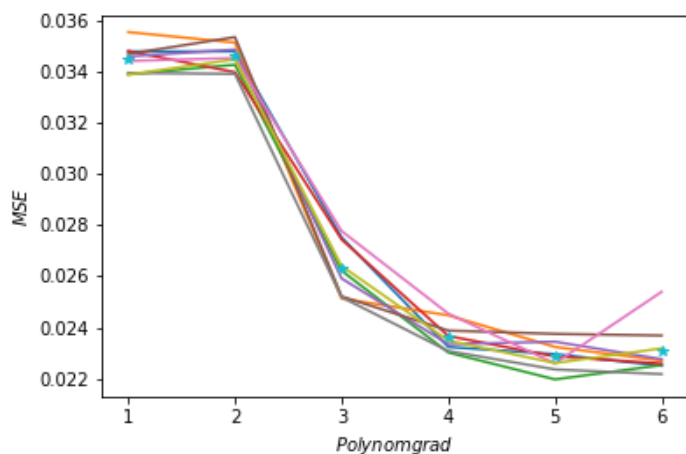


Figure 2: Her ser vi de forskjellige trenings datane MSE for de forskjellige gradene 1 til 6 (x^g, y^g). Der de b lyseblåe stjernene er gjennomsnitte av grad 1,2,...,6

På figur over ser vi de forskjellige verdien til MSE og gjenoomsnit verdien til de forskjellige treningses daten. vi kan se at vi har en veldig høy verdi for MSE,

men den blir lavere etter hvert som høyere grader av x og y blir lagt til i \hat{X} . Vi ser også at vi får en liten stigning fra grad 5 til 6. Det kan hende at MSE blir verre om vi legger til høyere grader av x og y. Vi skal ha en MSE verdi som er nærmest null. Gjennomsnittsverdien er cirka 0.225 noe som er ganske høyt.

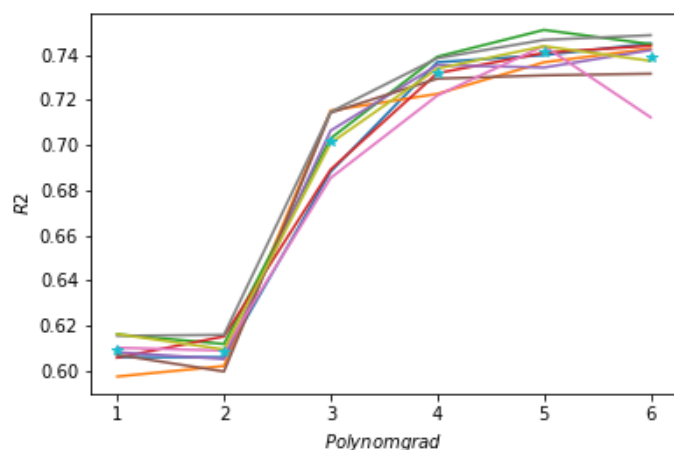


Figure 3: På denne figuren ser vi R2 skoren. Med trenings daten og gjennomsnitt verdien (blå stjerne)

Som forventet ser vi at R2 scoren blir bedre desto større graden på x og y. Scoren er ikke noe å skryte av, siden scoren skal være nærmest 1 som mulig. Og på figur 3 ser vi at flater ut ved 0.74 (grad 5, og 6)

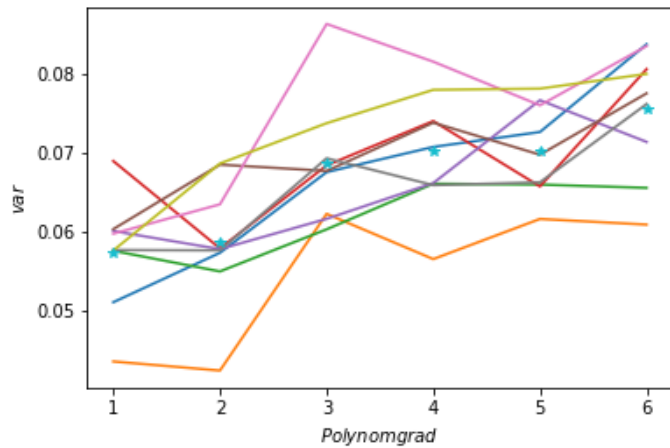


Figure 4: Her ser vi Variansen, for de forskjellige gradene 1 til 6(x^g, y^g) og gjennomsnitts verdien

Ser at variansen holder seg lav. Det er bra at variansen er positiv. Og at variansen lav betyr at datan ikke varierer så mye fra gjennomsnitte

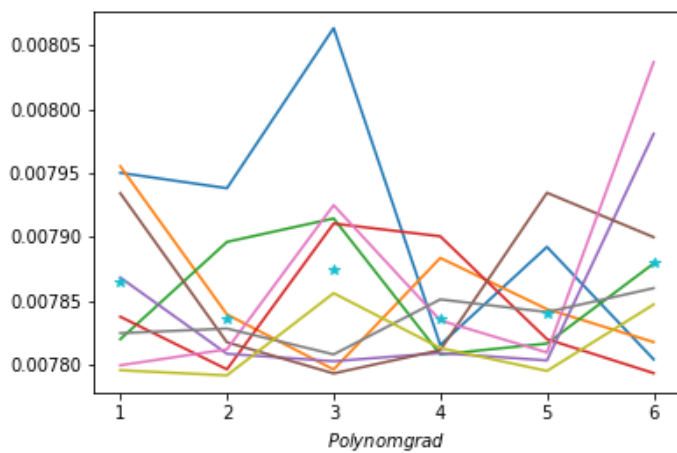


Figure 5: Her ser vi $Bias^2$, for de forskjellige gradene 1 til 6(x^g, y^g) og gjennomsnitts verdien

$Bias^2$ holder seg ganske likt over alle gradene. Bias er jo summen av forskjellen den spådegjennomsnitt verdien minus den ekte verdien min, delt på antalle verdier. Jeg ville jo tro att denne verdien ville synke voldsomt etter at man nærmer seg den ekte y-verdien.

Diskusjon

Vi vet at $MSE = Bias^2 + var$. Ut ifra grafene kan vi se at dette ikke stemmer med min data. Noe jeg synes er veldig rart. MSE og r2 har greier verdier. hvia vi ser på 3d bilde så ser vi at den har samme form som og ligner mye på den ekte funksjonen

Ridge Regression

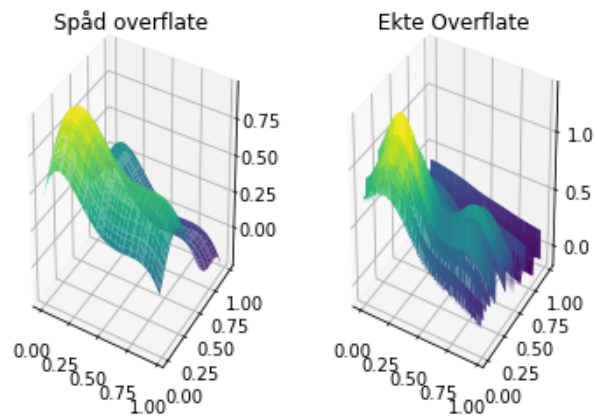


Figure 6: Her ser vi 3D plot av frankfunksjone med støy normalt fordelt.

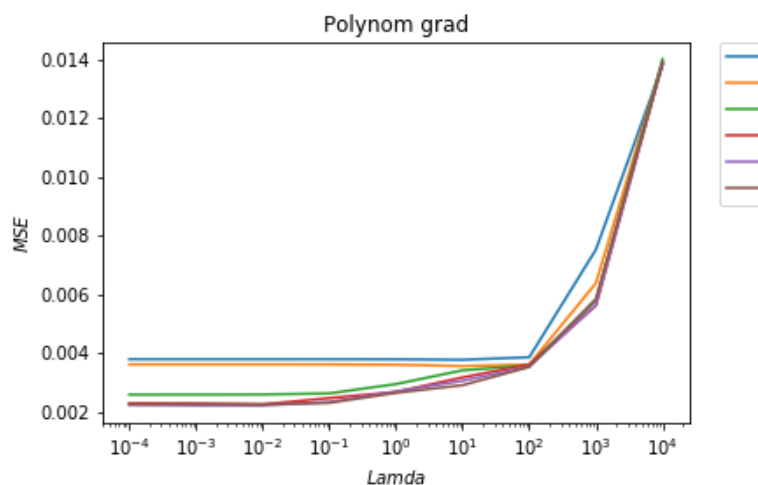


Figure 7: MSE vs $\log(\lambda)$, Gradene varier fra 1-6

Hvis vi ser på figure 7 kan vi se at jo høyere lamda verdier λ blir MSE verdiene høyere. Ser også at det er grad 5 og 6 som gjelder. Den halverer MSE scoren.

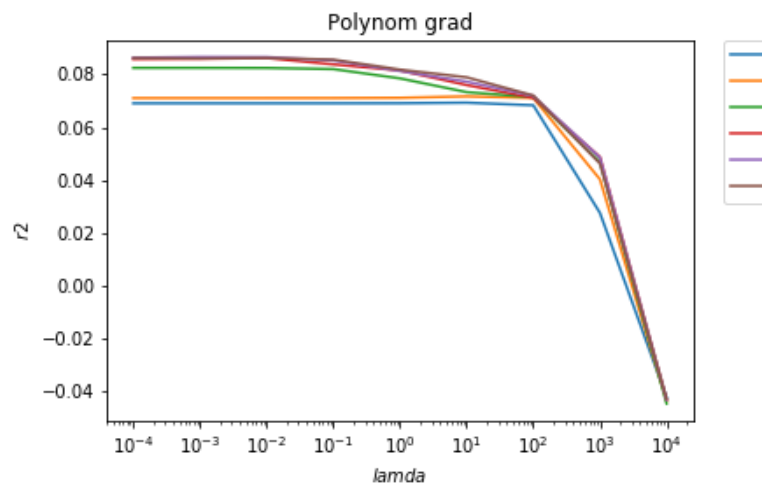


Figure 8: R2 vs $\log(\lambda)$, Gradene varier fra 1-6

R2 scoren oppfører seg gnskel likt so med MSE figur(8). Høye λ gir dårlig R2 score. Viser også at det er grad 5-6 som gir best resultat

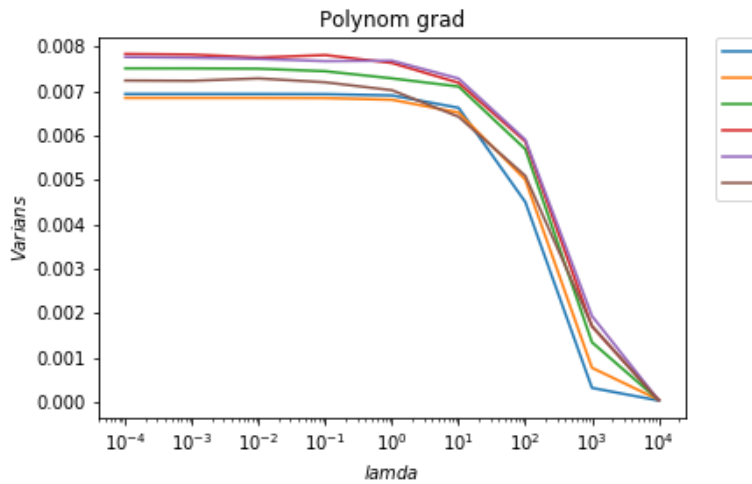


Figure 9: var vs $\log(\lambda)$, Gradene varier fra 1-6

På figur 9 kan vi se variansen for Rige. Den holder seg posetiv som er bra. Og veridene er veldig lave, noe som igjen tydeer på at verdien ikke varerer serlig fra gjennomsnittet. Noe som er veldig interesangt er at jo større λ blir mindre blir variangsen. Dette betyr kansje β -ene blir så store at de jevner ut Z . Hvis alle Z like vil jo variansen være null.

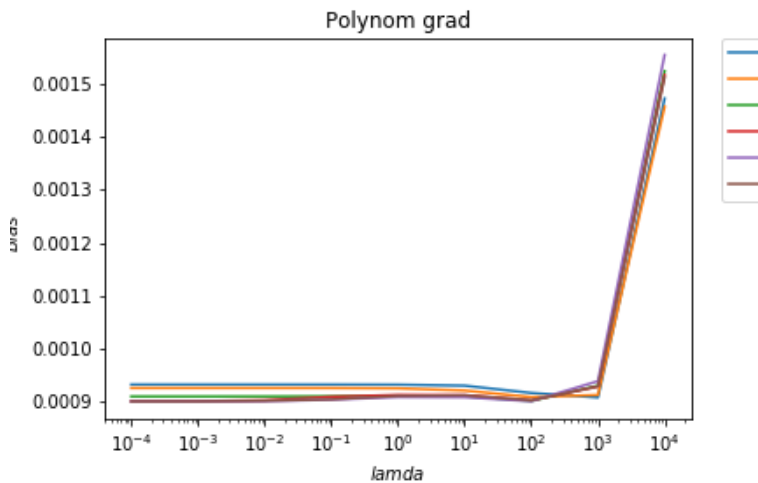


Figure 10: $Bias^2$ vs $\log(\lambda)$, Gradene varier fra 1-6
Bias holder seg meget la intill $\lambda = 10^3$

Diskusjon Rige

Her ser vi også at $MSE = Bias^2 + var(x)$ ikke holder. MSE holder seg meget lav rundt 0.003, mens varians ligger på 0.008 og $bias^2 = 0.0009$ spørålet her er jo at varians er 8 ganger større enn biasen. Dette tyder på at algorytmen overfitter. jeg har jo lagt normalfordelt støy på $0.1 \times sty$. hvis støy blir for stor vil man fitte mer etter denne enn selve de faktiske verdien.

Lasso Regression

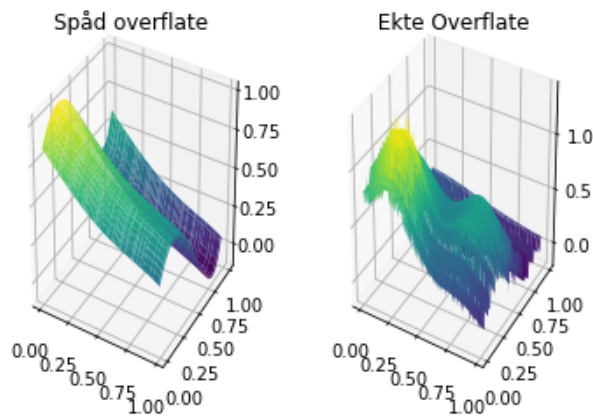


Figure 11: MSE vs $\log(\lambda)$, Grad varier fra 5.

Vi ser på fiur 11 att 3d plottet til Lasso ikke følger helt formen til den ekte overflaten. Vi kan se med en gang at det er en forskjell.

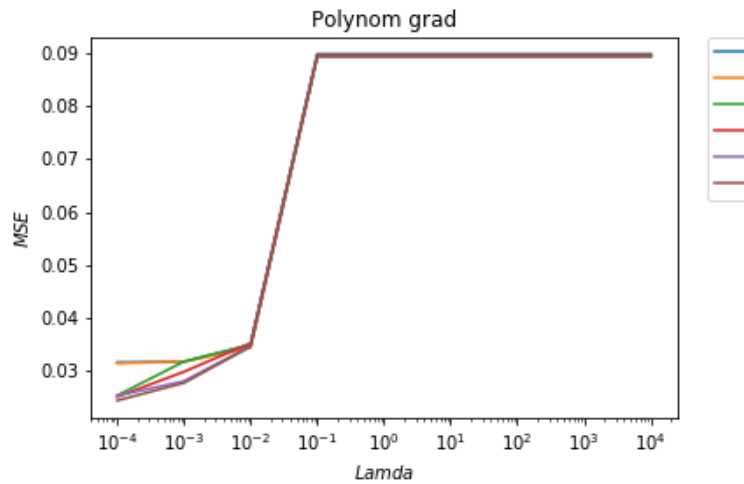


Figure 12: MSE vs $\log(\lambda)$, Gradene varier fra 1-6

Vi ser her også hvordan verdien er avhengig av λ . Svært små λ gir svært gode MSE verdier. Greie.Mello, 10^{-2} og 10^{-1} skjer det noe da får vi en svært hopp i MSE verdier fra rundt 0,03 til 0,09. 0,09 er jo fremdeles en veldig go skår

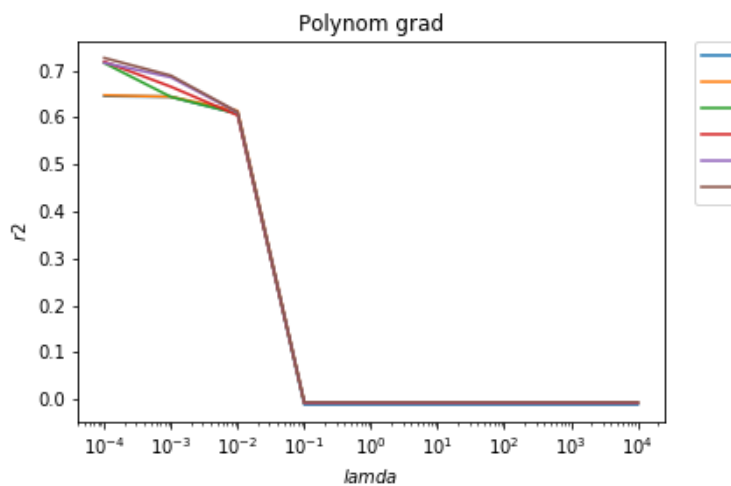


Figure 13: R^2 vs $\log(\lambda)$, Gradene varier fra 1-6

R^2 skåren til lasso kan vi sse på figur 13. Lasso har akseptable verdier intil $\lambda = 10^{-1}$. En skår mellom 0.7 og 0.6 er helt ok verdier.

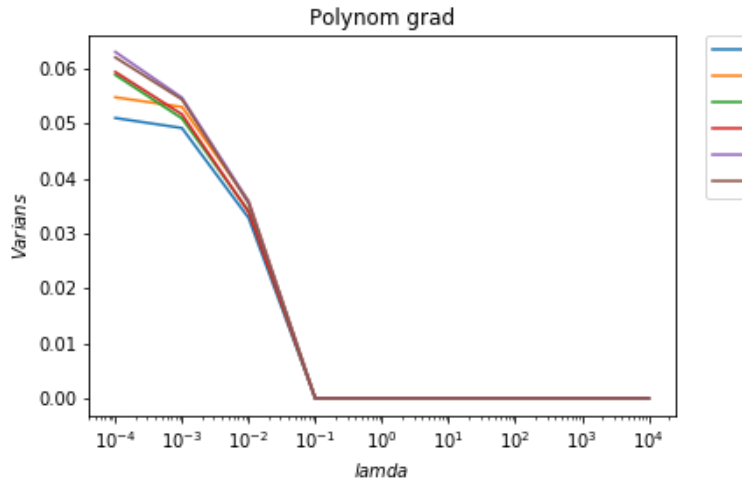


Figure 14: var vs $\log(\lambda)$, Gradene varier fra 1-6

Variansen(figur 14) følger r^2 scorens mønster ved at den gpr fa akseptable verdier til ikke akseptable verdier fra $\lambda 10^{-1}$

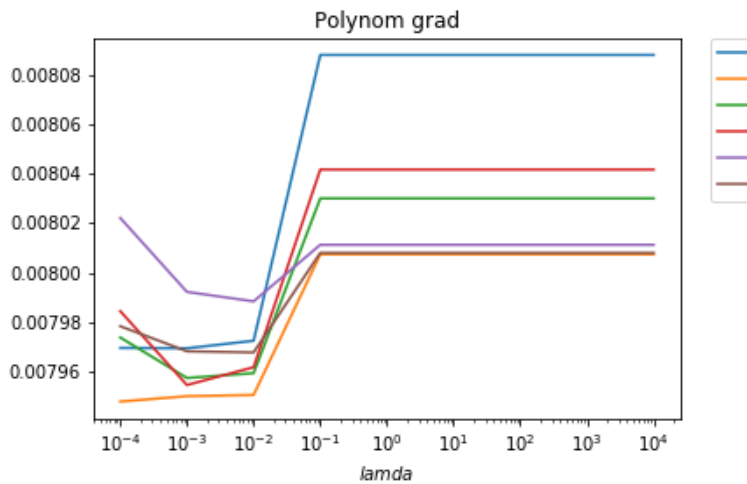


Figure 15: $Bias^2$ vs $\log(\lambda)$, Gradene varier fra 1-6

Bias for lasso på figur 15 følger også mønstret til var og r^2 . Den endrer seg ved 10^{-1} , men den er lavest ved $[10^{-3}, 10^{-2}]$. Den har en overaskende oppførsel etter som gradene ikke nødvendigvis har den lavest biasen, men nå betyr ikke det at den lavste bias er den beste tilpasningen.

diskusjon Lasso

Som vi ser så er ikke 3d projeksjonen noe serlig bra. den følger foren. men det virker som detaljen ike følger. Her har vi også noe som ligner å være en underfitt etter som variansen er mye større en bias. $MSE \neq bias^2 + var.$ Mulig jeg har gjort noe feil med lignigene?? Siden dette gjentar seg.

Diskusjon OLS vs Rige vs LASSO

Fra både verdier og grafisk 3d fremstilling, Så kommer Rige for meg klart på topp MSE og R2 verdier som ligger på 0.003 og 0.8. Er mye bedre enn hva OLS og LASSO kommer med. Rige er også mindre følsom når det kommer λ sine verdier. Selv om Rige har en laver $bias^2$ så virker ikke det som at det gjør at den legger mer merke til støyen. Ser nesten ut som det er lasso som blir mer påvirket av støyen. etter den får noe høyere på verdier på grafen. helt opp mot 1. Hvis jeg plotter lasso grad 10 så begynner den å nærme seg formen som Rige og ols klare og etterligne.

Real data

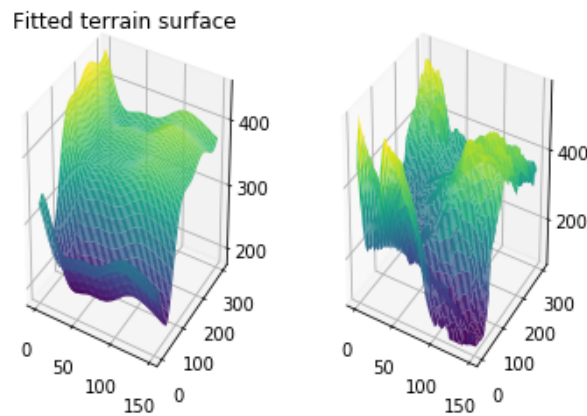


Figure 16: OLSE, grad 11, utrag 300×150

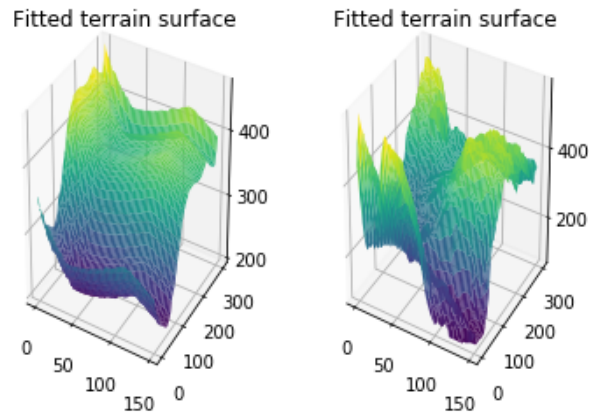


Figure 17: Riddge, grad11, utrag 300×150

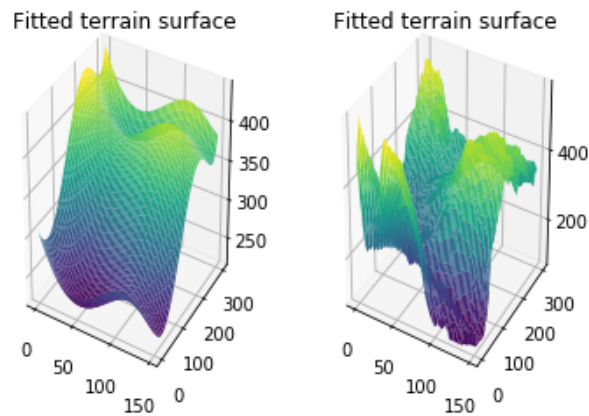


Figure 18: Lasso, grad11, utrag 300×150

Vi kan se OLS, Rige og lasso ikke like greit greier å fange detaljene til Delen jeg har valgt ut her. De greier å fange de viktige overflaten. Men detaljer og noe viktig informasjon forsvinner. Det er vanskelig å Bedømme 3d plottene. og bestteme hvem som er beast her.

MSE

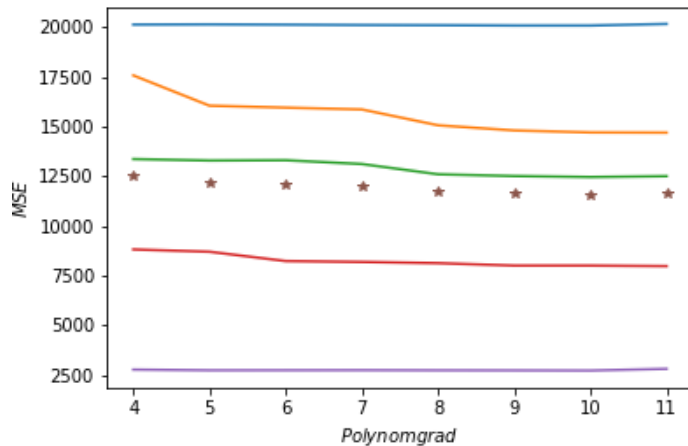


Figure 19: MSE OLS, brune stjerner er gjennomsnittet. for de forskjellige gradene

Verdien varierer mer med hvilken patch det vi har valgt isteden for, hvilken grad vi har valgt. Som vi vet skal jo MSE ha verdier som nærmer seg null. disse verdiene havner jo langt over 0. Så jeg vet egentlig ikke helt hva hvordan jeg skal tolke dette her.

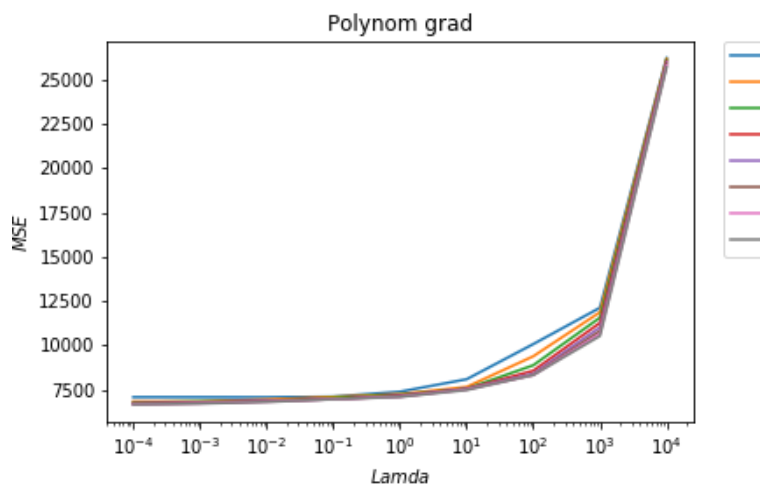


Figure 20: MSE av Rlge av 5 forskjellige deler av ekte data som jeg valgte

På Rlge MSE figur 20 ser vi samme mønstret som vi har sett tidligere bare

at vi har verdiene er mange ganger større verdien vi hadde på framfunksjonen. Så lenge λ -ene holder seg lave så er de stabiel. Det er heller ikke stor variasjon i gradene.

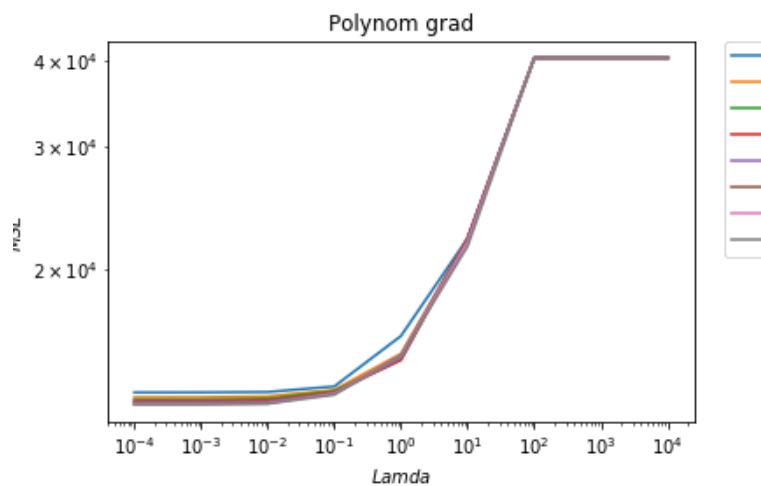


Figure 21: MSE vs λ . Tatt gjennomsnitte av de 5 forskjellige delene, for hver λ av ekte data som jeg valgte

Lasso har også verdier som er sky høye. Den har samme oppfører seg likt med λ . Den stiger kraftig når λ begynner å stige

R2

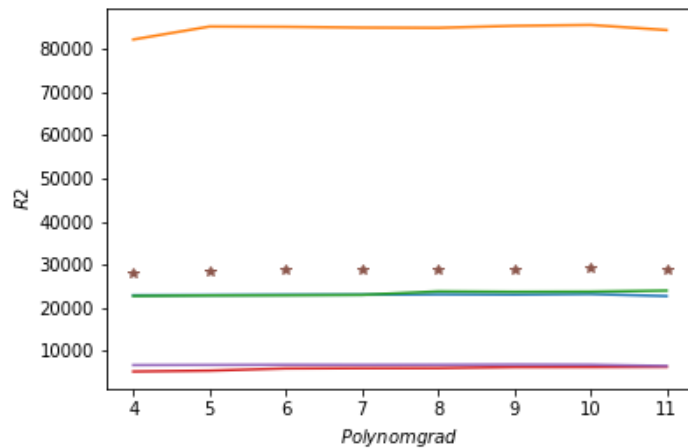


Figure 22: R2 vs graden til x, y

Verdie er ikke noe bedre her. De er skyhøye. Vet ikke hvordan jeg skal tolke det uten om at $y_i - \tilde{y}$ den diferangsen må være ganske stor og at $y - \bar{y}$ må være liten.

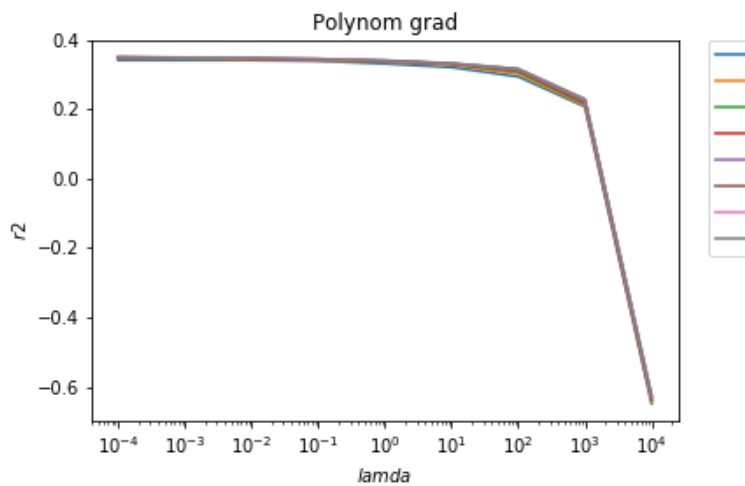


Figure 23: R2 vs $\log(\lambda)$, Gradene varier fra 4-11. Rige

Det er var jo ganske fasinerende å se at vi får en r2 verdi som faktisk er innefor en realistiske verider. De er jo ikke noe gode verdier. Det er ogsp ineresangt å se at når lambda blir stor så blir verdien negativ.

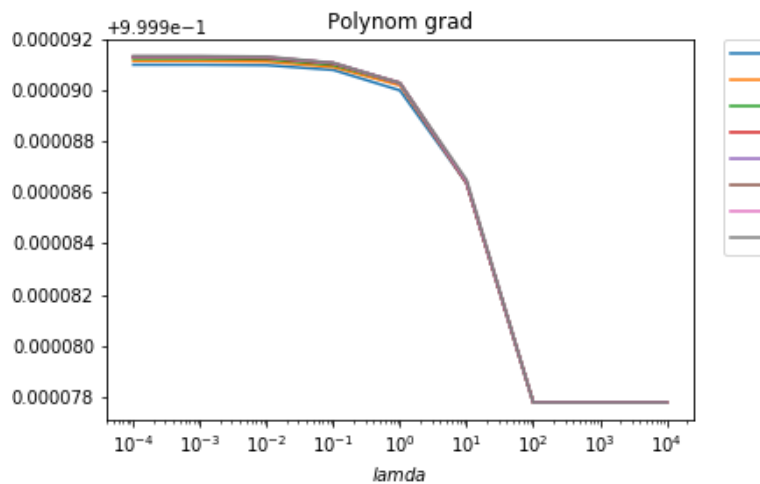


Figure 24: R^2 vs $\log(\lambda)$, Gradene varier fra 4-11. Lasso

Det er jo ganske fasinerende for her har vi faktisk verdier som er gode r^2 verdier figur 24. $0.9999 + 0.000092 = 0.999992$. Den oppfører seg også likt som i fraktfunksjonen hvor verdien faller. Men holder seg gode.

Varianc

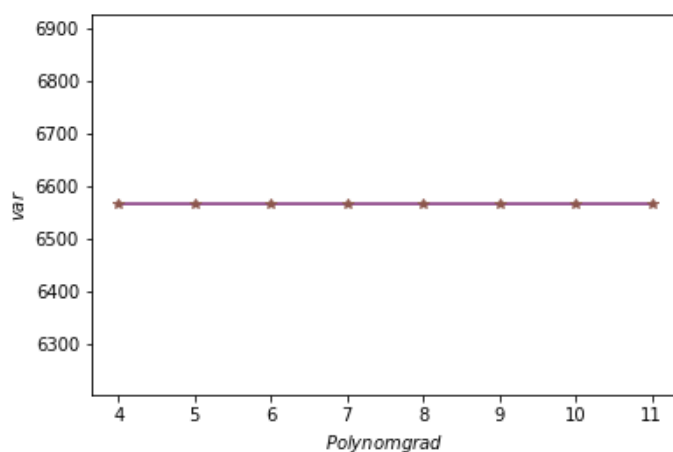


Figure 25: Varians vs gradenen til x , y , OLS

Sondig att variansen ikke varier??

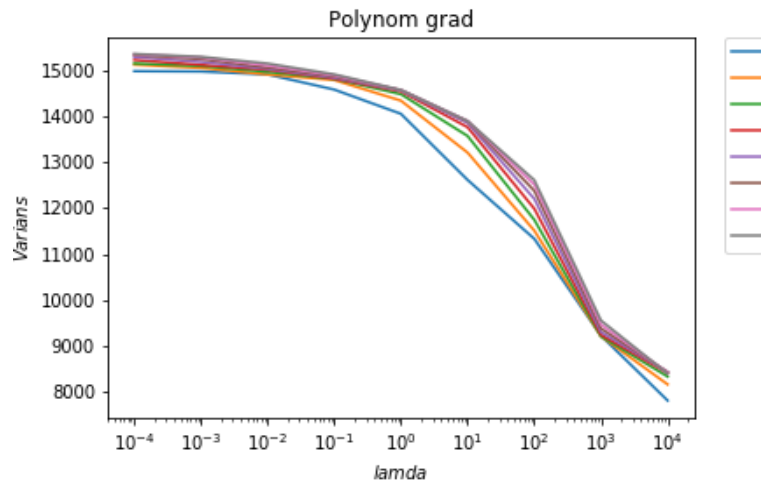


Figure 26: var vs $\log(\lambda)$, Gradene varier fra 4-11. Med Rige

Følger samme mønstret som frankfunsjonen, Når λ blir større synker variansen, men denne gangen noe sakkere.

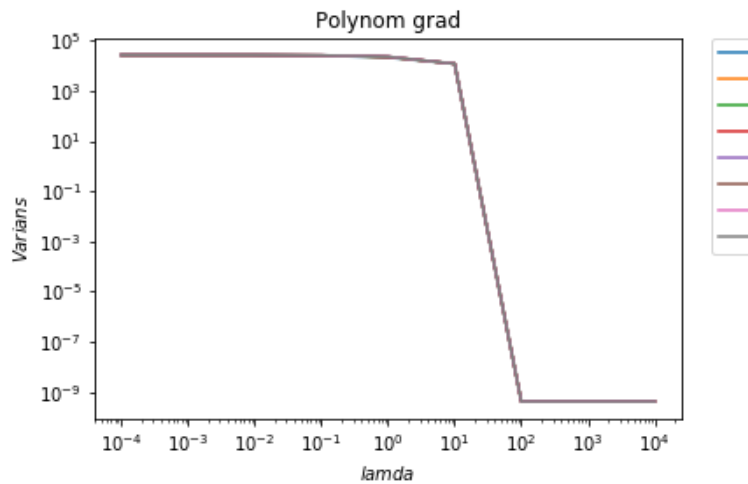


Figure 27: var vs $\log(\lambda)$, Gradene varier fra 4-11, Med Lasso

Følger samme mønstret som frankfunsjonen, Når λ blir større synker variansen, men den har en helt ekstrem endring fra helle 100000 til 0,0000009. Dette skjønner jeg faktisk. Dette kommer av at når λ blir stor vel den nominer så mye av at alle verdien til z blir like og vi ender med null varians. Så meget lav varians er faktisk en dårlig ting.

$Bias^2$

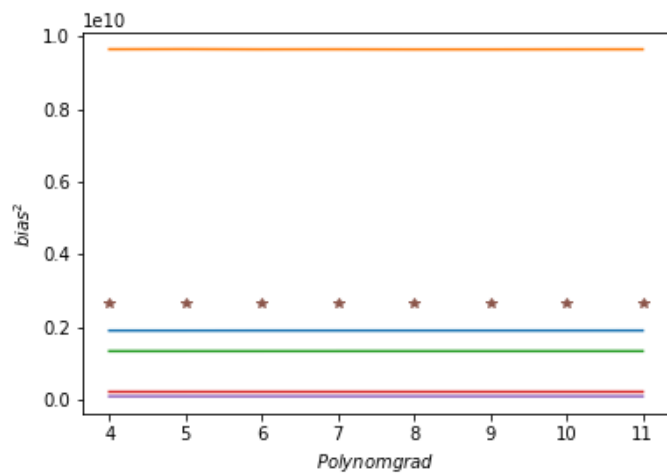


Figure 28: Bias vs Graden(x^g, y^g), $g=[4,11]$

Veldig høye verdier her. Det varier noen i fra data sett til data sett.

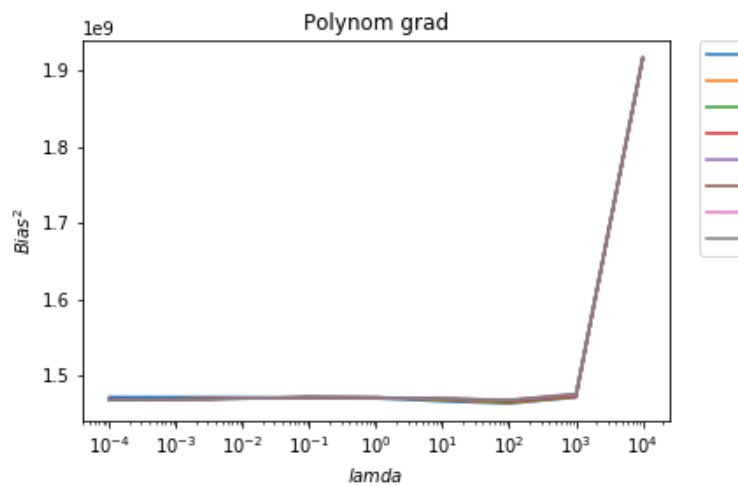


Figure 29: Bias2 vs λ , Rige. gradene går fra 4 til 11

Veldig høye verdier her

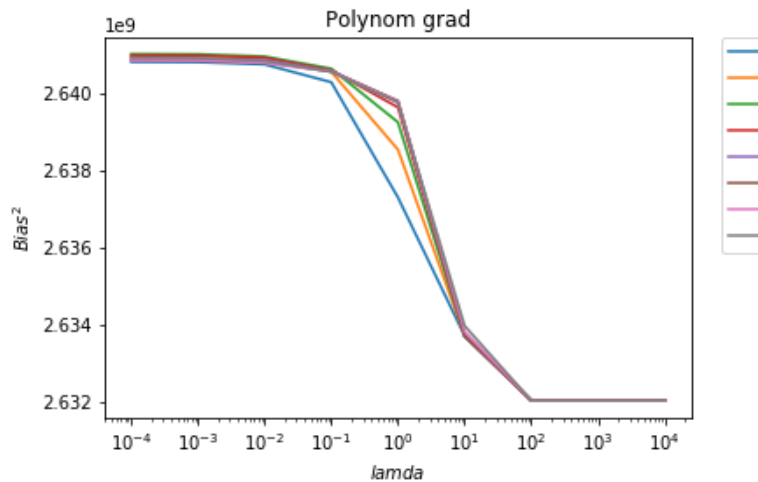


Figure 30: Bias2 vs λ , Lasso. gradene går fra 4 til 11

Sotore verdier på biasen her også

Diskusjon

Mye rare tall men at formen nogen lunde kommer frem er klart. Noen av tallen kan forklares, men jeg greier heller ikke her å få MSE som stemmer med bias og varians. Vi får så klart ekstremt forskjellige verdier fra den vi har delen fra datan, men viser at vi kan gjenskape følelse av den såkalte funksjonen til datan. Selv om vi bommer langt fra verdiene.

Diskusjon, data og frankfunksjon

Det jeg vet ikke helt hvordan jeg skal tolke disse tallene de er altfor høye, og det er veldig merkelig at r^2 verdien for lasso og ridge holder seg i det området som er akseptabelt. Vi kan se fra 3d plottene i figurene 16-18 at det har likheter med den originale delene. Men de er ganske langt fra originalen. Det er da som man kan forvente at verdien ikke er helt lik som den funksjonen. 3d grafen er jo ikke like.

Fra testen på frankfunksjon kan vi si lignet mer på den funksjonen

konklusjon

Jeg tror jeg kan komme frem til en konklusjon Ridge er mer stabil i forhold til det jeg har gjort. Dette baserer jeg på den har gode verdier på frankfunksjo-

nen, men også gode visuelle 3d plot. På ekte data har jo ridge nogenlunde likt 3d plot som ols og lasso men greie aksptable verdier på r^2 . Alle har syke verdier på real data. Men det virker som att det faktisk er mulig å kunne bruke denne metoden på for kunne spå $z(x,y)$. Men jeg kan ikke konkluder så veldig mye med de verdien jeg har fått På data delen. Noe jeg ikke skjønner hvorfor jeg ikke får til at $MSE = bias^2 + var$

jeg tror jeg også kan konkludere med empirisk av studert rige og lasso av at det lav λ som gir de beste verdien. Iverfall på de data delen jeg sett på og frankfunksjonen

hjelpemidler

<<https://compphysics.github.io/MachineLearning/doc/web/course.html>>

<<https://en.wikipedia.org/wiki/Bias\T1\textendashvariance-tradeoff>>

<<https://www.springer.com/gp/book/9780387848570f>>

<<https://piazza.com/class/ji78s1cduul39a?cid=75>> Mer på at jeg har brukt piatzza til løse oppgavene. A high-bias, low-variance introduction to Machine Learning for physicists Pankaj Mehta, Ching-Hao Wang, Alexandre G. R. Day, and Clint Richardson