Universidade Federal do Rio de Janeiro
Instituto Alberto Luiz Coimbra de
Pós-Graduação e Pesquisa de Engenharia

Departamento de Engenharia Elétrica

COE782 - Introdução ao Aprendizado de Máquina
Prof. Dr. Markus Vinícius Santos Lima

Lista 3 de exercícios

Luiz Henrique Souza Caldas
email: lhscaldas@cos.ufrj.br

9 de junho de 2024

1 Exercício 1

Considere o experimento computacional denominado "Polynomial Curve Fitting", usado diversas vezes no livro texto (veja páginas 4 e 5 do livro, bem como Apêndice A), considerando a ordem do modelo sendo M = 9 e o tamanho da amostra sendo N = 10. Faça:

(a) Calcule a solução de mínimos quadrados (LS) $\mathbf{w_{LS}}$;

(b) Calcule a solução via regressão ridge (escolha um fator de regularização razoável) $\mathbf{w_{ridge}}$;

(c) Calcule a solução via regressão lasso (escolha um fator de regularização razoável) $\mathbf{w_{lasso}}$;

(d) Monte uma tabela exibindo os 10 coeficientes $\mathbf{w}$ para as 3 soluções obtidas nos itens acima e comente/compare os resultados;

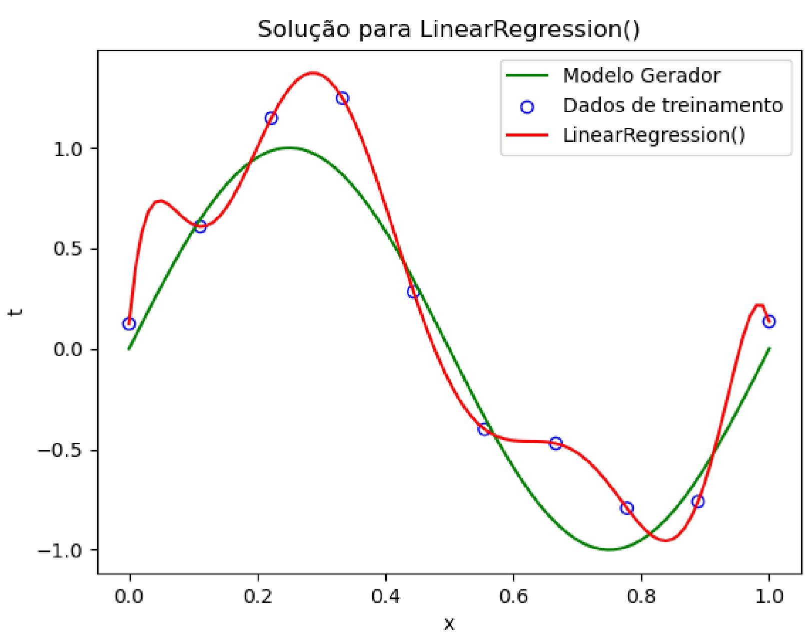
(e) Plote uma figura contendo o processo gerador em verde (a senoide), e suas estimativas y_{LS}, y_{ridge}, e y_{lasso} em preto, azul e vermelho, respectivamente;

(f) Repita todos os itens anteriores para $N = 20$ e $N = 50$.

1.1 Resposta do item (a)

Para responder esse item foi utilizada a classe *LinearRegression*() da biblioteca *sklearn* para linguagem *Python*, que realiza uma regressão linear utilizando mínimos quadrados. Foi escolhido um polinômio de ordem 9 como modelo.

Figura 1: Solução para mínimos quadrados (LS)

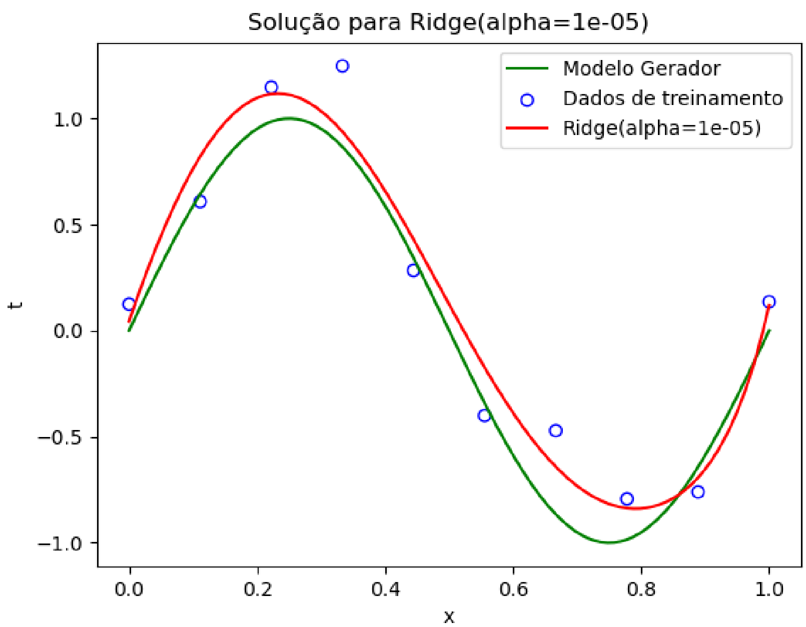


A regressão linear por mínimos quadrados não introduz nenhuma regularização e por isso a curva vermelha passa exatamente pelos dados de treinamento, mostrando que eles foram decorados overfitting.

1.2 Resposta do item (b)

Para responder esse item foi utilizada a classe *Ridge*() da biblioteca *sklearn* para linguagem *Python*, que realiza uma regressão linear utilizando mínimos quadrados e regularização de norma L_2 (Ridge). Foi escolhido um polinômio de ordem 9 como modelo e o fator de regularização λ, que na classe *Ridge*() é chamado de *alpha*, que melhor adaptou a curva ao modelo gerador foi $1^{-0.5}$.

Figura 2: Solução para Ridge com $\lambda = 1^{-0.5}$

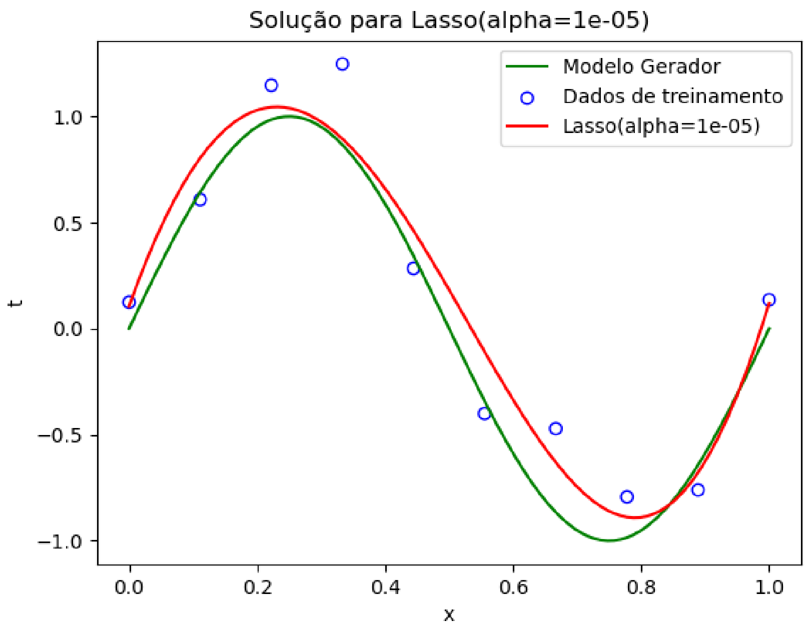


A regressão Ridge introduz uma penalização nos coeficientes através do fator de regularização, ajudando a evitar overfitting.

1.3 Resposta do item (c)

Para responder esse item foi utilizada a classe *Lasso*() da biblioteca *sklearn* para linguagem *Python*, que realiza uma regressão linear utilizando mínimos quadrados e regularização de norma L_1 (Lasso). Foi escolhido um polinômio de ordem 9 como modelo e o fator de regularização λ, que na classe *Lasso*() é chamado de *alpha*, que melhor adaptou a curva ao modelo gerador foi $1^{-0.5}$.

Figura 3: Solução para Lasso com $\lambda = 10^{-0.6}$



A regressão Lasso, assim como a Ridge, introduz uma penalização nos coeficientes através do fator de regularização, ajudando a evitar overfitting.

1.4 Resposta do item (d)

A tabela abaixo exibindo os coeficientes para as três soluções permite comparar diretamente o impacto da regularização nos coeficientes.

Tabela 1: Coeficientes para $N = 10$

Modelo	w0	w1	w2	w3	w4	w5	w6	w7	w8	w9
LS	0.00	33.81	-652.62	5410.54	-21973.73	48026.65	-58345.99	37810.21	-10973.23	664.37
Ridge	0.00	8.87	-14.84	-21.96	25.51	24.50	-3.61	-23.28	-16.17	21.04
Lasso	0.00	8.50	-20.39	3.66	5.68	3.20	0.97	-0.19	-0.66	-0.76

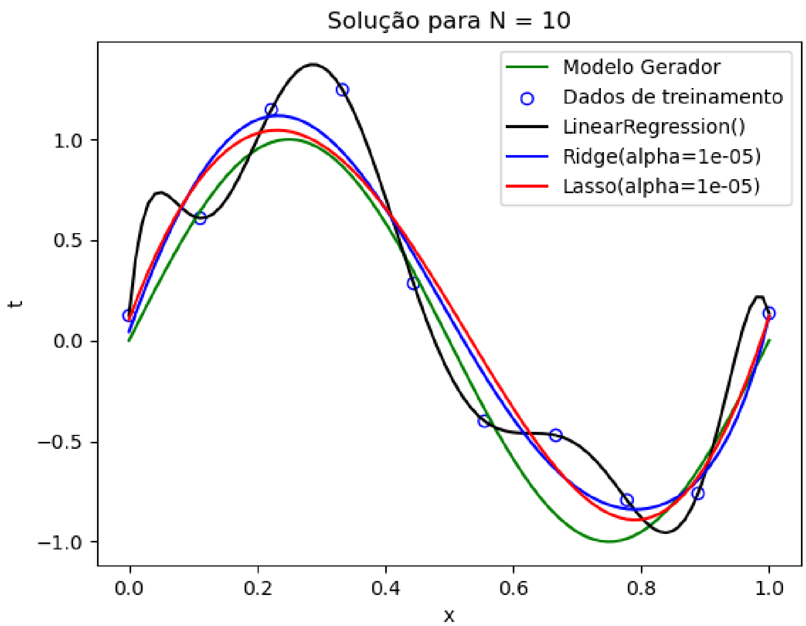
A regressão por mínimos quadrados sem regularização tende a produzir coeficientes maiores, um dos sinais que indicam overfitting ou pelo menos uma alta sensibilidade aos altos de treinamento. Enquanto que os coeficientes produzidos pela Ridge e pela Lasso são menores.

Além disso, é possível observar a presença de alguns coeficientes praticamente nulos para a Lasso. Esse resultado era esperado, uma vez que a Lasso, por utilizar a norma L_1, tende a produzir uma seleção mais esparsa, selecionando atributos mais importantes.

Enquanto isso, a regressão Ridge, apesar de não selecionar atributos como a Lasso, tende a penalizar, e consequentemente reduzir, ainda mais os coeficientes grandes.

1.5 Resposta do item (e)

Figura 4: Comparação entre as soluções para $N = 10$



1.6 Resposta do item (f)

Figura 5: Comparação entre as soluções para $N = 20$

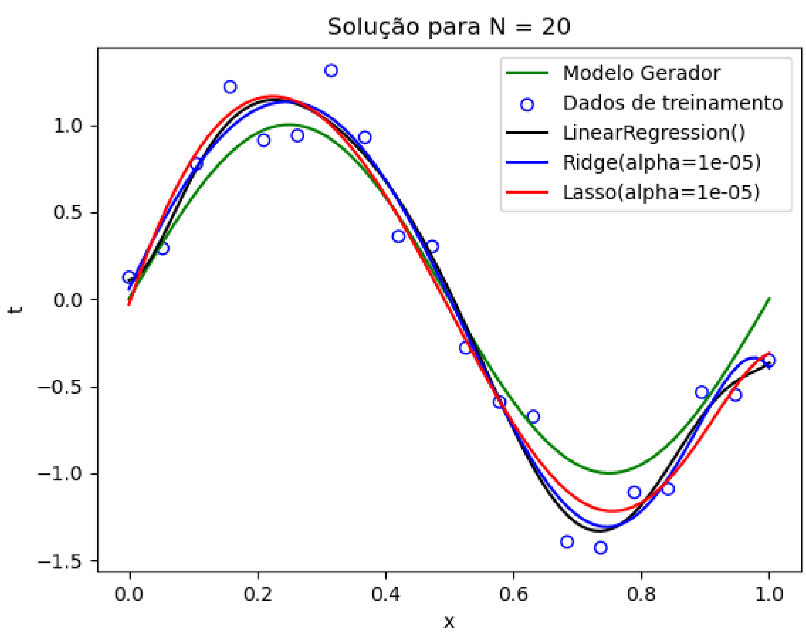


Tabela 2: Coeficientes para $N = 20$

Modelo	w0	w1	w2	w3	w4	w5	w6	w7	w8	w9
LS	0.00	0.13	133.29	-1042.60	3592.94	-6598.01	6040.15	-1620.50	-1137.52	631.65
Ridge	0.00	7.94	-10.61	-17.66	2.91	14.55	13.52	4.83	-5.11	-10.82
Lasso	0.00	11.00	-26.25	2.05	8.34	7.01	3.69	0.35	-2.11	-4.36

Figura 6: Comparação entre as soluções para $N = 50$

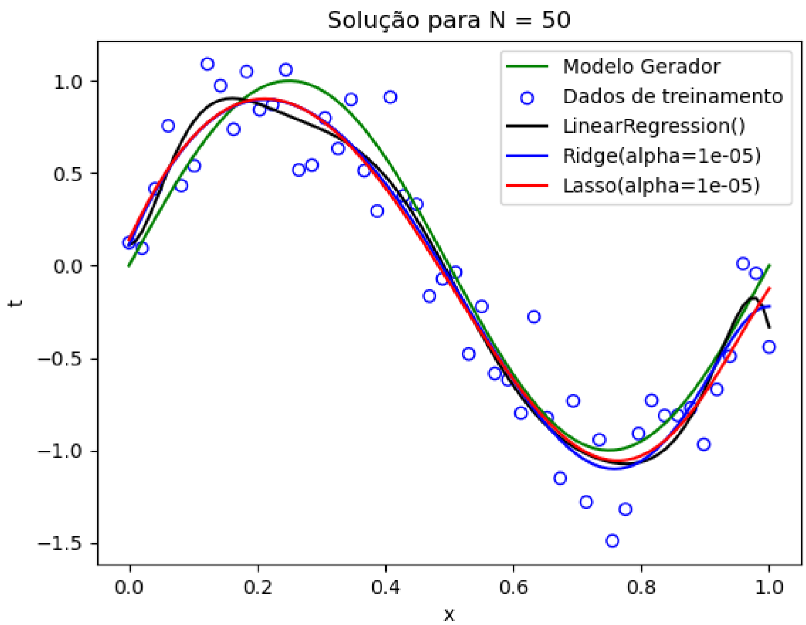


Tabela 3: Coeficientes para $N = 50$

Modelo	w0	w1	w2	w3	w4	w5	w6	w7	w8	w9
LS	0.00	-0.06	228.03	-2656.39	13838.07	-40023.24	67930.77	-67287.01	36037.64	-8068.26
Ridge	0.00	8.06	-23.82	17.17	-5.30	-8.80	5.11	16.26	10.42	-19.44
Lasso	0.00	7.42	-18.67	1.59	5.49	4.71	2.67	0.51	-1.09	-2.90

2 Exercício 2

Data Source:

- (info) `https://web.stanford.edu/~hastie/ElemStatLearn/datasets/prostate.info.txt`
- (database) `https://web.stanford.edu/~hastie/ElemStatLearn/datasets/prostate.data`

"The data for this example come from a study by Stamey et al. (1989). They examined the correlation between the level of prostate-specific antigen (lpsa) and a number of clinical measures in men who were about to receive a radical prostatectomy. The variables are log cancer volume (lcavol), log prostate weight (lweight), age, log of the amount of benign prostatic hyperplasia (lbph), seminal vesicle invasion (svi), log of capsular penetration (lcp), Gleason score (gleason), and percent of Gleason scores 4 or 5 (pgg45)."

Considere a variável (lpsa) como 'target' e as variáveis (lcavol), (lweight), (age), (lbph), (svi), (lcp), (gleason) e (pgg45) como 'entradas'. Siga o roteiro abaixo:

(a) Padronize os atributos de entrada para que eles tenham média 0 e variância 1;

(b) Divida o dataset em dois conjuntos, treinamento e teste, conforme indicado nos índices da última coluna (T = treinamento, F = teste);

(c) Encontre o modelo linear de regressão ótimo no critério de mínimos quadrados (solução LS);

(d) Implemente modelos lineares regularizados pelos métodos 'Ridge' e 'Lasso' que minimizam a função objetivo $L(w) = \frac{1}{2N}RSS(w) + \lambda||w||^q$. Apresente resultados para $\lambda = 0.25$;

(e) Aplicando as regressões 'Ridge' e 'Lasso' e utilizando k-fold cross-validation, é possível selecionar um valor para λ que resulta em um modelo com melhor capacidade de generalização. Isso é feito selecionando o λ relativo à menor estimativa do erro de predição quadrático médio (usualmente chamado de validation score) ao longo dos k-folds. Também é possível selecionar um valor de λ que seleciona o modelo mais simples dentro de uma tolerância da estimativa do erro de predição quadrático médio. Isso é particularmente útil quando se deseja encontrar soluções esparsas (no caso do Lasso) ou de menor norma L2 (no caso do Ridge). Para tal, um critério comumente adotado é a 'Regra de 1 desvio padrão', onde escolhe-se o maior λ cujo validation score seja igual ou pouco menor do que o 'score mínimo' + '1 desvio padrão do score mínimo'.

- Monte as curvas de validation score de k-fold cross-validation em função de λ para os modelos regularizados por 'ridge' e 'lasso' (Sugestão: use k = 10, e procure λ em um intervalo [0, 0.5]);
- Calcule o desvio padrão do 'score' mínimo em cada respectiva curva e desenhe-o como barra de erro em torno daquele ponto;
- Determine o λ que resulta no modelo mais simples de acordo com a 'Regra de 1 desvio padrão';
- Treine o modelo final 'ridge' e 'lasso' utilizando todos os dados (de treinamento) e o respectivo λ encontrado e apresente os resultados;

(f) Utilizando o conjunto de teste construído no item (b), calcule a estimativa do erro de predição quadrático médio do conjunto de teste para cada modelo (mínimos quadrados, 'ridge' e 'lasso'). Disserte sobre os resultados obtidos.

(g) (Bônus) Estime o desvio padrão dos coeficientes do modelo obtido pelo método de bootstrap dos resíduos;

Dica: Veja os slides 9 e 10 de `http://www.est.ufmg.br/~cristianocs/MetComput/Aula8.pdf`

Códigos