Chapter 17: Lasso, Ridge, and Elastic Net

Modern Clinical Data Science
Chapter Guides
Bethany Percha, Instructor

How to Use this Guide

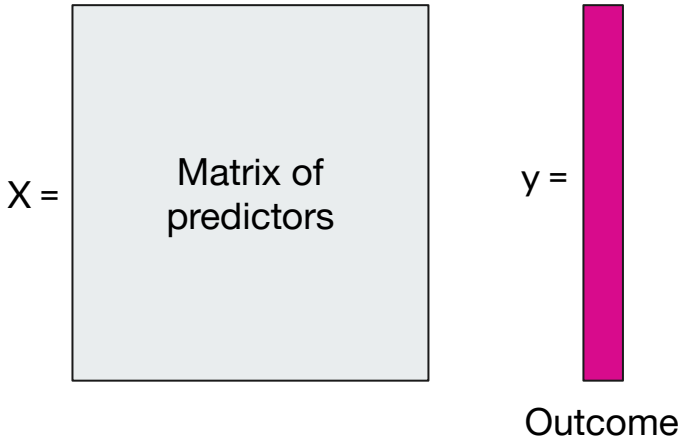
- Read the corresponding notes chapter first
- Try to answer the discussion questions on your own
- Listen to the chapter guide (should be 30 min, max) while following along in the notes

Every supervised learning method has three things:

1. Loss function (objective function)

2. Search space (set of possible mappings)

3. Optimization algorithm



Revisiting linear and logistic regression

$$\text{loss} = \sum_{i=1}^{N} (y^{(i)} - \beta^T x^{(i)})^2$$

$$\text{loss} = -\sum_{i=1}^{n} \left[y^{(i)} \beta^T x^{(i)} - \log\left(1 + \exp(\beta^T x^{(i)})\right) \right]$$

$$\text{penalty}_{\text{Lasso}} = \lambda \sum_{j=1}^{p} |\beta_j|$$

$$\text{penalty}_{\text{ridge}} = \lambda \sum_{j=1}^{p} \beta_j^2$$

$$\text{penalty}_{\text{EN}} = \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2 \right]$$

Example: Predicting blood pressure

	sexf	sexm	age	obesity	blood_pressure
1	0	1	54	1	123
2	1	0	66	0	111
3	1	0	23	0	98
4	0	1	59	1	154
5	0	1	76	1	199
6	1	0	33	0	101
7	0	1	35	1	91
8	1	0	54	0	133
9	1	0	21	0	116
10	0	1	26	0	121

Question 17.3

Which values for λ and α correspond to: (a) unregularized linear regression, (b) pure Lasso, (c) pure ridge, (d) elastic net with an even combination of Lasso and ridge penalties? What happens to the values of the coefficients in each case?

	lambda	alpha	beta_sexf	beta_sexm	beta_age	beta_obesity
1	0.0	0.0	-24.61	-0.00	1.15	-13.78
2	1.0	0.0	-11.19	10.71	1.09	-9.97
3	5.0	0.0	-8.25	8.24	0.91	-2.11
4	10.0	0.0	-6.97	6.99	0.77	1.66
5	0.0	0.2	-24.61	-0.00	1.15	-13.78
6	1.0	0.2	-10.35	9.73	1.07	-7.84
7	5.0	0.2	-6.81	6.72	0.88	0.00
8	10.0	0.2	-6.10	6.07	0.76	0.00
9	0.0	0.5	-24.61	-0.00	1.15	-13.78
10	1.0	0.5	-9.09	8.04	1.04	-4.38
11	5.0	0.5	-5.65	5.50	0.87	0.00
12	10.0	0.5	-3.84	3.77	0.71	0.00
13	0.0	1.0	-24.61	-0.00	1.15	-13.78
14	1.0	1.0	-13.15	0.00	1.00	0.00
15	5.0	1.0	-6.93	0.00	0.84	0.00
16	10.0	1.0	0.00	0.00	0.62	0.00

Question 17.3

Which values for λ and α correspond to: (a) unregularized linear regression, (b) pure Lasso, (c) pure ridge, (d) elastic net with an even combination of Lasso and ridge penalties? What happens to the values of the coefficients in each case?

	lambda	alpha	beta_sexf	beta_sexm	beta_age	beta_obesity
1	0.0	0.0	-24.61	-0.00	1.15	-13.78
2	1.0	0.0	-11.19	10.71	1.09	-9.97
3	5.0	0.0	-8.25	8.24	0.91	-2.11
4	10.0	0.0	-6.97	6.99	0.77	1.66
5	0.0	0.2	-24.61	-0.00	1.15	-13.78
6	1.0	0.2	-10.35	9.73	1.07	-7.84
7	5.0	0.2	-6.81	6.72	0.88	0.00
8	10.0	0.2	-6.10	6.07	0.76	0.00
9	0.0	0.5	-24.61	-0.00	1.15	-13.78
10	1.0	0.5	-9.09	8.04	1.04	-4.38
11	5.0	0.5	-5.65	5.50	0.87	0.00
12	10.0	0.5	-3.84	3.77	0.71	0.00
13	0.0	1.0	-24.61	-0.00	1.15	-13.78
14	1.0	1.0	-13.15	0.00	1.00	0.00
15	5.0	1.0	-6.93	0.00	0.84	0.00
16	10.0	1.0	0.00	0.00	0.62	0.00

Question 17.3

Which values for λ and α correspond to: (a) unregularized linear regression, (b) pure Lasso, (c) pure ridge, (d) elastic net with an even combination of Lasso and ridge penalties? What happens to the values of the coefficients in each case?

	lambda	alpha	beta_sexf	beta_sexm	beta_age	beta_obesity
1	0.0	0.0	-24.61	-0.00	1.15	-13.78
2	1.0	0.0	-11.19	10.71	1.09	-9.97
3	5.0	0.0	-8.25	8.24	0.91	-2.11
4	10.0	0.0	-6.97	6.99	0.77	1.66
5	0.0	0.2	-24.61	-0.00	1.15	-13.78
6	1.0	0.2	-10.35	9.73	1.07	-7.84
7	5.0	0.2	-6.81	6.72	0.88	0.00
8	10.0	0.2	-6.10	6.07	0.76	0.00
9	0.0	0.5	-24.61	-0.00	1.15	-13.78
10	1.0	0.5	-9.09	8.04	1.04	-4.38
11	5.0	0.5	-5.65	5.50	0.87	0.00
12	10.0	0.5	-3.84	3.77	0.71	0.00
13	0.0	1.0	-24.61	-0.00	1.15	-13.78
14	1.0	1.0	-13.15	0.00	1.00	0.00
15	5.0	1.0	-6.93	0.00	0.84	0.00
16	10.0	1.0	0.00	0.00	0.62	0.00

Question 17.3

Which values for λ and α correspond to: (a) unregularized linear regression, (b) pure Lasso, (c) pure ridge, (d) elastic net with an even combination of Lasso and ridge penalties? What happens to the values of the coefficients in each case?

	lambda	alpha	beta_sexf	beta_sexm	beta_age	beta_obesity
1	0.0	0.0	-24.61	-0.00	1.15	-13.78
2	1.0	0.0	-11.19	10.71	1.09	-9.97
3	5.0	0.0	-8.25	8.24	0.91	-2.11
4	10.0	0.0	-6.97	6.99	0.77	1.66
5	0.0	0.2	-24.61	-0.00	1.15	-13.78
6	1.0	0.2	-10.35	9.73	1.07	-7.84
7	5.0	0.2	-6.81	6.72	0.88	0.00
8	10.0	0.2	-6.10	6.07	0.76	0.00
9	0.0	0.5	-24.61	-0.00	1.15	-13.78
10	1.0	0.5	-9.09	8.04	1.04	-4.38
11	5.0	0.5	-5.65	5.50	0.87	0.00
12	10.0	0.5	-3.84	3.77	0.71	0.00
13	0.0	1.0	-24.61	-0.00	1.15	-13.78
14	1.0	1.0	-13.15	0.00	1.00	0.00
15	5.0	1.0	-6.93	0.00	0.84	0.00
16	10.0	1.0	0.00	0.00	0.62	0.00

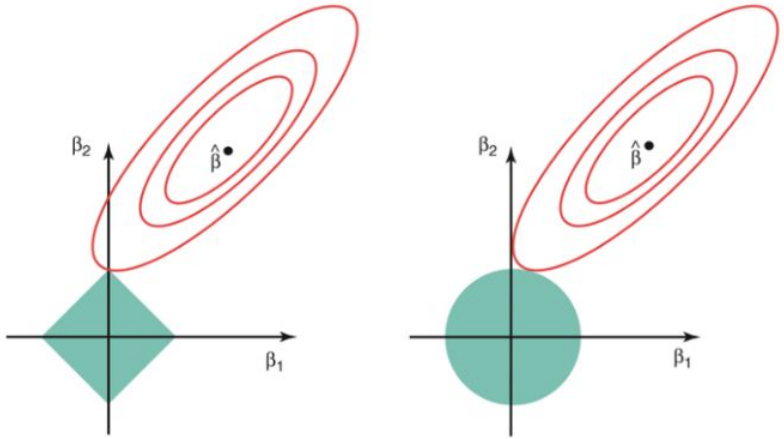
Question 17.3

Which values for λ and α correspond to: (a) unregularized linear regression, (b) pure Lasso, (c) pure ridge, (d) elastic net with an even combination of Lasso and ridge penalties? What happens to the values of the coefficients in each case?

	lambda	alpha	beta_sexf	beta_sexm	beta_age	beta_obesity
1	0.0	0.0	-24.61	-0.00	1.15	-13.78
2	1.0	0.0	-11.19	10.71	1.09	-9.97
3	5.0	0.0	-8.25	8.24	0.91	-2.11
4	10.0	0.0	-6.97	6.99	0.77	1.66
5	0.0	0.2	-24.61	-0.00	1.15	-13.78
6	1.0	0.2	-10.35	9.73	1.07	-7.84
7	5.0	0.2	-6.81	6.72	0.88	0.00
8	10.0	0.2	-6.10	6.07	0.76	0.00
9	0.0	0.5	-24.61	-0.00	1.15	-13.78
10	1.0	0.5	-9.09	8.04	1.04	-4.38
11	5.0	0.5	-5.65	5.50	0.87	0.00
12	10.0	0.5	-3.84	3.77	0.71	0.00
13	0.0	1.0	-24.61	-0.00	1.15	-13.78
14	1.0	1.0	-13.15	0.00	1.00	0.00
15	5.0	1.0	-6.93	0.00	0.84	0.00
16	10.0	1.0	0.00	0.00	0.62	0.00

Question 17.1

This picture, from *Elements of Statistical Learning* (Figure 6.7) is a geometric picture of what happens to the coefficients under the Lasso and ridge penalties.



Which picture is which? What are the red ellipses? What are the blue shapes? Do you see why Lasso is more likely than ridge to set some of the coefficients to zero?