Feature Selection, Regularization, Ridge and LASSO Regression

Machine Learning Unit 5

How do users rate a product?

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor
1. The relevance of this topic to me was	4	3	2	1
2. The usefulness of materials was	4	3	2	1
3. The effectiveness of the presenter was	4	3	2	1
4. I expect the future usefulness of this topic to be	4	3	2	1
5. My overall evaluation of this session is	4	3	2	1

Your Account > Packaging Feedback

Rate Amazon's Packaging

Did the packaging protect your items adequately? ☆☆☆☆☆ Protection 1 star = Poor; 5 stars = Excellent

Was the box size and packaging appropriate for the items?
- ◯ Too Small
- ◯ About Right
- ◯ Too Big
- ◯ Way Too Big

Rate Item's Packaging

☆☆☆☆☆ Ease of Opening 1 star = Very Difficult; 5 stars = Very Easy

Central Railway

Annexure E3 (A)

FEEDBACK FORM

"On-Board Housekeeping Services" - Indian Railways

AC COACH S. No:________

Dear Passenger,

Our endeavor is to provide you the most hygienic On Board Housekeeping Services. Your valuable feedback would help us improve further.

Kindly spare few minutes in rating the areas as given in table below:
Ratings

5 = Excellent, 4 = Very Good, 3 = Good, 2 = Average, 1 = Poor

Passenger Feedback - AC Coaches

Sr. No.	Areas of Cleaning / Services	5	4	3	2	1
	Please mark (✓) in space					
1	Cleaning / Washing of Toilet floor and commode pan					
2	Dry Cleaning of Toilet Floor					
3	Cleaning of Mirror, shelf, wall panels and other fittings in Toilets					
4	Cleaning of Wash Basin in Toilets and Doorways					
5	Cleaning of Doorway Area					
6	Cleaning of Vestibule Area including entrance to toilets					
7	Cleaning of Passenger compartments					
8	Cleaning of Passenger aisle area					
9	Cleaning of Window Glasses on Platform side					
10	Cleaning of Dust Bins of coaches					
11	Disinfection and provision of Deodorant in toilets					
12	Spraying of air freshener in compartments					
13	Spraying of Mosquito Repellent					
14	Replenishment of Liquid Soap in Coach toilets					
15	Replenishment of Tissue Paper Roll in Western style Coach toilets					
16	Collection of Garbage and disposal in Poly Bags duly segregate as Biodegradable / Non biodegradable					
17	Behaviour of Janitors / Supervisor					
18	Hygiene & Cleanliness of Janitors / Supervisor including their uniform					
	Scores*					
	Passenger Satisfaction Index (PSI)*					

*Not to be filled by the passenger

Image source: Google Images

How do users rate a product?

User 1:

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	2
2. The usefulness of materials was	4	3	2	1	2
3. The effectiveness of the presenter was	4	3	2	1	5
4. I expect the future usefulness of this topic to be	4	3	2	1	2
5. My overall evaluation of this session is	4	3	2	1	4

User 2:

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	4
2. The usefulness of materials was	4	3	2	1	4
3. The effectiveness of the presenter was	4	3	2	1	1
4. I expect the future usefulness of this topic to be	4	3	2	1	3
5. My overall evaluation of this session is	4	3	2	1	2

For both users, feature 3 seems to play a major role in deciding the overall evaluation, other features have smaller impact

Image source: Google Images

How do users rate a product?

User 1:

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	1
2. The usefulness of materials was	4	3	2	1	2
3. The effectiveness of the presenter was	4	3	2	1	5
4. I expect the future usefulness of this topic to be	4	3	2	1	1
5. My overall evaluation of this session is	4	3	2	1	5

User 2:

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	5
2. The usefulness of materials was	4	3	2	1	4
3. The effectiveness of the presenter was	4	3	2	1	1
4. I expect the future usefulness of this topic to be	4	3	2	1	5
5. My overall evaluation of this session is	4	3	2	1	1

For both users, feature 3 seems to be the only factor in deciding the overall evaluation, other features do not matter

Image source: Google Images

Feature Selection

- Linear regression model: $y_i = \sum_j w_j x_{ij} + b_i$, i.e. all feature ratings contribute to the final rating
- But in the examples, only a small number of features seem to influence the final rating, other features have little importance
- In case 1: One element in "w" will have high value, other elements will have small values
- In case 2: All elements except one in "w" have 0 value, i.e. "w" is sparse!

Feature Selection

- Feature selection: the task of identifying the “important” features
- Important feature: those which strongly influence the final ratings
- In the given examples, feature selection is easy by manual inspection
- Large dataset: many examples, many dimensions, noisy ratings, manual inspection impossible
- Can linear regression itself solve the feature selection problem?
- It can, if it returns a suitable “w”!

Sparse Regression for Feature Selection

- Case 1: we want “w” such that most of its elements are small
- Case 2: we want “w” such that most of its elements are 0
- Can we convert these demands into mathematical formulations?

Sparse Regression for Feature Selection

- Case 1: we want "w" such that most of its elements are small
- Case 2: we want "w" such that most of its elements are 0
- Can we convert these demands into mathematical formulations?
- General recipe: find a regularization function f(w)
- f(w) should have low value for suitable "w", high value for unsuitable "w"

Sparse Regression for Feature Selection

- Case 1: we want “w” such that most of its elements are small
- Case 2: we want “w” such that most of its elements are 0
- Can we convert these demands into mathematical formulations?
- General recipe: find a regularization function f(w)
- f(w) should have low value for suitable “w”, high value for unsuitable “w”

- Find (w,b) to minimize $L(w,b) + \lambda f(w)$
- First term to find w that fits data, second term to find “w” that is suitable, λ to balance them!

Ridge Regression

- f(w) : how to choose?
- Simplest $f(w) = ||w||_2^2$
- The L_2-norm of vector "w", $||w||_2^2$

  $= w^T w = \sum_j w_j^2$
- Limits the distance of "w" from origin

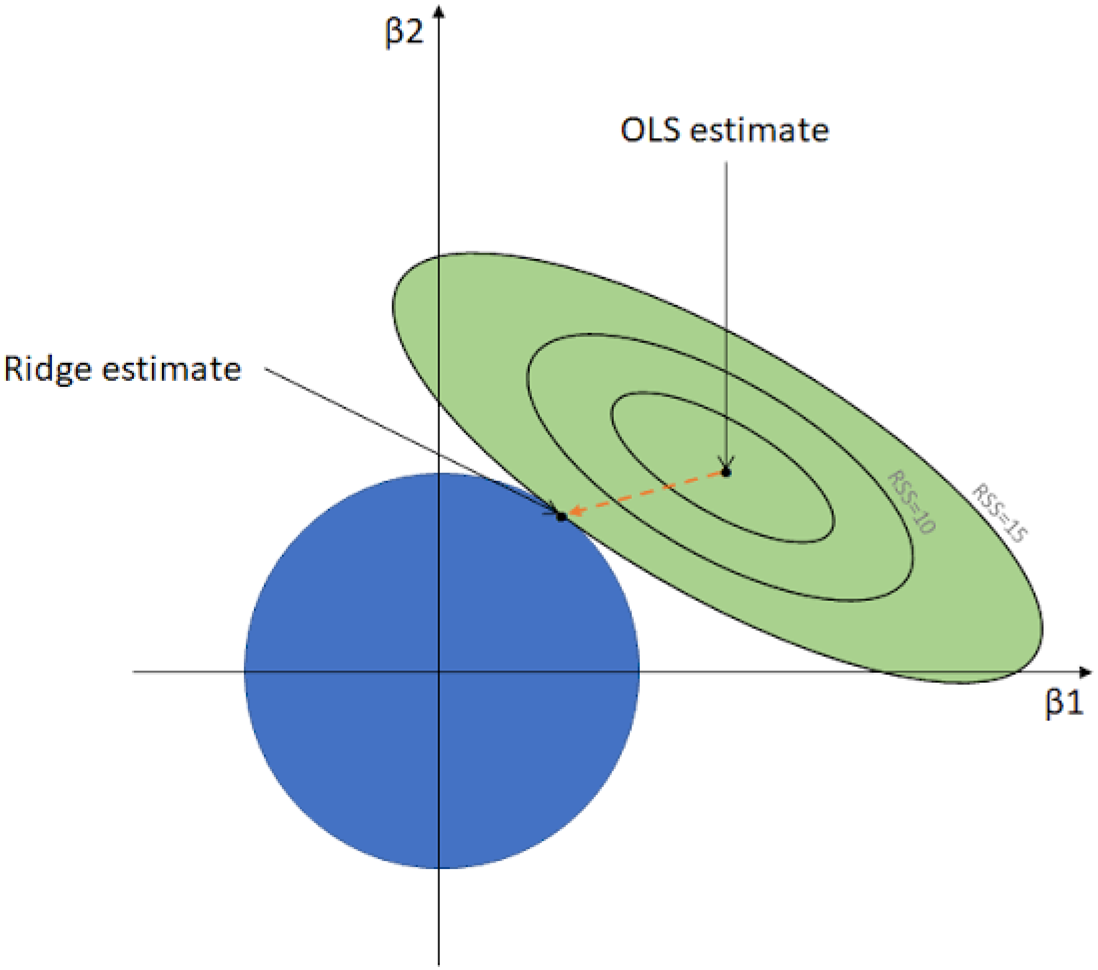


Image source: Google Images

Ridge Regression

- f(w) : how to choose?
- Simplest $f(w) = ||w||_2^2$
- The L2-norm of vector "w", $||w||_2^2$

  $= w^T w = \sum_j w_j^2$
- Limits the distance of "w" from origin

  i.e. constrains the different dimensions
- Low value of $||w||_2^2$ indicates that all features will have restricted weights.
- Popularly known as "ridge regression"

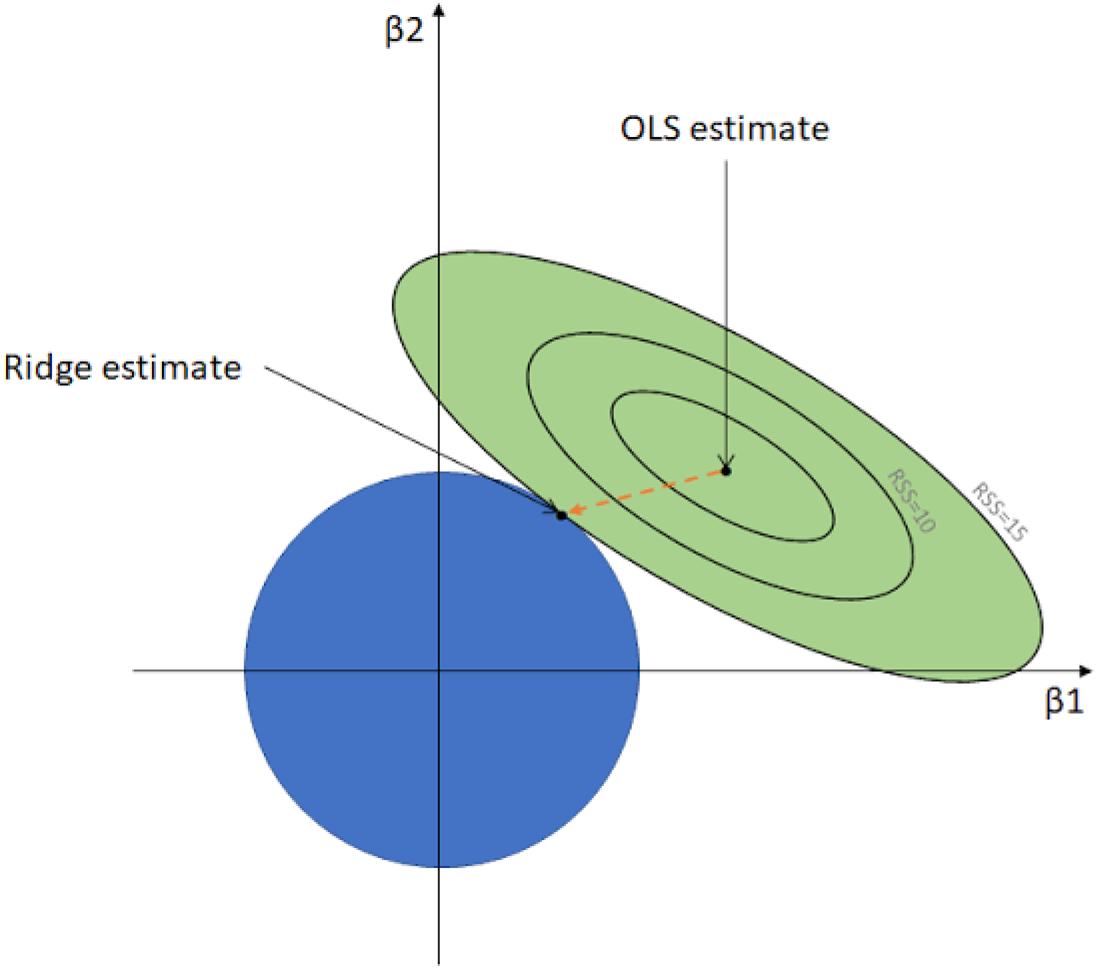


Image source: Google Images

Ridge Regression: Mathematics

Loss function $L(w, b) = \sum_i (y_i - w^T x_i - b)^2$

Regularization $f(w) = \frac{1}{2}||w||_2^2 = w^T w$

Objective function $\mathcal{L}(w, b) = L(w, b) + \lambda f(w)$

$\frac{dL}{dw} = 0 \implies \sum_i (y_i - w^T x_i - b) x_i + \lambda w = 0$

$\frac{dL}{db} = 0 \implies \sum_i (y_i - w^T x_i - b) = 0$

Ridge Regression: Mathematics

Solving these equations, we get

$b = \bar{y} - w^T \bar{x}$

$w = (\sum_i (\tilde{x}_i \tilde{x}_i{}^T) + \lambda I)^{-1} (\sum_i \tilde{x}_i \tilde{y}_i)$

where $\bar{x} = \frac{1}{N} \sum_i x_i$, $\bar{y} = \frac{1}{N} \sum_i y_i$, $\tilde{x}_i = x_i - \bar{x}$

Here, all the additions are vector additions
I is the $D \times D$ identity matrix

Notice the similarity with linear regression!

The role of λ-parameter

- λ decides the relative importance of fitting error and regularizer
- Small value of λ: regularization not important!

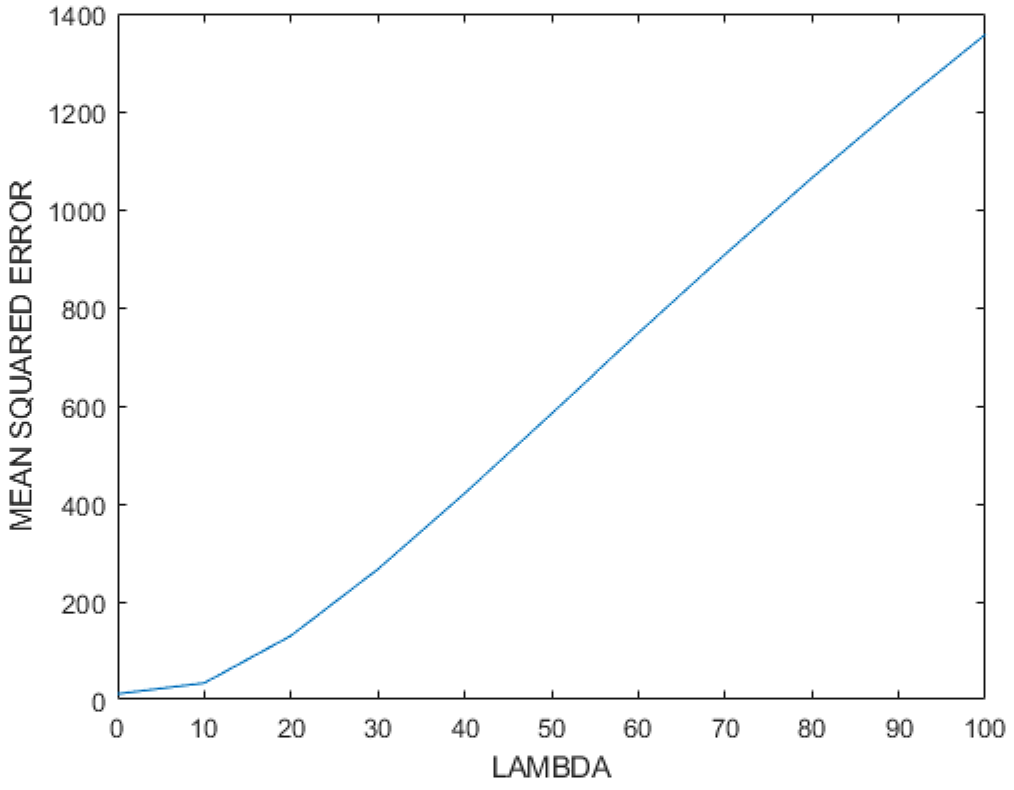
  low error, “w” vector may contain large values!

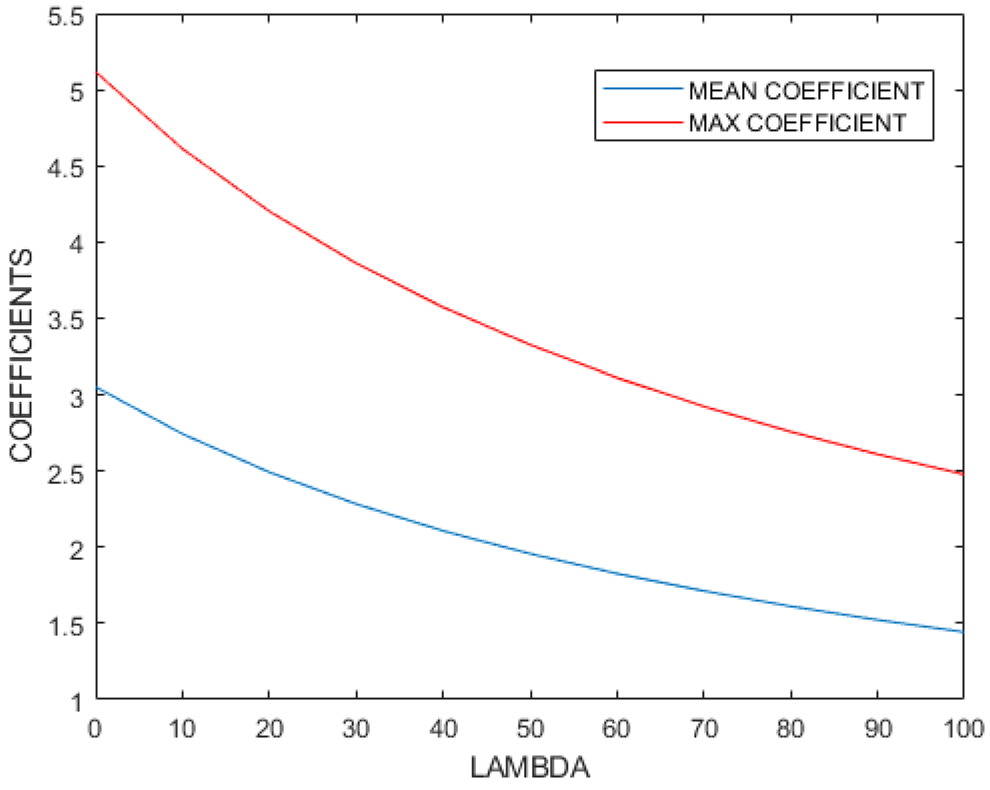
  result similar to linear regression!
- Large value of λ: fitting error not important!

  high error, but “w” contains small values

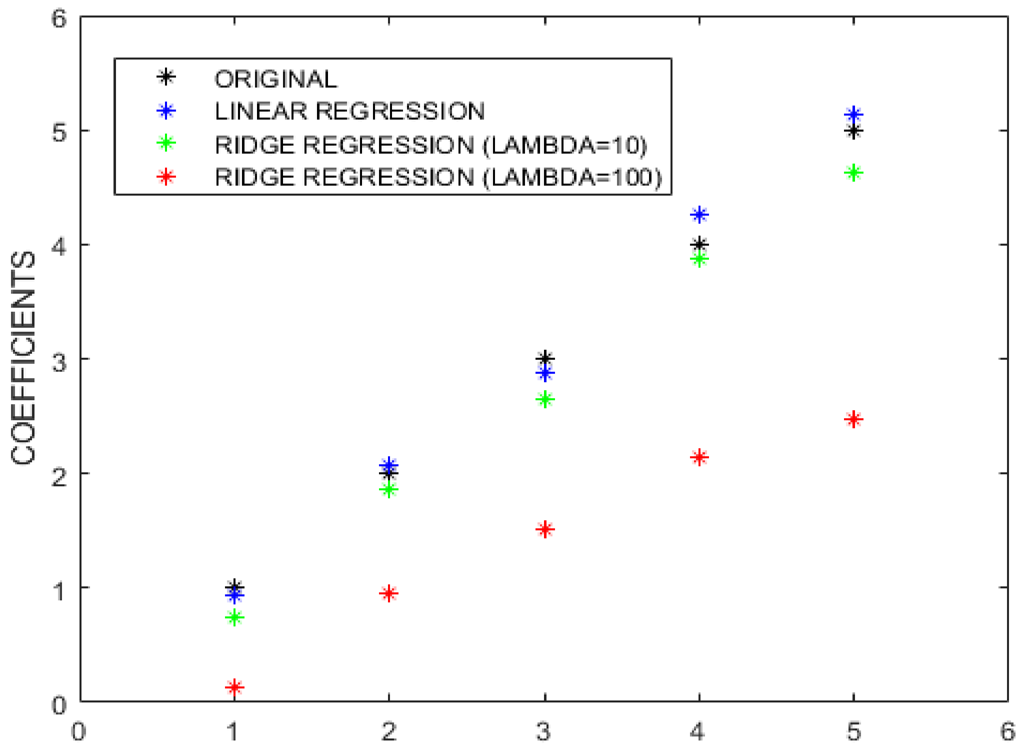
  result different from linear regression!

The role of λ-parameter

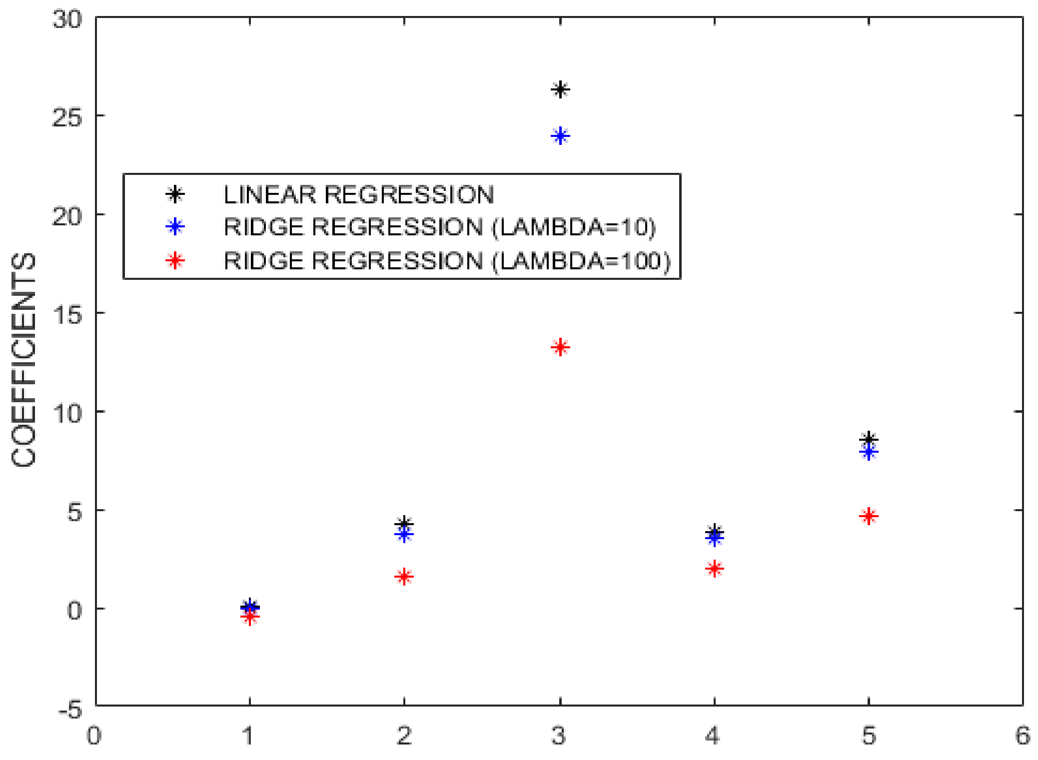




Ridge Regression vs Linear Regression



Original function: linear



Original function: non-linear

LASSO regression

- Our original aim: “sparse w”!
- The L_0-norm of vector “w”: number of non-zero elements
- Regularizer $f(w) = ||w||_0$ promotes sparse “w”!
- New problem: $L(w,b) + \lambda f(w)$
- Non-differentiable function!!!

LASSO regression

- Our original aim: "sparse w"!
- The L_0-norm of vector "w": number of non-zero elements
- Regularizer $f(w) = ||w||_0$ promotes sparse "w"!
- New problem: $L(w,b) + \lambda f(w)$
- Non-continuous function!!!
- Relaxation: $f(w) = ||w||_1 = \sum_j |w_j|$ = sum of absolute values of elements!
- Low value of $||w||_1$: most values of w "close to 0"
- "Almost sparse" w!

LASSO vs Ridge Regression

- Both are compromise between squared loss minimization and feasible region

  Feasible region shape different in both cases

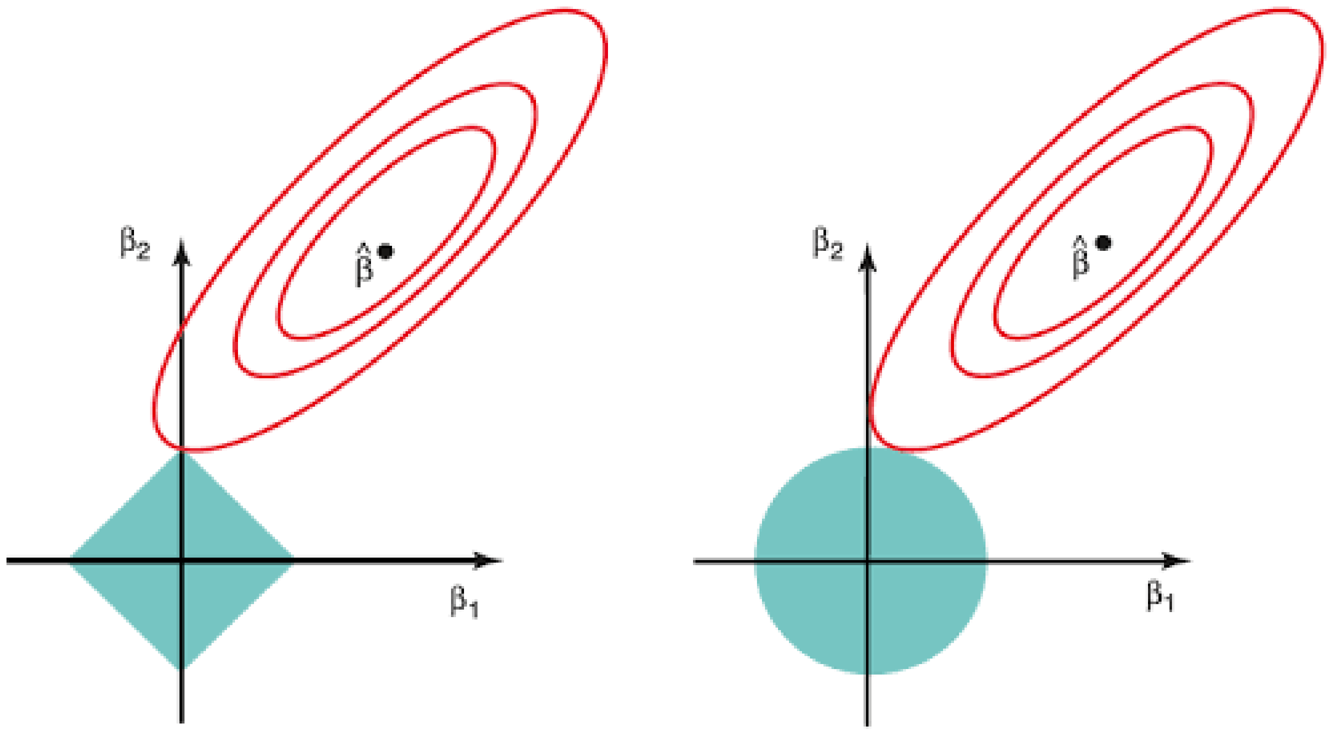


Image source: Google Images

LASSO regression

- Objective function: $\sum_i (y_i - w^T x_i - b)^2 + \lambda ||w||_1$
- Difficult to solve by differentiation!

- Alternative: use numerical method instead of analytical!
- Gradient Descent: to be covered later!

Python Implementation using sklearn

In [64]:
```
TrainX=np.asarray(X)
TrainY=np.asarray(Y)

type(NewX)
```

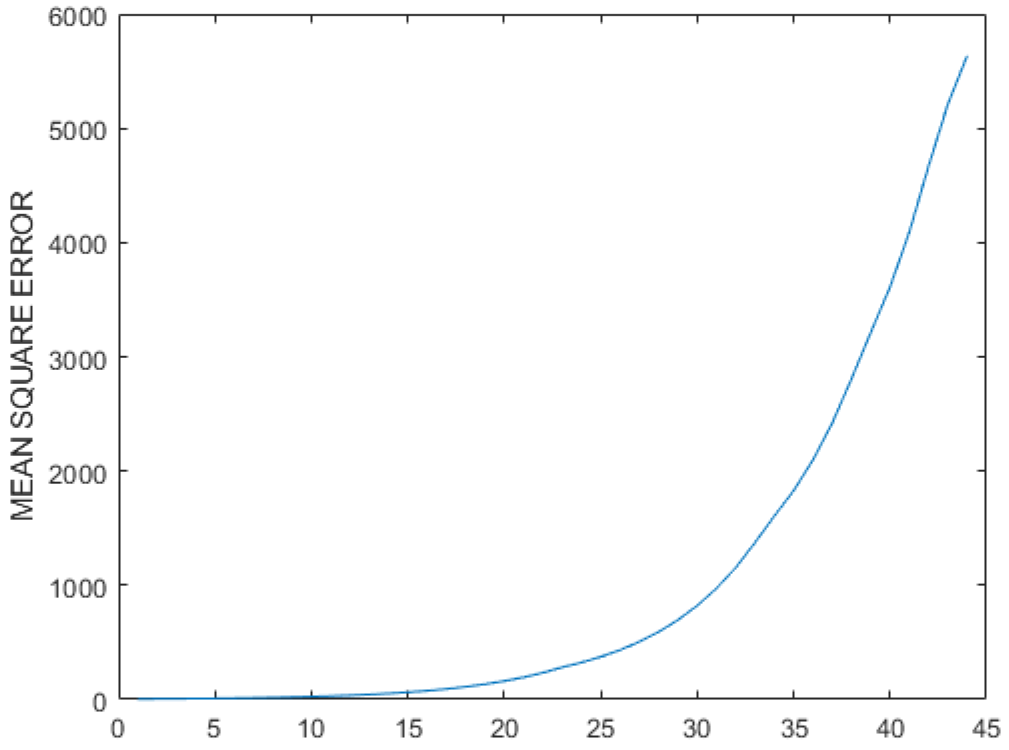
Out[64]: numpy.ndarray

In [0]:
```
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model  import Lasso
from sklearn.linear_model  import Ridge
```

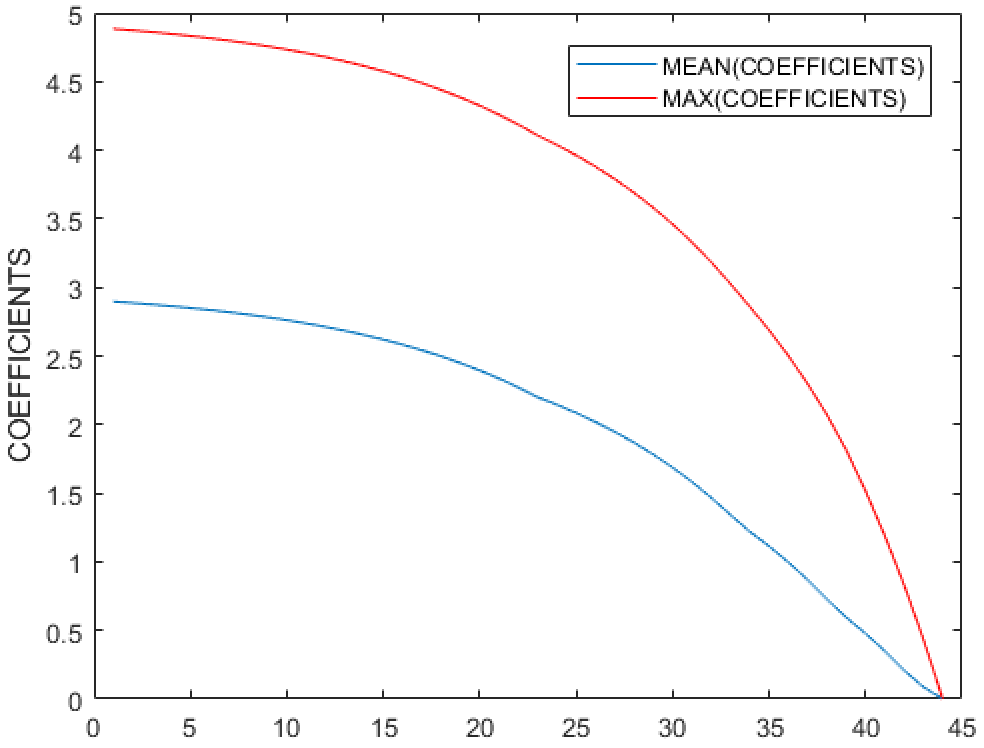
In [73]:
```
lasso=Lasso()
parameters={'alpha': [0.001,0.01,0.1, 0.5,1]}
lassoReg=GridSearchCV(lasso,parameters,scoring='neg_mean_squared_error',cv=3)    #using gridsearch for cross validation
lassoReg.fit(TrainX.reshape(-1,1),TrainY.reshape(-1,1))    # training

ridge=Ridge()
parameters={'alpha': [0.1, 0.5,1]}
ridgeReg=GridSearchCV(ridge,parameters,scoring='neg_mean_squared_error',cv=3)    #using gridsearch for cross validation
ridgeReg.fit(TrainX.reshape(-1,1),TrainY.reshape(-1,1))    # training
```

LASSO regression



LAMBDA



LAMBDA