Machine Learning

Dr Changjiang He, Dr Kuo-Ming Chao
Computer Science| School of Art
University of Roehampton

Lesson 2.3

Lasso Regression

Contents

- Multi-input linear regression
- Overfitting problem
- Lasso regression

Review of Single Input Linear Regression

In the simplest case, the regression model allows for a linear relationship.

$$y = m \times x + b$$

y is the forecast variable – rental price

x is the predictor - size

m, b are parameters;

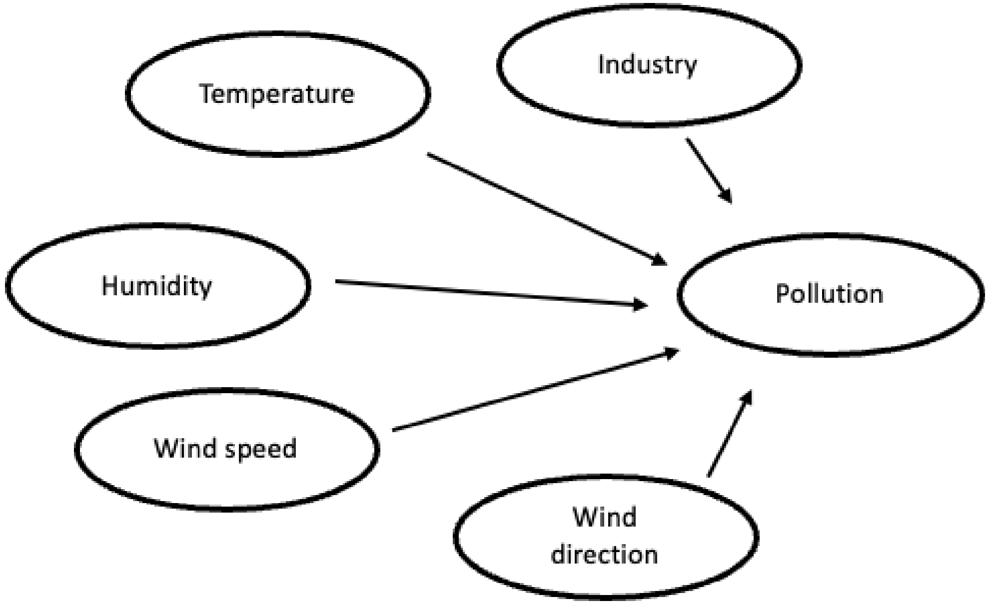
ID	Size	Rental Price
1	500	320
2	550	380
3	620	400
4	630	390
5	665	385
6	700	410
7	770	480
8	880	600
9	920	570
10	1,000	620

Multi-input Regression

When there are several predictors and one forecast variable,

the model can be extended to includes a number of predictor.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_M x_M + e$$



Predictive model of air pollution

Multi-input Regression

How do we estimate the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_M$ in this case?

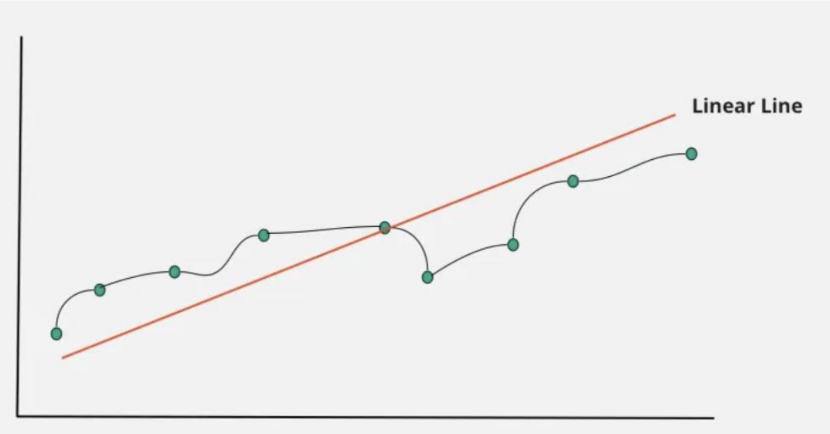
We can still use least square estimation to minimise mean squared error (MSE).

For $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_M x_M + e$, we want to minimise:

$$\frac{1}{n}\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1,i} - \beta_2 x_{2,i} - \cdots - \beta_M x_{M,i})^2$$

Overfitting Problem

The line in the above graph represents the linear regression model.



You can see how well the model fits the data.

It looks like a good model, but sometimes the model fits the data too much, resulting in overfitting.

This often could happen when there are irrelevant predictors in multi-input regression problem.

Overfitting Problem

Overfitting happens when the model learns the data as well as the noises in the training set.

Wrong parameters get calculated if there is a lot of irrelevant data and noise in the training set.

Overfitting causes low model accuracy.

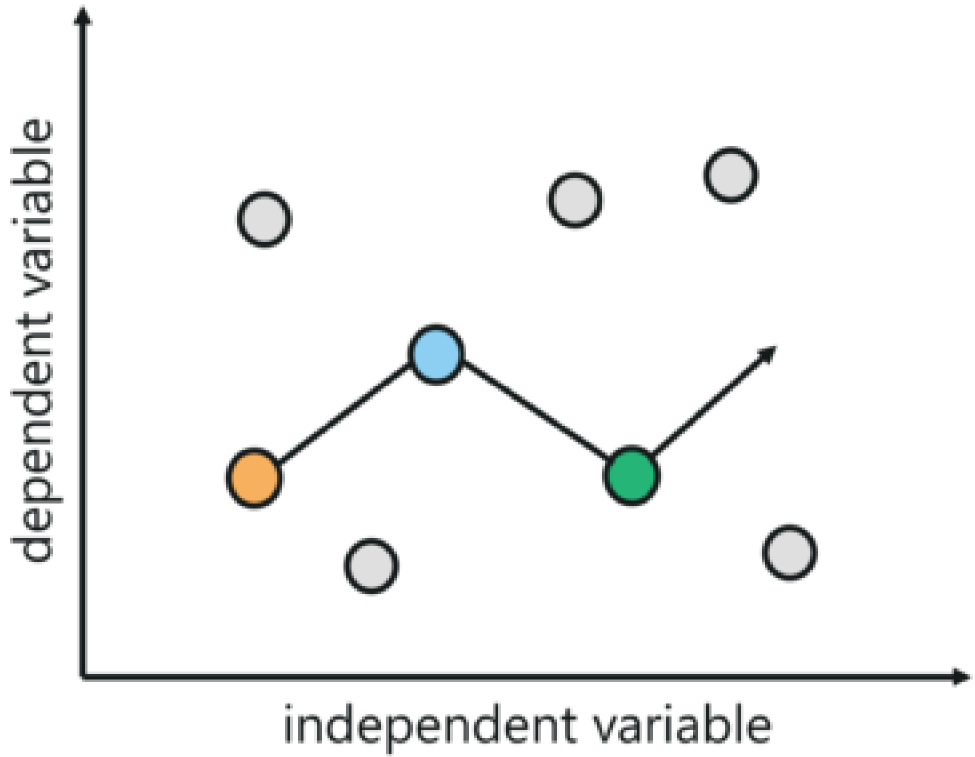
This will not go well for model predictions in the future.

Regularization

Regularization solves the overfitting problem.

In cases like this, we can use regularization to regularize or shrink these wrongly learned parameters to zero.

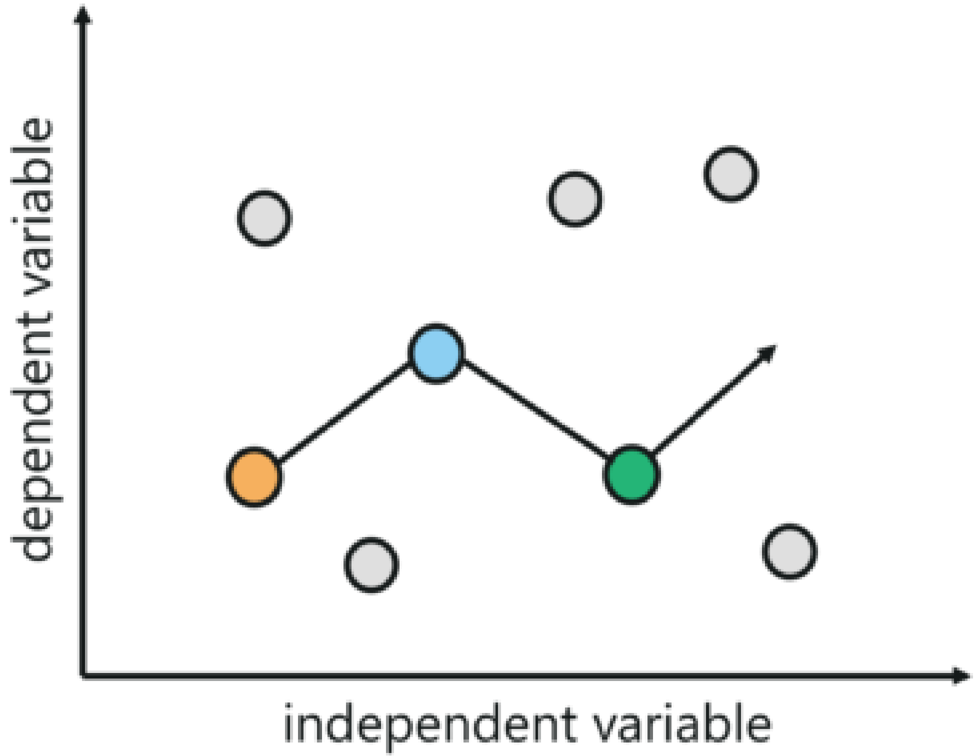
There are two main regularization techniques, namely Ridge Regression and Lasso Regression.



Regularization

If a regression model uses the L1 Regularization technique, then it is called Lasso Regression.

If it used the L2 regularization technique, it's called Ridge Regression



Lasso regression

L1 regularization adds a penalty that is equal to the absolute value of the magnitude of the coefficient.

Lasso regression minimises the following function:

$$\underbrace{\sum_{i=1}^{n}(y_i - \sum_j x_{ij}\beta_j)^2}_{\text{MSE}} + \underbrace{\lambda \sum_{j=1}^{p} |\beta_j|}_{\text{penalty}}$$

where λ denotes the amount of shrinkage.

Lasso regression

- $\lambda = 0$ implies all features are considered and it is equivalent to the linear regression where only the residual sum of squares is considered to build a predictive model
- $\lambda = \infty$ implies no feature is considered i.e, as λ closes to infinity it eliminates more and more features
- The bias increases with increase in λ
- variance increases with decrease in λ

Lasso regression

- This regularization type can result in sparse models with few coefficients.
- Some coefficients might become zero and get eliminated from the model.
- Larger penalties result in coefficient values that are closer to zero.
- This ideal for producing simpler models.

Extra: Ridge regression

University of Roehampton London

- On the other hand, L2 regularization (Ridge Regression) does not result in any elimination of sparse models or coefficients.
- Thus, Lasso Regression is easier to interpret as compared to the Ridge.
- After this lesson, find some resources (online, library, etc) and explore how to add penalty of Ridge regression.