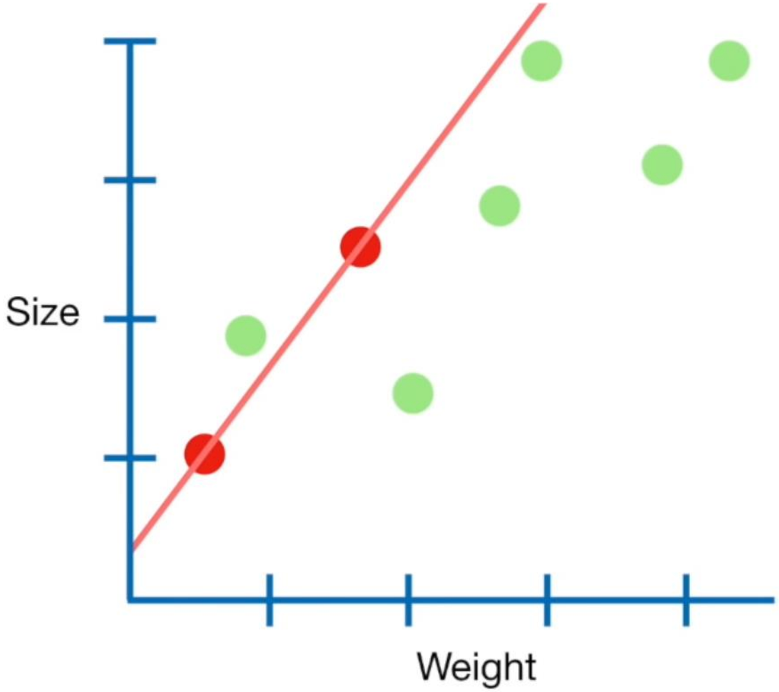
Ridge and Lasso Regression

Víctor Acevedo Vitvitskaya

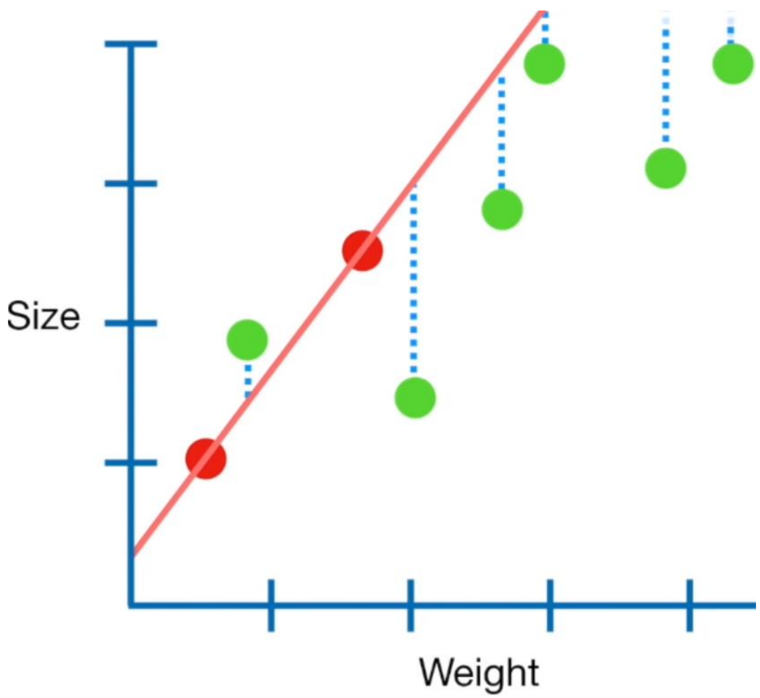
Ridge regression

- Red points: training data
- Green points: test data



Ridge regression

- Red points: training data
- Green points: test data



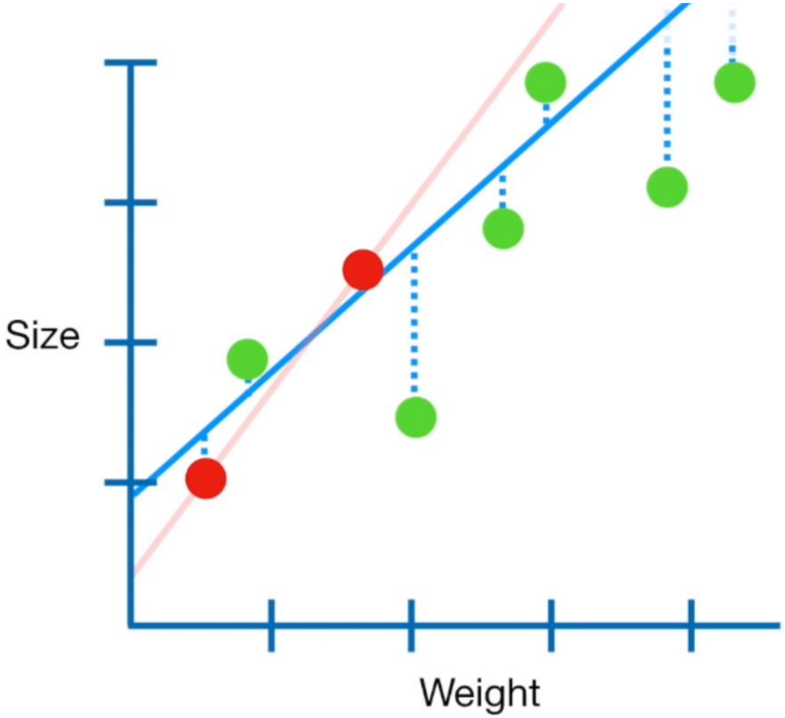
The red line is good for the training data, but it seems to be a bad estimation for the test data. This is called overfitting.

Ridge regression finds a line that does not feet training data as well, but is a much better approximation to the real data.

This is the classic trade off bias-variance.

Ridge regression

- Red points: training data
- Green points: test data



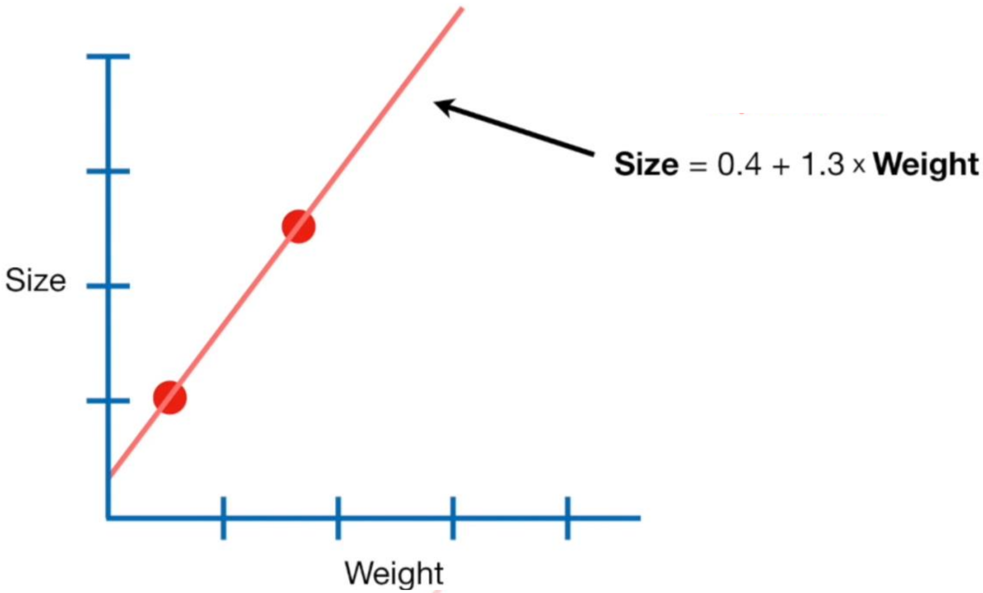
The red line is good for the training data, but it seems to be a bad estimation for the test data. This is called overfitting.

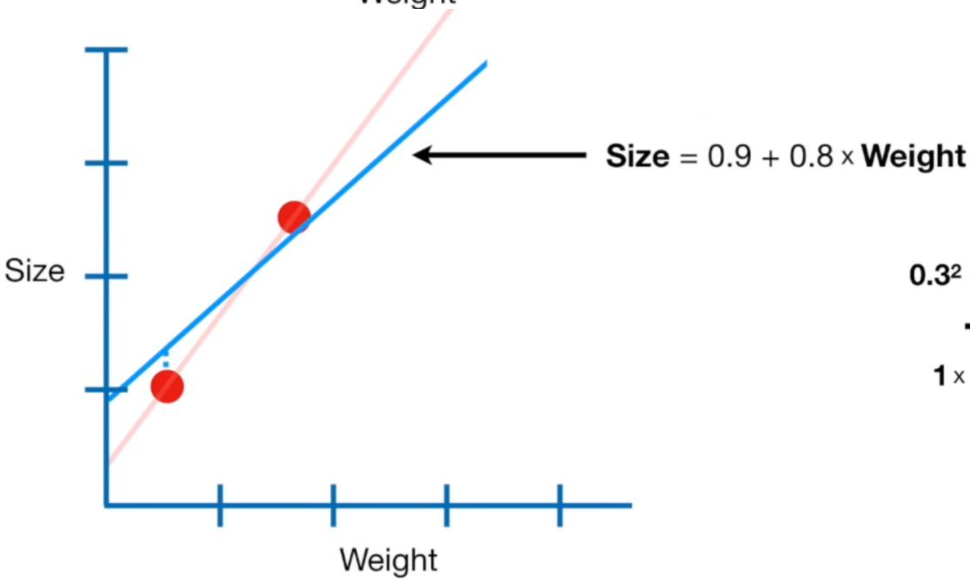
Ridge regression finds a line that does not feet training data as well, but is a much better approximation to the real data.

This is the classic trade off bias-variance.

Now, the loss function contains a penalty known as lambda.

Ridge regression





$$\begin{array}{c} 0.3^2 + 0.1^2 \\ + \\ 1 \times 0.8^2 \end{array} = 0.09 + 0.01 + 0.64$$

$$\boxed{= 0.74}$$

Lasso regression

Instead of squaring the slope for the loos function, we take the absolute value. Lambda steel can be any value from zero to infinity and it is determined using CV.

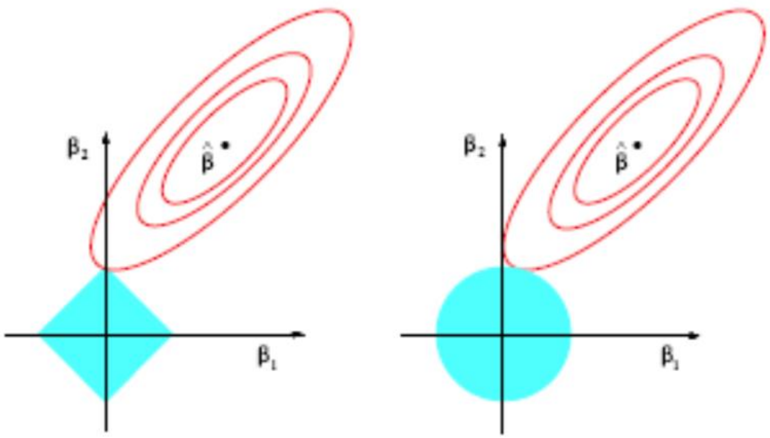


Figure 3.12: *Estimation picture for the lasso (left) and ridge regression (right). Shown are contours of the error and constraint functions. The solid blue areas are the constraint regions $|\beta_1| + |\beta_2| \leq t$ and $\beta_1^2 + \beta_2^2 \leq t^2$, respectively, while the red ellipses are the contours of the least squares error function.*

https://newonlinecourses.science.psu.edu/stat508/lesson/5/5.4