A general form of Ridge regression optimization looks like:

$$\text{Minimize}: \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \sum_{j=1}^{p}\hat{\beta}_j x_j\right)^2 + \lambda\sum_{j=1}^{p}\hat{\beta}_j^2$$

In this case, $\hat{\beta}_0 = 0$ and $n = p = 2$.
So the optimization looks like:

$$\text{Minimize}: (y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2 + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)$$

Since According to this setting

$x_{11} = x_{12} = x_1$ and $x_{21} = x_{22} = x_2$

The ridge regression seeks to Minimize

$$\boxed{(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)}$$

(b) Taking derivative of the expression obtained in (a) wrt β_1 and β_2 and equating to 0, we get :

$$\cancel{\hat{\beta}_1 x_1^2 + y_1^2}$$

$$\hat{\beta}_1 (x_1^2 + x_2^2 + \lambda) + \hat{\beta}_2 (x_1^2 + x_2^2) = y_1 x_1 + y_2 x_2 \quad (1)$$

and

$$\hat{\beta}_1 (x_1^2 + x_2^2) + \hat{\beta}_2 (x_1^2 + x_2^2 + \lambda) = y_1 x_1 + y_2 x_2 \quad (2)$$

Subtracting (1) and (2) we get

$$\hat{\beta}_1 - \hat{\beta}_2 = 0$$

or $\boxed{\hat{\beta}_1 = \hat{\beta}_2}$

Scanned by CamScanner

(c) The lasso optimization looks like

Minimize : $(y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2 + \lambda(|\hat{\beta}_1| + |\hat{\beta}_2|)$

$\because x_{11} = x_{12} = x_1$ and $x_{21} = x_{22} = x_2$ we get

$$\boxed{\text{Minimize}: (y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 + \lambda(|\hat{\beta}_1| + |\hat{\beta}_2|)}$$

Scanned by CamScanner

The lasso constraint takes the form $|\hat{\beta}_1| + |\hat{\beta}_2| < s$,
which when plotted takes the shape of a diamond centered at origin $(0,0)$.

Next consider the squared optimization constraint
$(y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2$

We use the facts $x_{11} = x_{12}$, $x_{21} = x_{22}$, $x_{11} + x_{21} = 0$, $x_{12} + x_{22} = 0$ and $y_1 + y_2 = 0$ to simplify it

to minimize : $2(y_1 - (\hat{\beta}_1 + \hat{\beta}_2)x_{11})^2$

This optimization problem has a simple solution :
$$\hat{\beta}_1 + \hat{\beta}_2 = \frac{y_1}{x_{11}}$$

This is a parallel line to the edge of the Lasso-diamond $\hat{\beta}_1 + \hat{\beta}_2 = s$.

Now the solution to the original lasso optimization problem are contours of the function $(y_1 - (\hat{\beta}_1 + \hat{\beta}_2)x_{11})^2$ that touch the lasso diamond at $\hat{\beta}_1 + \hat{\beta}_2 = s$

Finally, as $\hat{\beta}_1$ and $\hat{\beta}_2$ vary along the line $\hat{\beta}_1 + \hat{\beta}_2 = y_1/x_{11}$ these contours touch the lasso diamond edge $\hat{\beta}_1 + \hat{\beta}_2 = s$ at different points. As a result, the entire edge $\hat{\beta}_1 + \hat{\beta}_2 = s$ is a potential solution to the lasso optimization.

Similar argument can be made for the opposite lasso diamond edge : $\hat{\beta}_1 + \hat{\beta}_2 = -s$
Thus the lasso problem doesnot have a unique solution. ~~The~~ gene

The general form of the solution is

$\hat{\beta}_1 + \hat{\beta}_2 = \delta$; $\hat{\beta}_1 \geq 0$; $\hat{\beta}_2 \geq 0$ and

$\hat{\beta}_1 + \hat{\beta}_2 = -\delta$; $\hat{\beta}_1 \leq 0$; $\hat{\beta}_2 \leq 0$