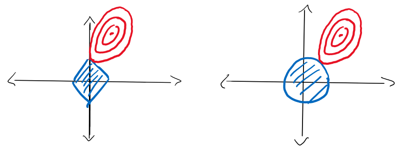
Contour Plots for Ridge and Lasso

We want to understand the following contour plots



Consider the square loss for a training set with n observations

$$\hat{R}(w) = \frac{1}{n} \sum_{i=1}^{n} (w^T x_i - y_i)^2$$

$$= \frac{1}{n} (Xw - y)^T (Xw - y)$$

Set

$$\hat{w} = (X^T X)^{-1} X^T y$$

Here we assume that X has full rank so we can invert the covariance matrix $X^T X$

Note that

$$(Xw-y)^T (Xw-y) = (X(X^TX)^{-1}X^Ty - y)^T (X(X^TX)^{-1}X^Ty - y)$$

$$= y^T X (X^TX)^{-1} X^T X (X^TX)^{-1} X^T y$$

$$- 2 y^T X (X^TX)^{-1} X^T y$$

$$+ y^T y$$

$$= - y^T X (X^TX)^{-1} X^T y + y^T y$$

Therefore

$$\hat{R}(w) = -\frac{1}{n} y^T X \hat{w} + \frac{1}{n} y^T y \qquad (1)$$

Exercise Recall from Homework 6 that

$$x^T M x - 2b^T x = (x - M^{-1}b)^T M (x - M^{-1}b) - b^T M^{-1} b \qquad (2)$$

Here we complete the square.

Let's apply these two expressions to better understand $\hat{R}$ for different inputs.

$$(Xw-y)^T (Xw-y) = w^T X^T X w - 2 y^T X w + y^T y$$

$$\overset{(2)}{=} (w - (X^TX)^{-1}X^Ty)^T X^T X (w - (X^TX)^{-1}X^Ty)$$

$$- (X^Ty)^T (X^TX)^{-1} X^T y + y^T y$$

$$= (w - \hat{w})^T X^T X (w - \hat{w})$$

$$- y^T X \hat{w} + y^T y$$

$$\overset{(1)}{=} (w - \hat{w})^T X^T X (w - \hat{w}) + n \hat{R}(\hat{w})$$

We conclude that

$$\hat{R}(w) - \hat{R}(\hat{w}) = (w - \hat{w})^T X^T X (w - \hat{w})$$