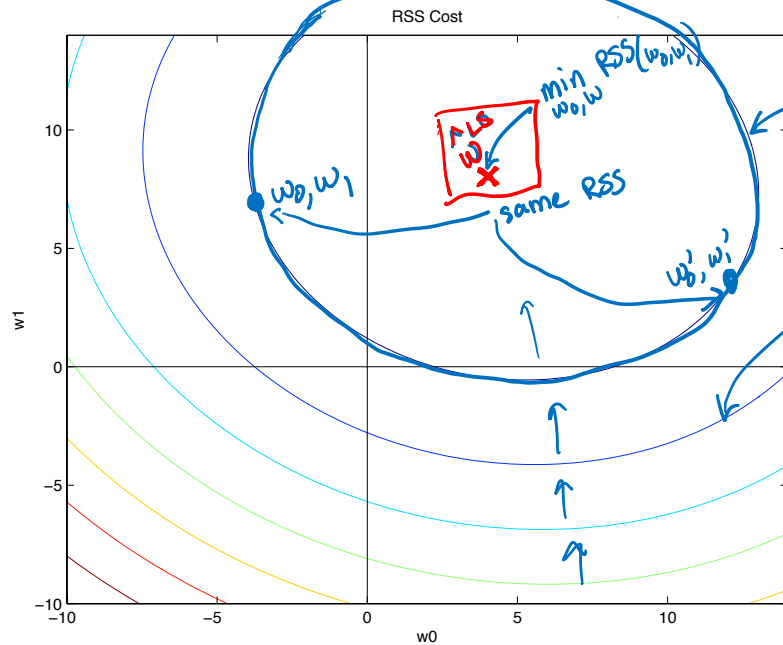


# Geometric intuition for sparsity of lasso solution



## Geometric intuition for ridge regression

# Visualizing the ridge cost in 2D



$RSS(w_0, w_1) = \text{const}_1$

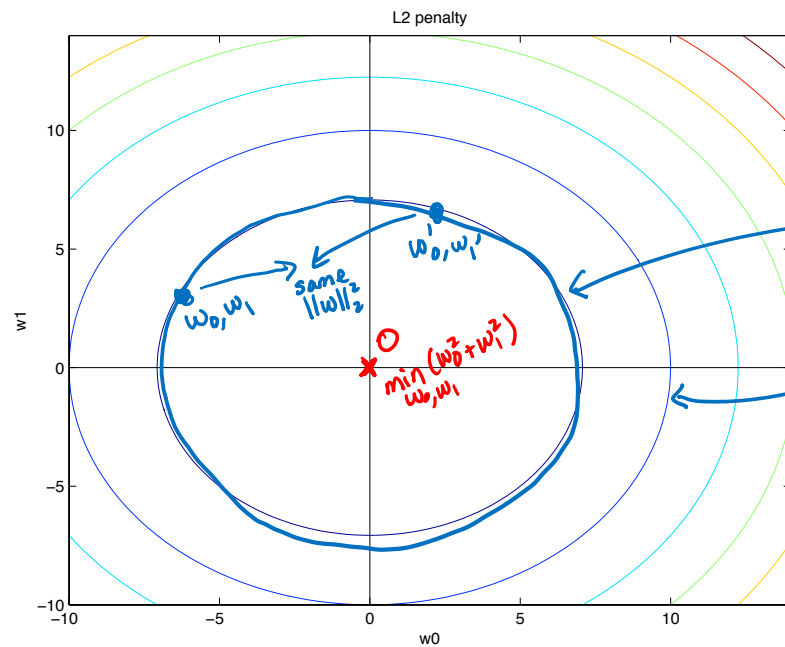
$RSS(w_0, w_1) = \text{const}_2 > \text{const}_1$   
 $\vdots$

$\sum_i y_i^2 + \underline{w_0^2 h_0^2} + \underline{w_1^2 h_1^2} + \text{cross terms} = \text{constant} +$   
ellipse

2 features for visualization sake

$$RSS(w) + \lambda ||w||_2^2 = \sum_{i=1}^N (y_i - w_0 h_0(x_i) - w_1 h_1(x_i))^2 + \lambda (w_0^2 + w_1^2)$$

# Visualizing the ridge cost in 2D

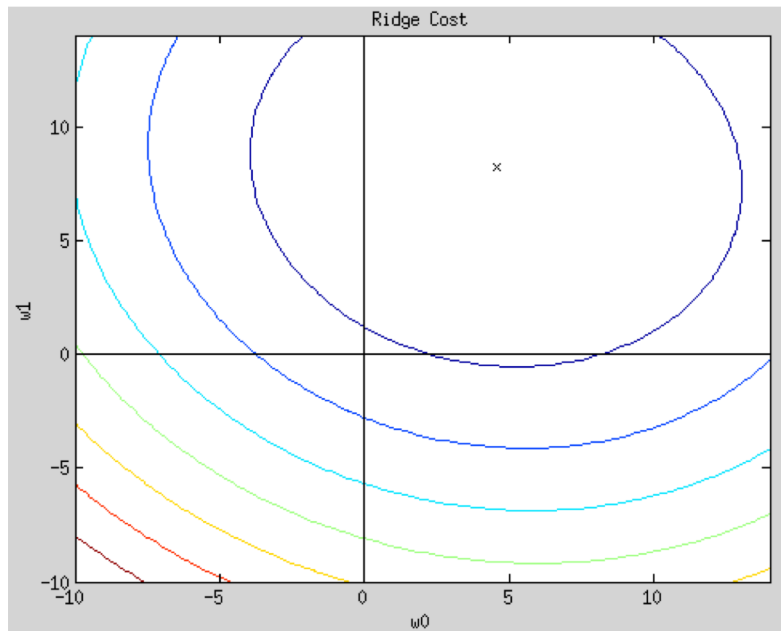


$\|w\|_2^2 = \text{const}_1$   
 $\|w\|_2^2 = \text{const}_2 > \text{const}_1$   
 $\vdots$

$w_0^2 + w_1^2 = \text{constant}$   
circle

$$\text{RSS}(w) + \lambda \|w\|_2^2 = \sum_{i=1}^N (y_i - w_0 h_0(x_i) - w_1 h_1(x_i))^2 + \lambda (w_0^2 + w_1^2)$$

# Visualizing the ridge cost in 2D



Add contour plots together

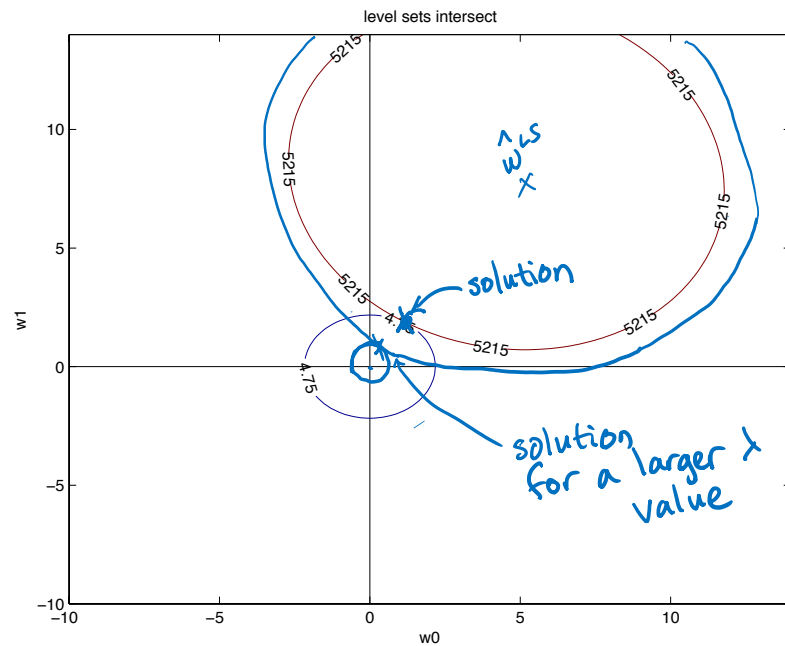
$RSS(w) + \lambda \|w\|_2^2$   
ellipses       $\uparrow$  circles  
weighting

Movie: function of increasing  $\lambda$

$\times$  mark optimal solution for a specific  $\lambda$

$$RSS(w) + \lambda \|w\|_2^2 = \sum_{i=1}^N (y_i - w_0 h_0(x_i) - w_1 h_1(x_i))^2 + \lambda (w_0^2 + w_1^2)$$

# Visualizing the ridge solution

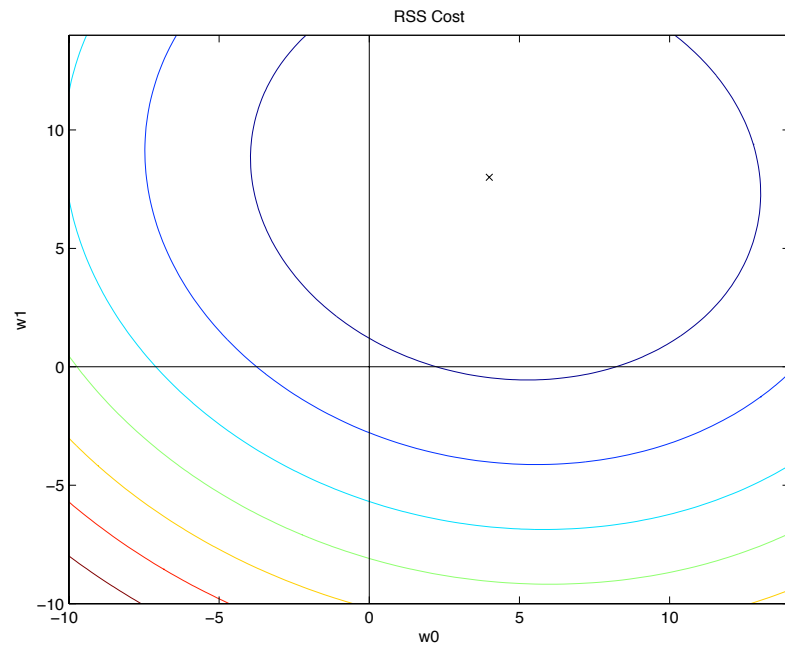


$$\text{RSS}(w) + \lambda \|w\|_2^2 = \sum_{i=1}^N (y_i - w_0 h_0(x_i) - w_1 h_1(x_i))^2 + \lambda (w_0^2 + w_1^2)$$



# Geometric intuition for lasso

# Visualizing the lasso cost in 2D

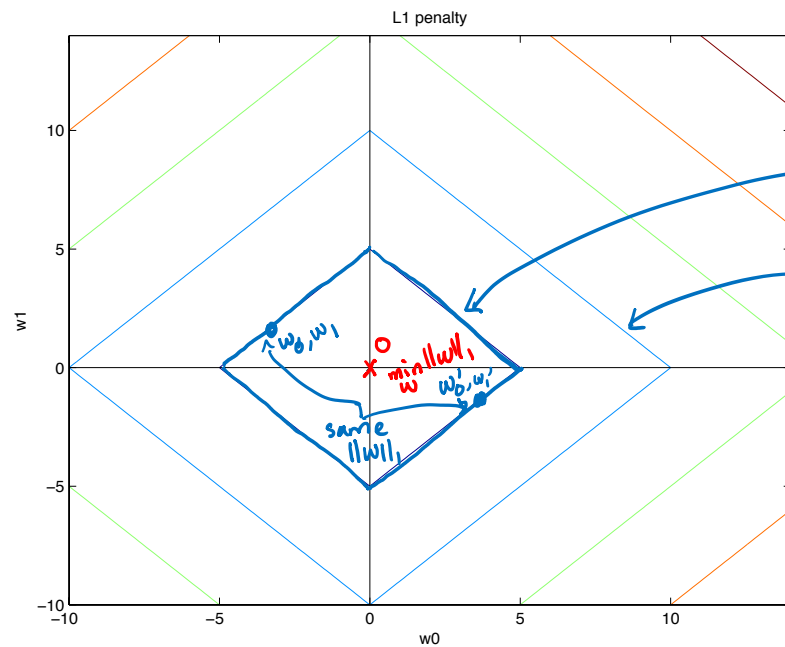


RSS contours for lasso are exactly the same as those for ridge!

$$\text{RSS}(\mathbf{w}) + \lambda \|\mathbf{w}\|_1 = \sum_{i=1}^N (y_i - w_0 h_0(x_i) - w_1 h_1(x_i))^2 + \lambda (|w_0| + |w_1|)$$



# Visualizing the lasso cost in 2D

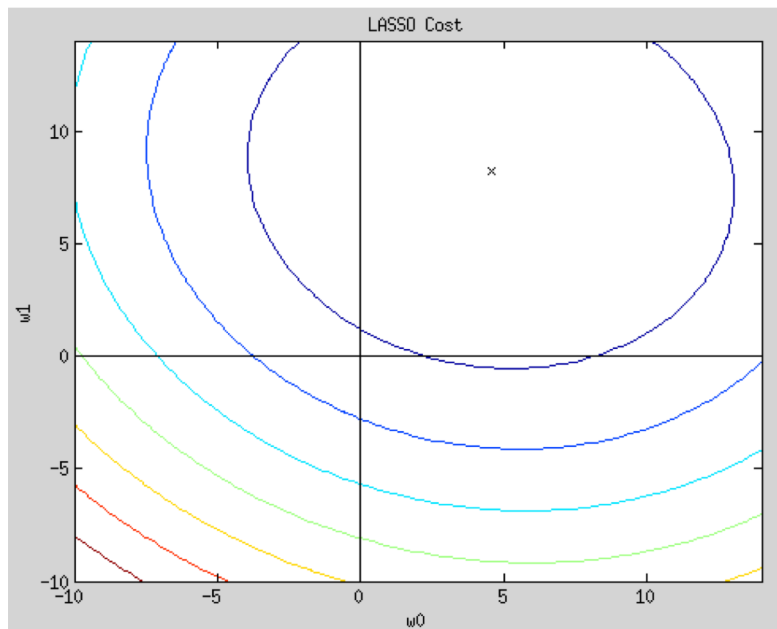


$\|w\|_1 = \text{const}_1$   
 $\|w\|_1 = \text{const}_2 \rightarrow \text{const}_1$   
 $\vdots$

$|w_0| + |w_1| = \text{constant}$   
diamond

$$\text{RSS}(w) + \lambda \|w\|_1 = \sum_{i=1}^N (y_i - w_0 h_0(x_i) - w_1 h_1(x_i))^2 + \lambda (|w_0| + |w_1|)$$

# Visualizing the lasso cost in 2D

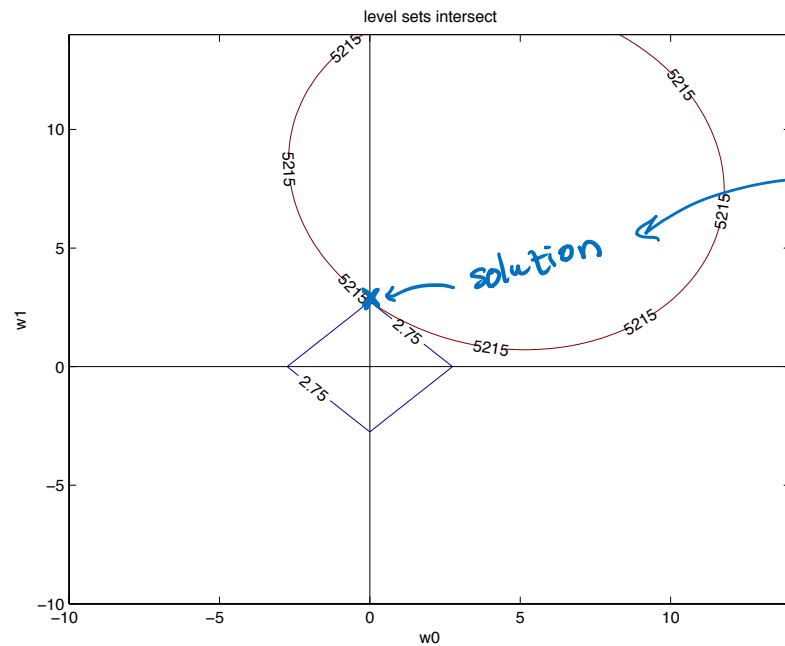


Adding  
RSS +  $\lambda \|w\|_1$   
ellipses                  diamonds

$x = \text{optimal } \hat{w} \text{ for a specific } \lambda$

$$\text{RSS}(w) + \lambda \|w\|_1 = \sum_{i=1}^N (y_i - w_0 h_0(x_i) - w_1 h_1(x_i))^2 + \lambda (|w_0| + |w_1|)$$

# Visualizing the lasso solution



$$\text{RSS}(\mathbf{w}) + \lambda \|\mathbf{w}\|_1 = \sum_{i=1}^N (y_i - \mathbf{w}_0 h_0(x_i) - \mathbf{w}_1 h_1(x_i))^2 + \lambda (|\mathbf{w}_0| + |\mathbf{w}_1|)$$