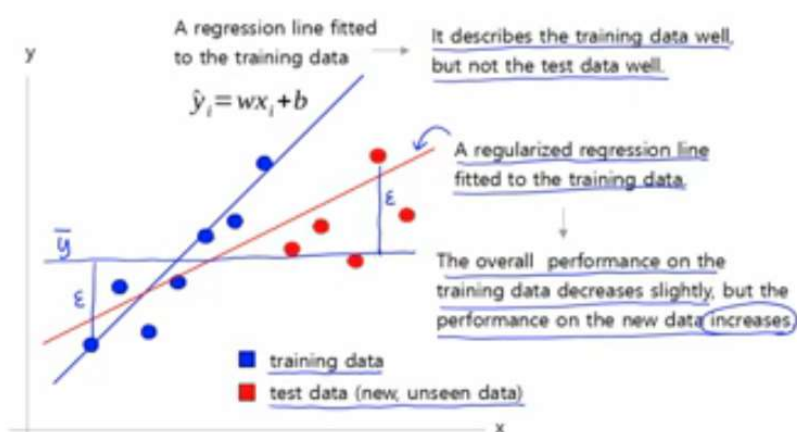


→ Regularization is a key technique that prevents overfitting by adding a penalty term to the model's loss function.

→ The penalty term prevents the loss from becoming too small.



▪ This produces a regression line that is slightly insensitive to x and not too sensitive.

Regularization: Lasso and Ridge

$$\hat{y}_i = w_1 x_{1,i} + w_2 x_{2,i} + \dots + w_k x_{k,i} + b$$

bias or intercept

$$\hat{y}_i = w_0 x_{0,i} + w_1 x_{1,i} + w_2 x_{2,i} + \dots + w_k x_{k,i} \quad (b = w_0, x_{0,i} = 1)$$

$$\hat{y}_i = \sum_{j=0}^k w_j x_{j,i} = \vec{w} \cdot \vec{x}$$

L1 regularization (LASSO)

$$\text{loss}(L_1) = \sum_{i=1}^n (y_i - \sum_{j=0}^k w_j x_{j,i})^2 + \lambda \sum_{j=0}^k |w_j|$$

subject to $\sum |w_j| \leq C$

L2 regularization (Ridge)

$$\text{loss}(L_2) = \sum_{i=1}^n (y_i - \sum_{j=0}^k w_j x_{j,i})^2 + \lambda \sum_{j=0}^k w_j^2$$

Assuming λ is a constant, λC is also a constant, and we get the following equation. λ is the regularization constant and a hyper-parameter.

subject to $\sum |w_j| \leq C_1$
 $\sum w_j^2 \leq C_2$
 Elastic-Net

Regularization constant $\dots + \lambda_1 \sum |w_j| + \lambda_2 \sum w_j^2$

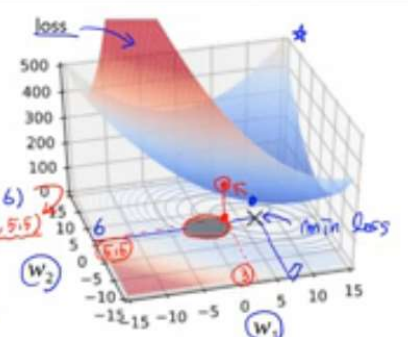
Constrained optimization (Lagrangian method)

$$\min_w \sum_{i=1}^n (y_i - \sum_{j=0}^k w_j x_{j,i})^2 \quad \text{subject to} \quad \sum_{j=0}^k w_j^2 \leq C$$

objective function
 variable
 constant value

$$\min_{w, \lambda} \left[\sum_{i=1}^n (y_i - \sum_{j=0}^k w_j x_{j,i})^2 + \lambda \left(\sum_{j=0}^k w_j^2 - C \right) \right] = \min_{w, \lambda} L(w, \lambda)$$

$$\frac{\partial L}{\partial w} = 0, \quad \frac{\partial L}{\partial \lambda} = 0 \rightarrow w^*, \lambda^*$$



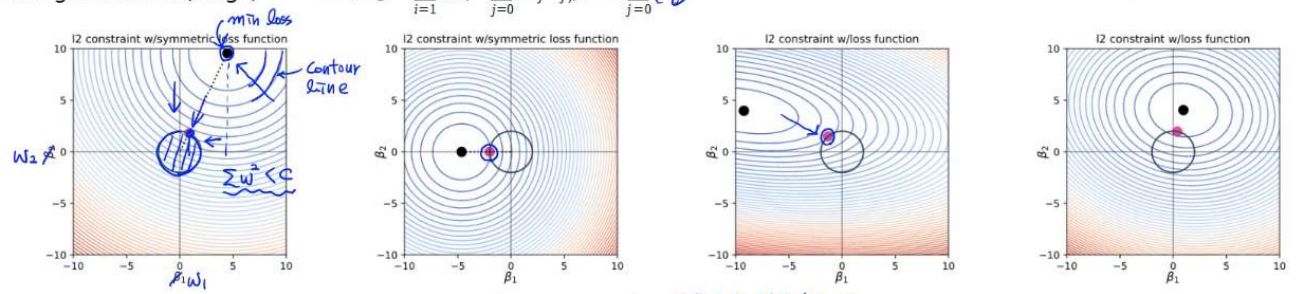
Source: <https://explained.ai/regularization/constraints.html>

■ Regularization: Lasso and Ridge

■ L2 regularization (Ridge)

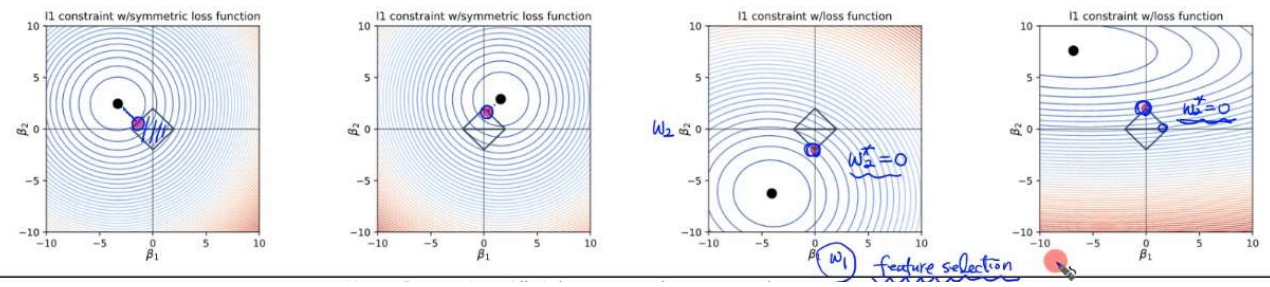
$$\text{loss}(L_2) = \sum_{i=1}^n (y_i - \sum_{j=0}^k w_j x_{j,i})^2 + \lambda \sum_{j=0}^k w_j^2$$

Source : <https://explained.ai/regularization/constraints.html>



■ L1 regularization (LASSO)

$$\text{loss}(L_1) = \sum_{i=1}^n (y_i - \sum_{j=0}^k w_j x_{j,i})^2 + \lambda \sum_{j=0}^k |w_j|$$



* No regularization on bias / intercept term