

Regularization Methods

How to avoid overfit problem:

1) Reduce number of features

2) Regularization

Any modification of the learning method to improve performance on the unseen datasets is called regularization. We need regularization to introduce Bias to the model and to decrease the Variance.

Peki Ne yapıyoruz?

Loss fonksiyonlarına (Log Loss, RSS v.b.) bir penalty Term ekliyoruz. Bu durumda fonksiyonumuz artık Regularized loss function oluyor. Artık gradient descent bu loss function'ı minimize etmeli.

$$\lambda \sum_{j=1}^p \beta_j^2 \Rightarrow \text{L2 Norm or L2 penalty}$$

$$\lambda \sum_{j=1}^p |\beta_j| \Rightarrow \text{L1 Norm or L1 penalty}$$

$$\lambda \sum_{j=1}^p (\alpha \beta_j^2 + (1-\alpha) |\beta_j|) \Rightarrow \text{Elastic Net}$$

↳ Controls the weight given to L1 or L2 penalty and its value between 0 and 1.

Not:

- The Linear Regression method that uses the L1 penalty is known as "Lasso Regression".
- The linear Regression method that uses the L2 penalty is known as "Ridge Regression".

.. / ... / ...

How to rank variables in terms of their importance?

Using the size of regression coefficients is a way to rank predictor variables. But! if the data is not normalized, we will get different scales for different variables.

↳ Therefore, we must normalize the data so that make all variables have the same scale. Then, use regression coefficients to obtain variable importances.