Linear Methods

Loss vs Metric

- Loss function:
  - Also known as a cost function.
  - What an algorithm aims to minimize during the training process.

- Metric:
  - Measure used to judge the performance of your model.

Linear Regression

- Finds proper hyperplane equation for predicting:

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

- Minimizes the loss function:

$$J(w) = MSE(w) = \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y)^2$$

Regularization

- Aims to prevent overfitting by penalizing model complexity.
- Adds a penalty to the loss function used to train the model.
- Discourages complex models with overly large coefficients.
- Typical linear models with regularization:
    - Lasso
    - Ridge
    - ElasticNet

Lasso

- L1 Regularization.
- Can shrink the coefficients of less important features to exactly zero.
- Beneficial in high-dimensional settings (num. features > num. observations).
- Minimizes the cost function:

$$J(w) = \mathrm{MSE}(w) + \alpha \sum_{i=1}^{n} |w_i|$$

Ridge

- L2 Regularization.
- Shrinks the coefficients towards zero without reaching it.
- Computationally efficient.
- Can handle high correlations among features (multicollinearity).
- Minimizes the cost function:

$$J(w) = \mathrm{MSE}(w) + \alpha \frac{1}{2} \sum_{i=1}^{n} w_i^2$$

ElasticNet

- L1 and L2 Regularizations.
- Can shrink coefficients like Lasso.
- Can handle multicollinearity like Ridge.
- Minimizes the cost function:

$$J(w) = \mathrm{MSE}(w) + r\alpha \sum_{i=1}^{n} |w_i| + \frac{1-r}{2}\alpha \sum_{i=1}^{n} w_i^2$$