Data

David Robinson

Data Processing

1. Remove features that are irrelevant or redundant.

2. Identify records with missing or outlier data and either drop the records or replace the data with the estimated value based on other information. For example, a human age of 10000 should be dropped or replaced with the mean age.

3. Engineer new features based on the existing features. For example, a boolean feature, isAdult, can be marked true for records with an age above 17.

Feature Selection Techniques

1. **Forward Selection**: Start with no variables and add them one by one based on improvement in the model.

2. **Backward Elimination**: Start with all variables and remove the least significant ones one by one.

3. **Lasso Regression**: Regularization method that automatically selects and removes irrelevant features by penalizing large coefficients.