Considerations

Train vs test

Permutation feature importance should be calculated using the **test** set or a **held out** sample.

If model over-fits, deriving feature importance from the train set is meaningless.

Deriving feature importance from the test set, shows its generalization importance.

➔ it would reflect the noise

Importance values

- The aggregated importance of every feature is greater than the drop in performance when shuffling all features.

- Shuffling is relatively costly.
  - Many features or repeating the shuffling multiple times may require some time.

- Randomness.

Susceptible to correlation

- When features are correlated, their importance is distributed across them, resulting in reduced individual importance compared to their uncorrelated counterparts.

- Diminished importance may indicate correlation rather than inherent feature significance.

Recursive elimination

- Combining permutation feature importance with recursive feature elimination may help mitigate the effect of correlation.
  - Removing a feature results in an increase in performance of the remaining variables.

- At the cost of increased computation.
  - Model is retrained after removing each feature.

Train In Data

THANK YOU

www.trainindata.com