

Univariate Imputation

Imputes values in each feature using only non-missing values or statistical estimates derived from that same feature.

E.g.,

- Mean and median imputation
- Mode imputation
- Random Sample imputation
- Etc. Section 4



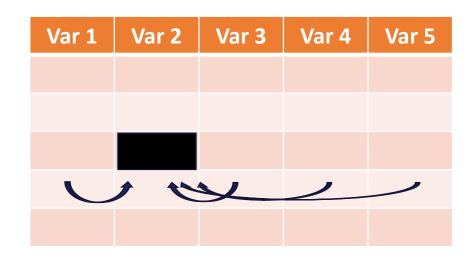
Multivariate Imputation

Uses other features in the dataset to estimate the missing values.





Multivariate Imputation



Use neighbouring variables to predict the missing value.

- KNN
- Regression

· Any ML model we like



Multivariate Imputation Techniques

KNN Imputation

MICE (multivariate Imputation of Chained Equations)

MissForest – extends MICE using Random Forests



Multivariate Imputation: Aim

Predict, as accurately as possible, the "real" values of those missing observations.

We are often not interested in those real values, instead in making accurate predictions of our target.



Multivariate Imputation: Considerations



Adds some complexity to the Machine Learning Pipeline





THANK YOU

www.trainindata.com