

Handling Missing Numerical Data

November 11, 2022 4:28 AM

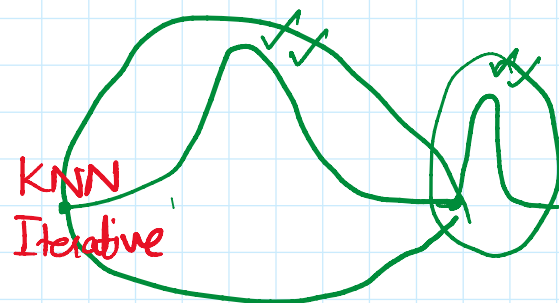
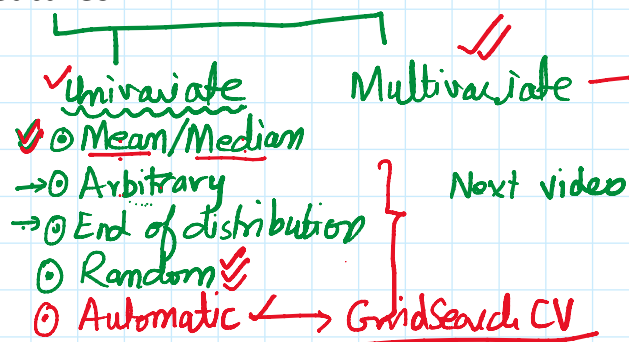
| x_1 | x_2 | x_3 |
|-------|-------|-------|
| ✓ | - | - |
| ⊗ | - | - |
| ✓ | - | ⊗ |

Complete Case Analysis →

drop all missing values.

Row → drop all rows $< 5\%$ → MCAR
columns → drop all columns missing $> 5\%$.

Numerical Features



- Pros
- Very easy to implement
- Cons
- Change the distribution of data
 - Change the relationship with other features.
- Covariance
→ Correlation
- It creates new outliers
- Box-whisker

