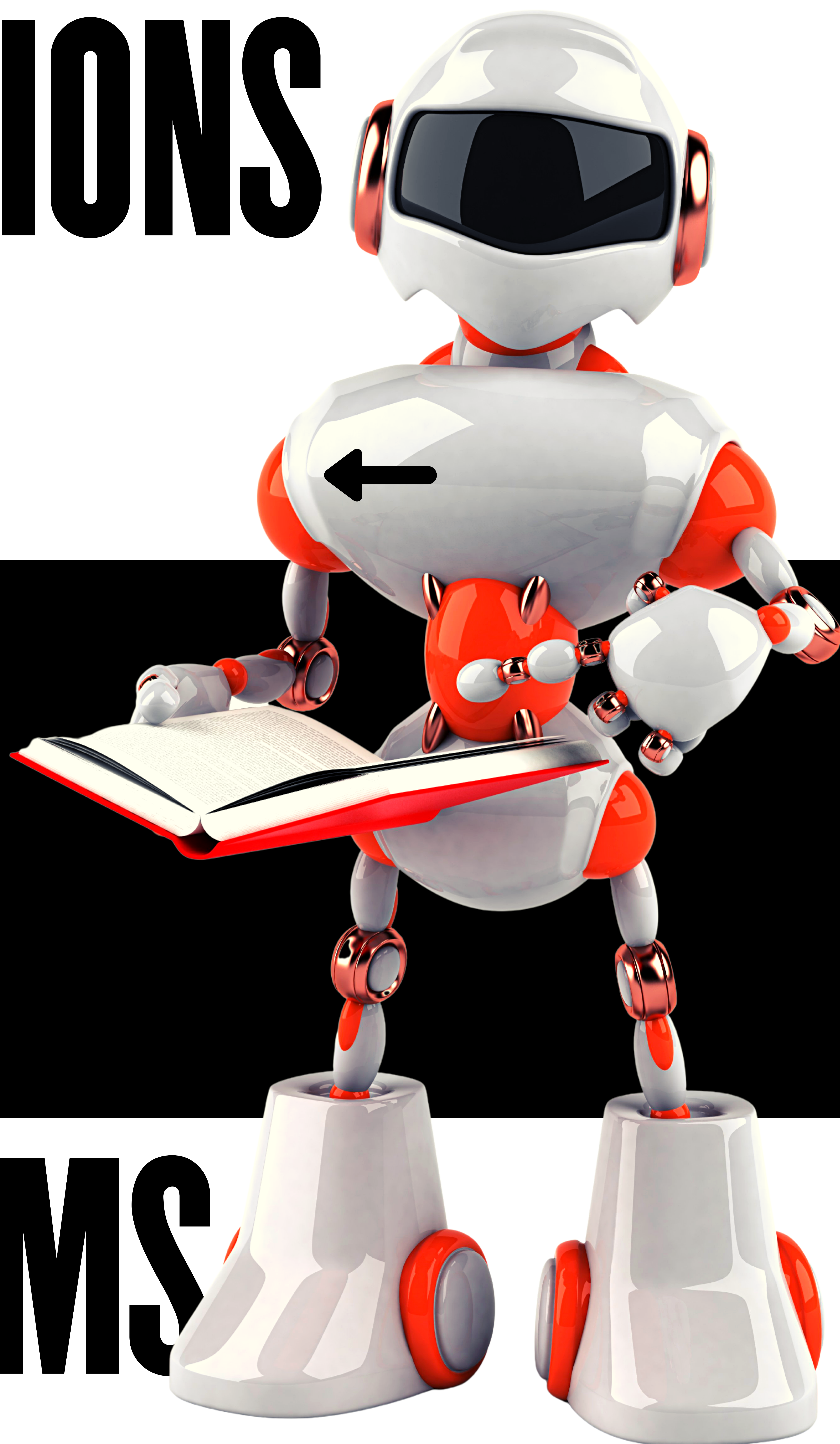


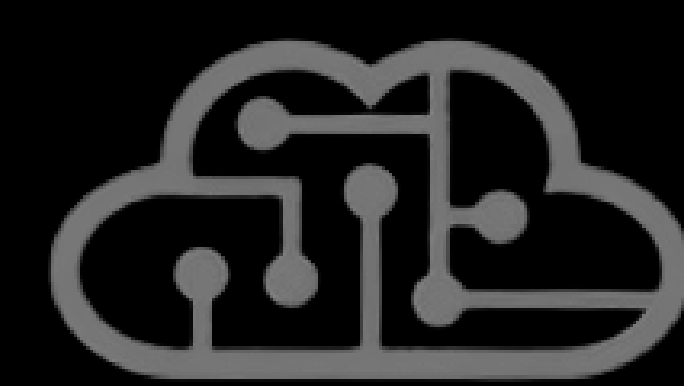
# EXPECTATIONS OF ALL

# MACHINE LEARNING

# ALGORITHMS

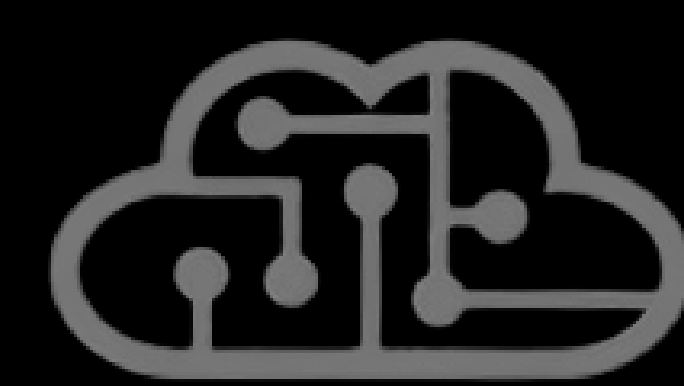






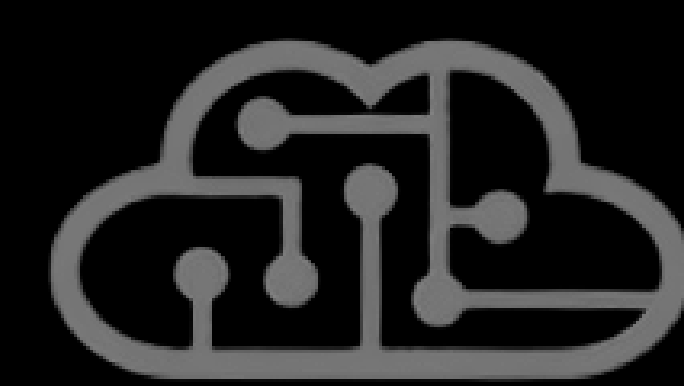
# Linear Regression Assumptions

- There should be a Linear Relationship between the Independent and Dependent Variables.
- There should be no or little multicollinearity
- No Autocorrelation
- Residuals should be Homoskedastic
- All Independent Variables are Normally Distributed



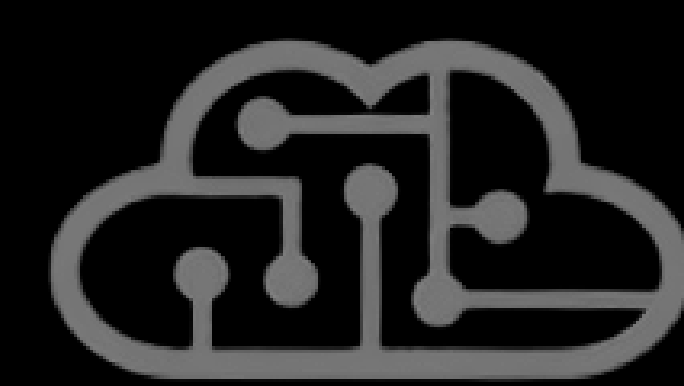
# Logistic Regression Assumptions

- Minimal or no multicollinearity among the independent variables.
- Observations to be independent of each other.
- Requires large data to predict properly.
- No Multicollinearity
- A linear relationship between the logit of the outcome and each predictor variables



## K-Nearest Neighbours Assumptions

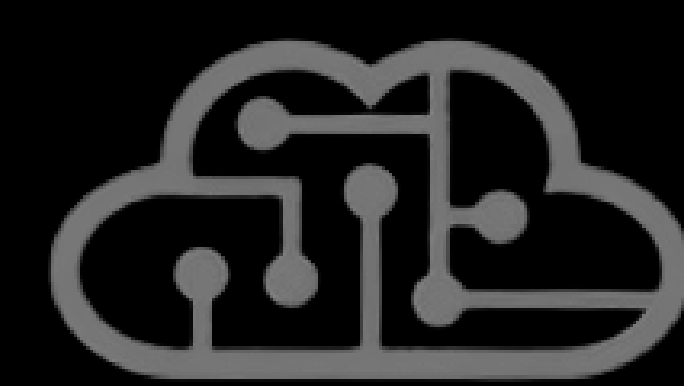
- KNN assumes that the data is in a feature space.
- Usually "K" is chosen as an odd number if the number of classes is 2.
- It does not make any assumptions on the underlying data distribution.



# Naive Bayes Assumptions

- The biggest and only assumption is the assumption of conditional independence.



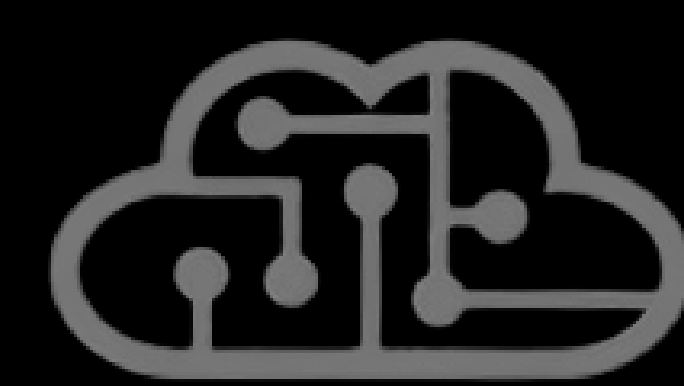


## Tree-based model Assumptions

- There are no model assumptions to validate.

## Support Vector Machine Assumptions

- There are no model assumptions to validate.



## K - means Assumptions

- It assumes the variance of the distribution of each attribute (variable) is spherical.
- All variables have the same variance.
- The prior probability for all  $k$  clusters are the same.
- Hierarchical clustering has got no separate assumption of its own.