

1) LINEAR MODEL

Simple linear regression defines relationship between dependant and independent variables by formula of straight line:

$$Y = mX + b$$

Where y is target variable, m is slope of line X is dependant variable and b is constant. Linear regression defines a line with least error which is calculated by perpendicular distance between predicted point and actual point or RMSE (Root mean square Error)

$$RMSE = \sqrt{\sum (\text{predicted} - \text{actual})^2}$$

Linear regression assumes that all variables are linearly correlated with target variable, there is no collinearity between dependant variables, and there is no correlation between independent variables.

STRENGTHS	WEAKNESSES
Can be adapted to almost any data	Does not do well with missing data
Provide estimates of strengths and size of relationship among features and the outcome	Makes strong assumptions about the data
	Require some knowledge of statistics to understand the model

2) REGRESSION TREES

In regression trees, split is performed by variable with most importance. it done using standard deviation reduction (SDR) where,

$$SDR = sd(X) - \sum (X_i/X) \times sd(X_i)$$

In regression trees, regression is performed at every split.

STRENGTHS	WEAKNESSES
Combines the strengths of trees with ability to model numeric data	Requires large amount of training data
May fit some type of data better than linear model where relation-ship between feature and target is not linear	Difficult to determine overall net effect of individual features on outcome
Does not require knowledge of statistics to interpret the model	Highly unstable, even minor addition of noise in the data may change outcome
Perform variable screening / feature selection	