Comparison of Methods

	Used For	Pros	Cons
Linear Regression	Predicting a continuous outcome (salary, price, number of votes, etc.)	• Simple, well recognized • Works on small and large datasets	• Assumes a linear relationship $Y = a\,\underbrace{\log(X)}+b$
Logistic Regression	Predicting a categorical outcome (Yes/No, Sell/Buy, Accept/Reject, etc.)	• Computes probabilities that can be used to assess confidence of the prediction	• Assumes a linear relationship

Comparison of Methods

	Used For	Pros	Cons
CART	Predicting a categorical outcome (quality rating 1--5, Buy/Sell/Hold) or a continuous outcome (salary, price, etc.)	• Can handle datasets without a linear relationship • Easy to explain and interpret	• May not work well with small datasets
Random Forests	Same as CART	• Can improve accuracy over CART	• Many parameters to adjust • Not as easy to explain as CART

Comparison of Methods

	Used For	Pros	Cons
Hierarchical Clustering	• Finding similar groups • Clustering into smaller groups and applying predictive methods on groups	• No need to select number of clusters a priori • Visualize with a dendrogram	• Hard to use with large datasets
k-means Clustering	Same as Hierarchical Clustering	• Works with any dataset size	• Need to select number of clusters before algorithm