@Sourabh Sahu

Algorithm	Best Fit Situation	Advantage	Disadvantage
Random Forest	1. Suited for at almost any machine learning problem 2. Ex. Bioinformatics	1. Can work in parallel 2. Susceptible to overfits 3. Automatically handles missing values if you impute using a special number 4. No need to transform any variable 5. No need to tweak parameters	6. Difficult to interpret 7. Weaker on regression when estimating values at the extremities of the distribution of response values 8. Biased in multiclass problems toward more frequent classes
Gradient Boosting	1. Apt at almost any machine learning problem 2. Search engines (solving the problem of learning to rank)	3. It can approximate most nonlinear function 4. Best in class predictor 5. Automatically handles missing values 6. No need to transform any variable	7. It can overfit if run for too many iterations 8. Sensitive to noisy data and outliers 9. Doesn't work at its best without parameter tuning
Linear Regression	1. Baseline prediction 2. Econometric predictions	1. Simple to understand and explain	1. You have to work hard to make it fit

	3. Modelling 4. marketing responses	It seldom overfits 2. Using L1 & L2 regularization is effective in feature selection 3. Fast to train 4. Easy to train on big data thanks to its stochastic version	nonlinear functions 2. Can suffer from outliers
Support Vector Machine	1. Character recognition 2. Image recognition 3. Text classification	1. Automatic non-linear feature creation 2. Can approximate complex non-linear functions 3. Works only with a portion of the examples (the support vectors)	1. Difficult to interpret when applying non-linear kernels 2. Suffers from too many examples, after 10,000 examples it starts taking too long to train
K Nearest Neighbor	1. Computer vision 2. Multilabel tagging 3. Recommender systems 4. Spell checking problems	1. Fast, lazy training 2. Can naturally handle extreme multiclass problems (like tagging text)	1. Slow and cumbersome in the predicting phase 2. Can fail to predict correctly due to the curse of dimensionality

Adaboost	Facedetection	1. Automatically handles missing values 2. No need to transform any variable 3. It doesn't overfit easily 4. Few parameters to tweak 5. It can leverage many different weak-learners	1. Sensitive to noisy data and outliers 2. Never the best in class predictions
Naive Bayes	1. Face recognition 2. Sentiment analysis 3. Spam detection 4. Text classification	1. Easy and fast to implement, doesn't require too much memory and can be used for online learning 2. Easy to understand 3. Takes into account prior knowledge	1. Strong and unrealistic feature independence assumptions 2. Fails estimating rare occurrences 3. Suffers from irrelevant features
Neural Networks	1. Image recognition 2. Language recognition and translation 3. Speech recognition	1. It can approximate any non-linear function	1. It requires you to define a network architecture

	4. Vision recognition	2. Robust to outliers 3. It can work with image, text and sound data	2. Difficult to tune because of too many parameters and you have also to decide the architecture of the network 3. Difficult to interpret 4. Easy to overfit
Logsitic Regression	1. Ordering results by probability 2. Modelling marketing responses	1. Simple to understand and explain 2. It seldom overfits 3. Using L1 & L2 regularization is effective in feature selection 4. The best algorithm for predicting probabilities of an event 5. Fast to train 6. Easy to train on big data thanks to its stochastic version	1. You have to work hard to make it fit non-linear functions 2. Can suffer from outliers

SVD	Recommendation System	1. Can restructure data in a meaningful way	1. Difficult to understand why data has been restructured in a certain way
PCA	1. Removing collinearity 2. Reducing dimensions of the dataset	1. Can reduce data dimensionality	1. Implies strong linear assumptions (components are a weighted summations of features)
K means	1. Segmentation	1. Fast in finding clusters 2. Can detect outliers in multiple dimensions	1. Suffers from multicollinearit 2. Clusters are spherical, can't detect groups of other shape 3. Unstable solutions, depends on initialization