Biostatistics 202: Machine Learning Guide

Supervised learning - prediction of a target

	Model	*Interpretability?*	*Stat or ML?*
Regression - continuous target			
	Linear regression	High	Both
Classification - categorical target			
	Logistic regression	High	Both
	Classification trees (decision trees)	High	ML
	K-nearest neighbors	Low	ML
	Support vector machines	Low	ML
	Artificial neural networks	Low	ML
Ensemble methods	Random forests (bagging)	Low	ML
	Adaptive boosted trees (boosting)	Low	ML
	Stacking	Low	ML

*Note: all classification methods may be adjusted to allow for prediction of continuous targets.

Unsupervised learning - clustering and/or data reduction

Clustering		
	k-Means	ML
	Heirarchical clustering	ML
Data Reduction		
	Principal component analysis	Both
	t-distributed stochastic neighbor embedding	ML