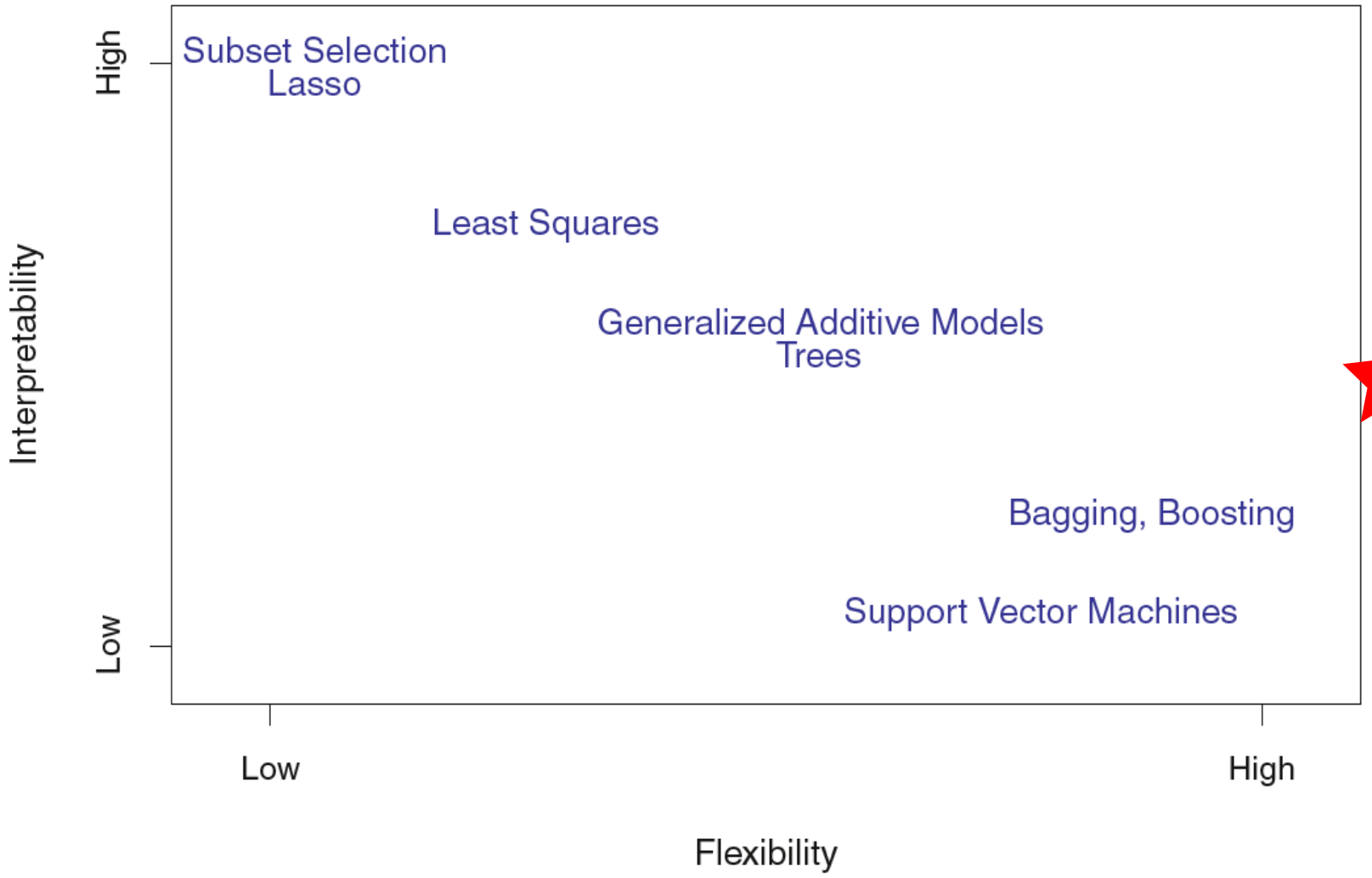
Model Selection



1 Choosing models before applying them saves time

Linear regression

~~Logistic regression~~

~~Feature selection~~

~~Subset selection~~

~~Shrinkage (Ridge & Lasso)~~

~~Dimension reduction (PCA & PLS)~~

~~Decision tree (and pruning)~~

Advanced tree

~~Cross-validation (K-fold & LOOCV)~~

Block building (Bagging, Random forest & Boosting)

Support Vector Machine

2 Data preparing defines limits of prediction

Keep as much information as possible (for random forest & xgboost)

Balance between information and Multicollinearity (for regression)

3 Combining different models' results to w more precise prediction

Arlene Huang (yh3235)