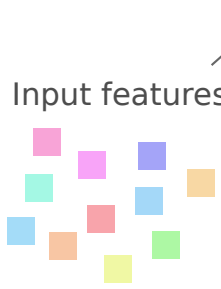
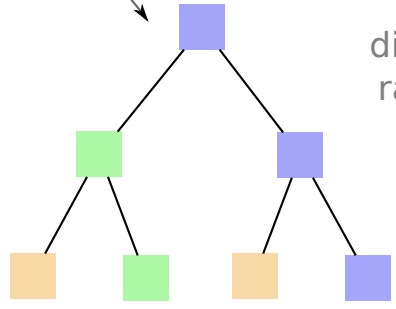
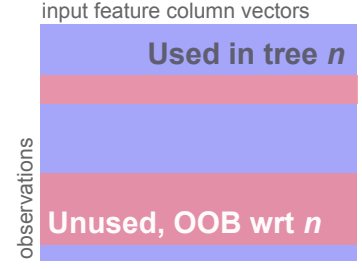




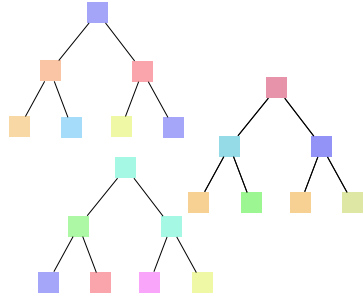
At each node, pick best discriminator from sample of random input features, with respect to **bootstrapped sample** of observations

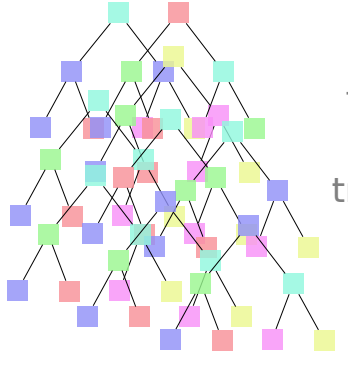


Random Forests

(in one slide)

Grow 200-500 trees this way





To predict for new ***y***, pass your features through all trees. Average the values in leaf nodes for a single prediction

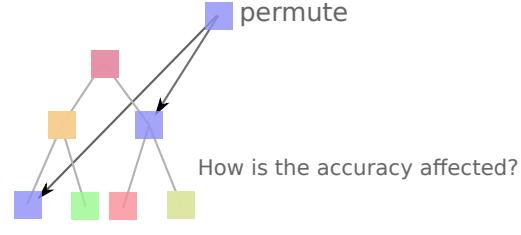
OK with extraneous variables and mixed inputs, captures non-linear relationships, classification or regression.

Parameter insensitive

mtry: helps prevent overfitting

ntrees: runtime vs. accuracy

Each tree uses a **subset** of rows, so the remainder (**OOB**: out-of-bag data), can be used to evaluate model performance as the forest is generated!



Similarly, use OOB to assess variable importance: **permute one feature** and compare the RMSE on OOB with that of unpermuted.