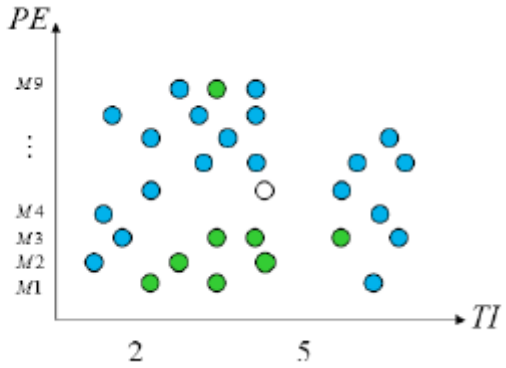
Ensemble models

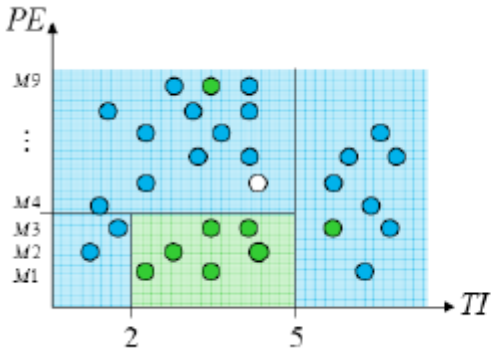
Decision tress involve greedy, recursive partitioning.

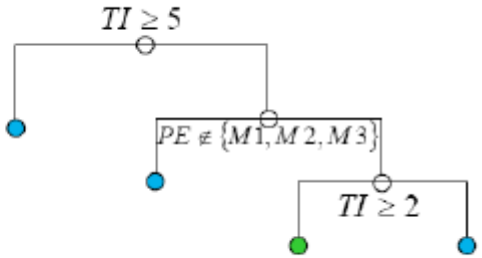
Simple dataset with two predictors

TI	PE	Response
1.0	M2	good
2.0	M1	bad
...	...	...
4.5	M5	?



- Greedy, recursive partitioning along TI and PE





Decision Tree Strengths and Weaknesses

STRENGTHS

Does not involve any distributional assumptions

Can work with missing data

Relatively computationally efficient

Transparent: produces easy to interpret output

WEAKNESSES

Greedy – locally optimal, can be globally suboptimal

Tends to overfit training data

Does not handle predictor interactions

Ensemble learning

Basic idea: combine predictions of several models

Random forest (bagging = bootstrap aggregation):

- Sample data & sample predictors
- Build multiple trees
- Model prediction = average prediction of the individual trees in the model
- Model building stops when performance on “out-of-bag” subsample stops improving

Random Forest Strengths & Weaknesses

STRENGTHS

Does not require data pre-processing

Computationally efficient

Generally improves accuracy over individual decision tree models

Provides an indication of feature importance

WEAKNESSES

Less transparent than decision tree models

Does not learn from errors

Bias towards multi-level variables

Boosted trees

Build multiple trees

“Learn” from mistakes by over-weighing misclassified cases when building the next tree

“Boosting” = improving model performance by adding models that focus on “errors” from previous rounds

Errors:

- For classification problems: misclassified cases
- For prediction problems: largest error vis-à-vis actual values