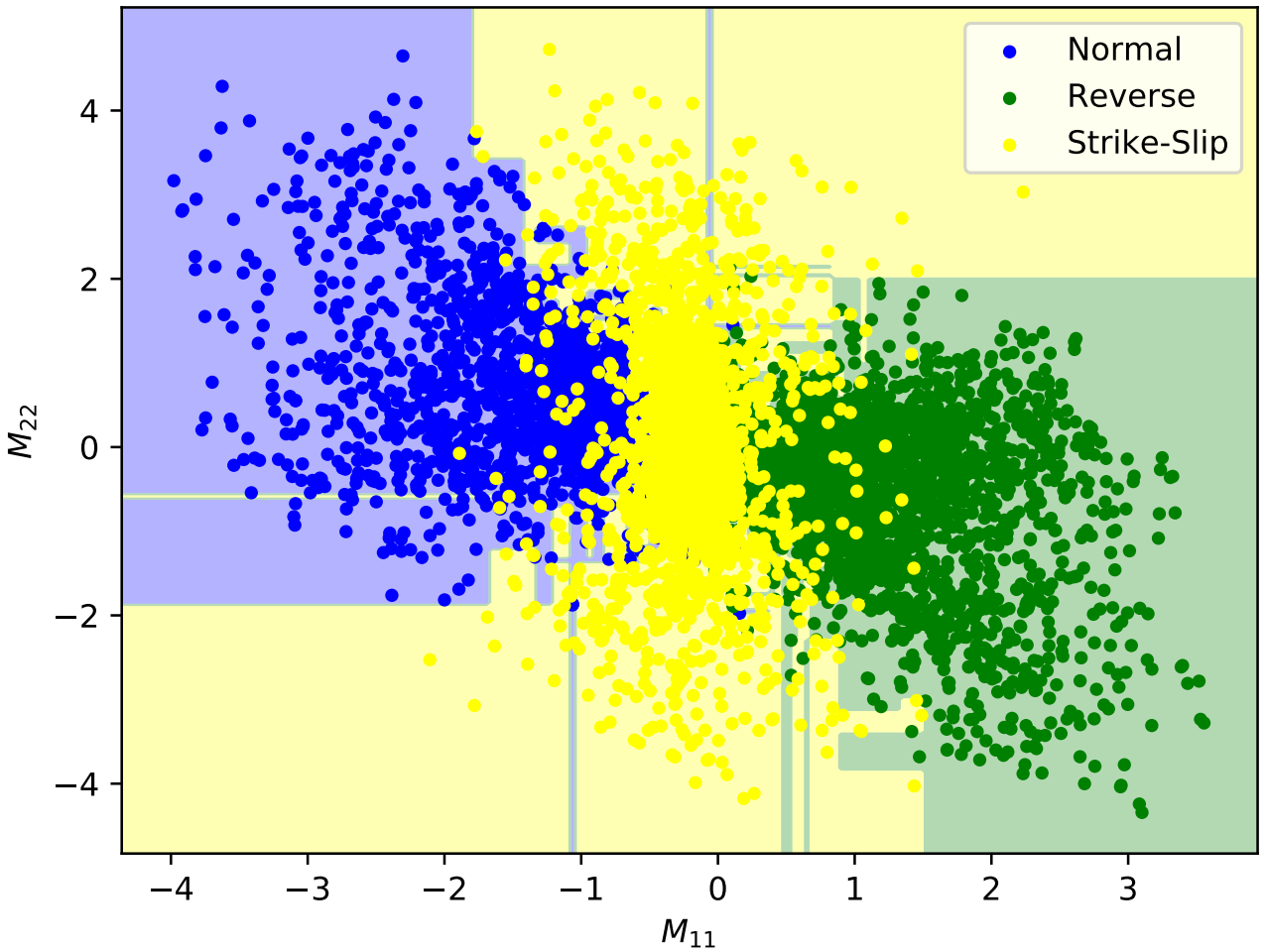
Machine Learning in Geophysics

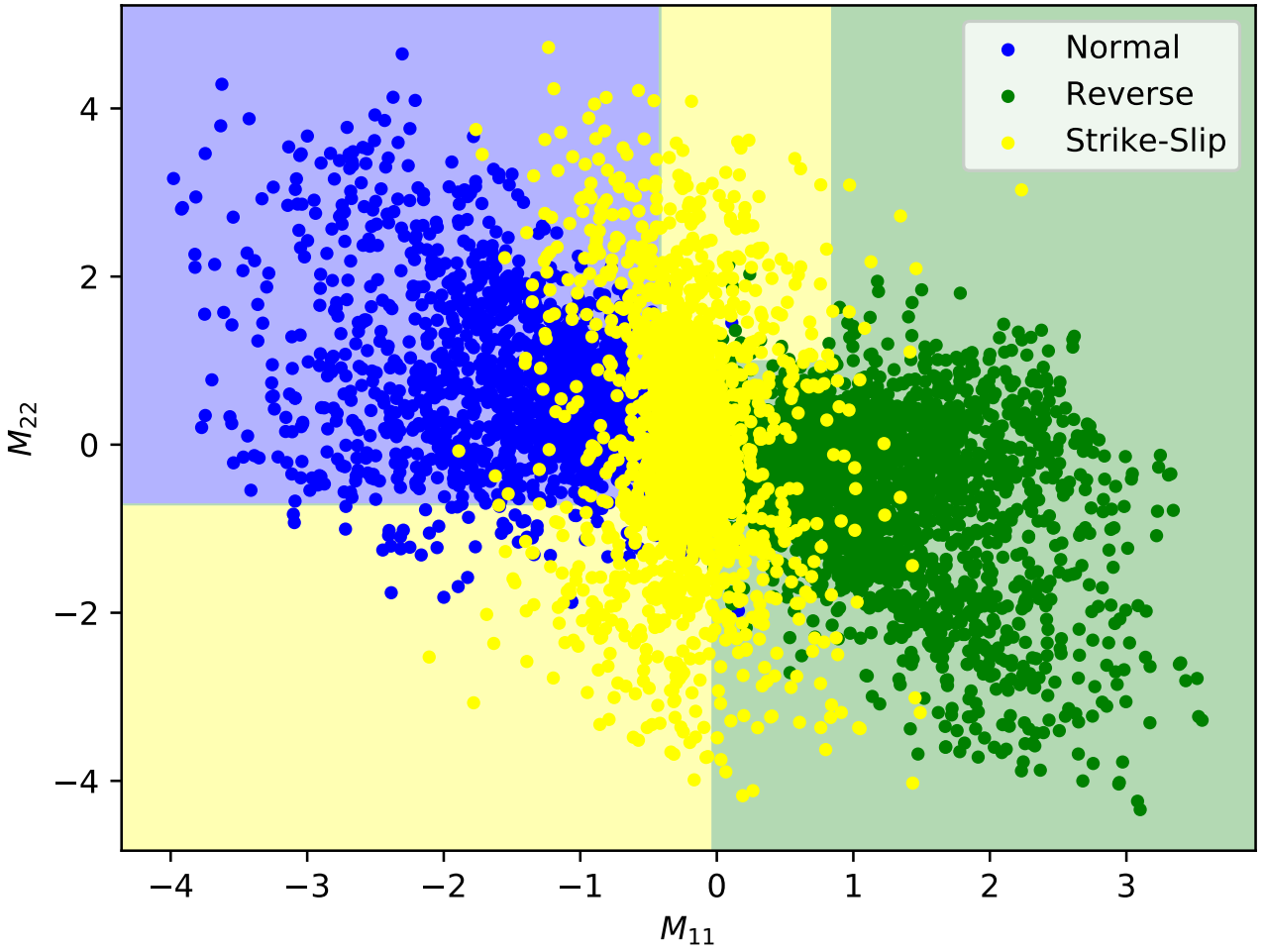
Lecture 6 – Decision Trees and Random Forests

Moment tensor example



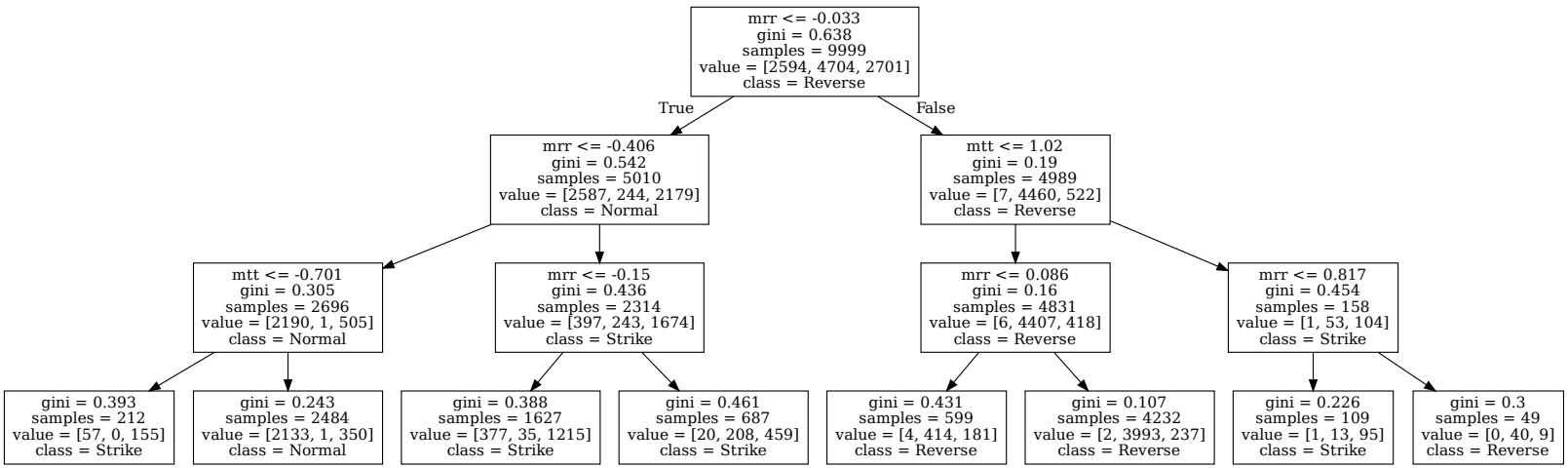
Decision tree with default parameters, prediction accuracy 1.0 for training data

Moment tensor example



Decision tree with Tree Depth 3, prediction accuracy 0.85 for training data

Decision tree



Decision tree with Tree Depth 3 , prediction accuracy 0.85
Gini measures the impurity of each node

$$G_i = 1 - \sum_{k=1}^{N} p_{i,k}^2 \text{ e.g. } 1 - \left(\frac{40}{49}\right)^2 - \left(\frac{9}{49}\right)^2 = 0.3$$

CART Algorithm

Classification and Regression Tree (CART) Algorithm

1. Find feature k, boundary t_k that minimize

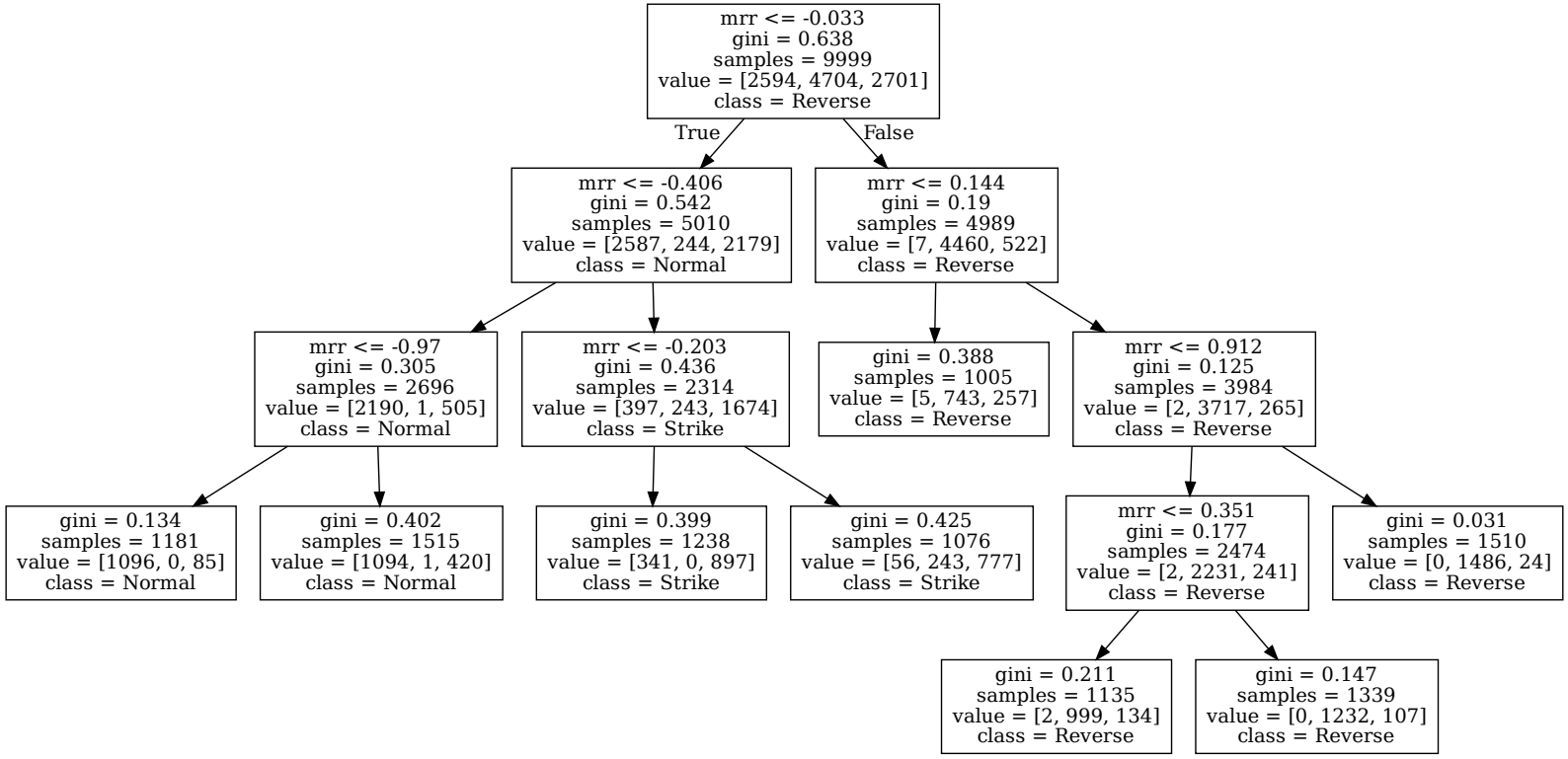
$$J(k, t_k) = \frac{m_{left}}{m} G_{left} + \frac{m_{right}}{m} G_{right}$$

2. Split dataset with (k, t_k)
3. Repeat 1 for every new node until maximum depth is reached or $G_i = 0$ for all i

Remarks

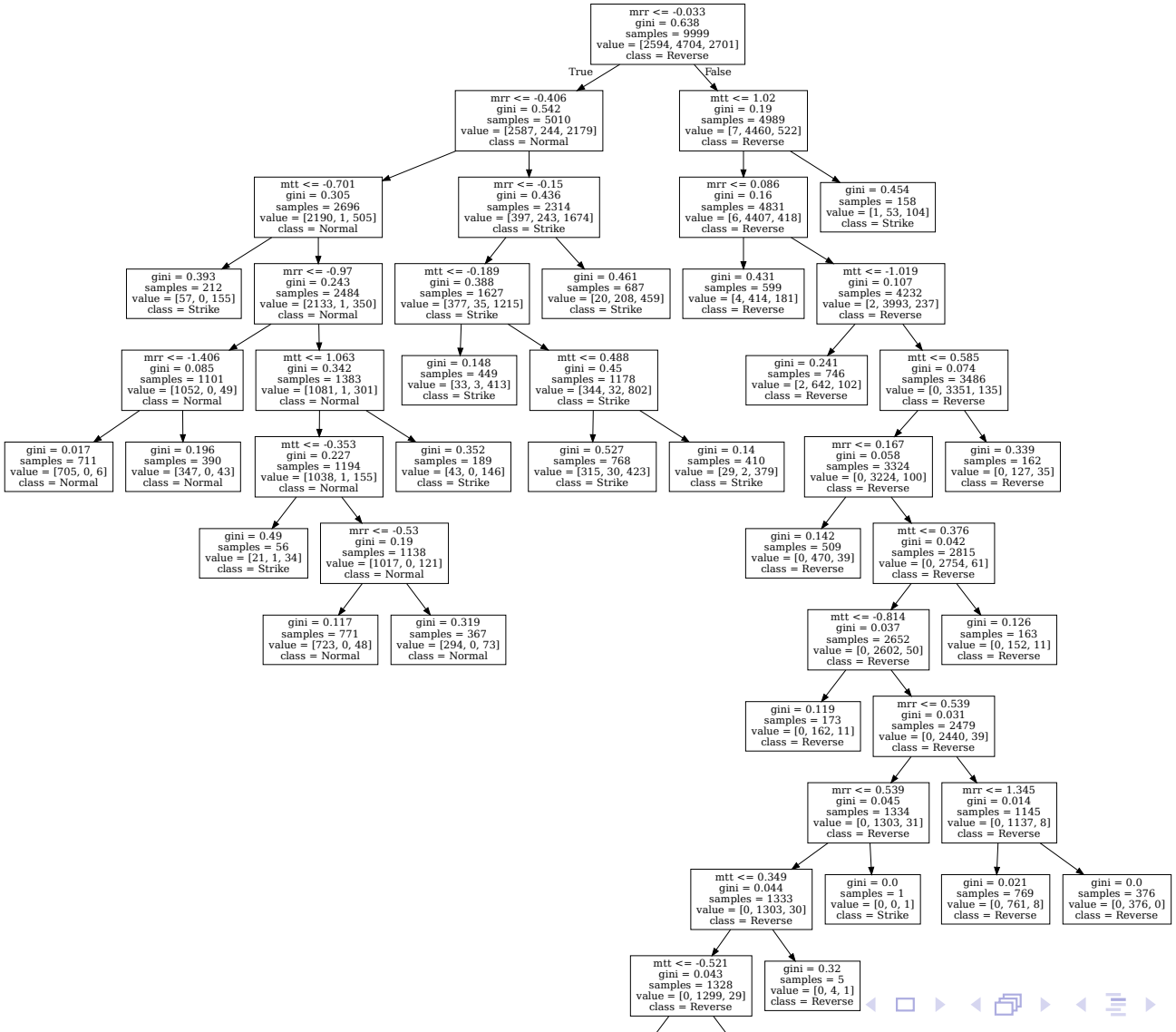
- The division at each level is performed without consideration for lower levels $\rightarrow$ tree might not be optimal, but optimal tree construction is NP-hard ($\mathcal{O}(\exp(m))$.
- Decision tree results can be easily understood *white box algorithm*
- Can estimate probability of a correct prediction $p_i = \frac{N_i}{N_{total}}$

Hyperparameters

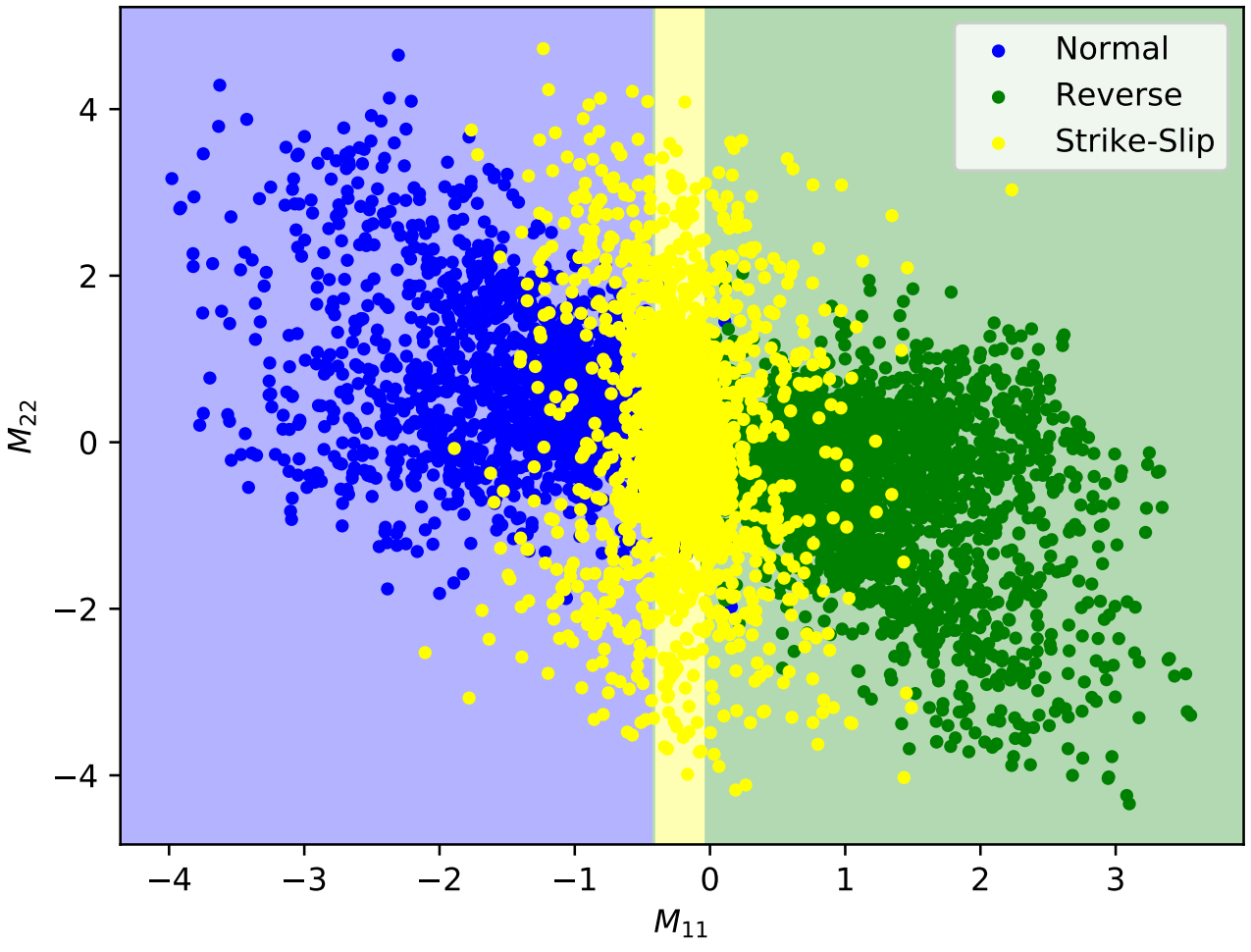


`min_sample_leaf`: Minimum number of samples in a leaf a node. `min_sample_split`: Minimum number of samples in a node that can be split.

min_sample_split: 1000

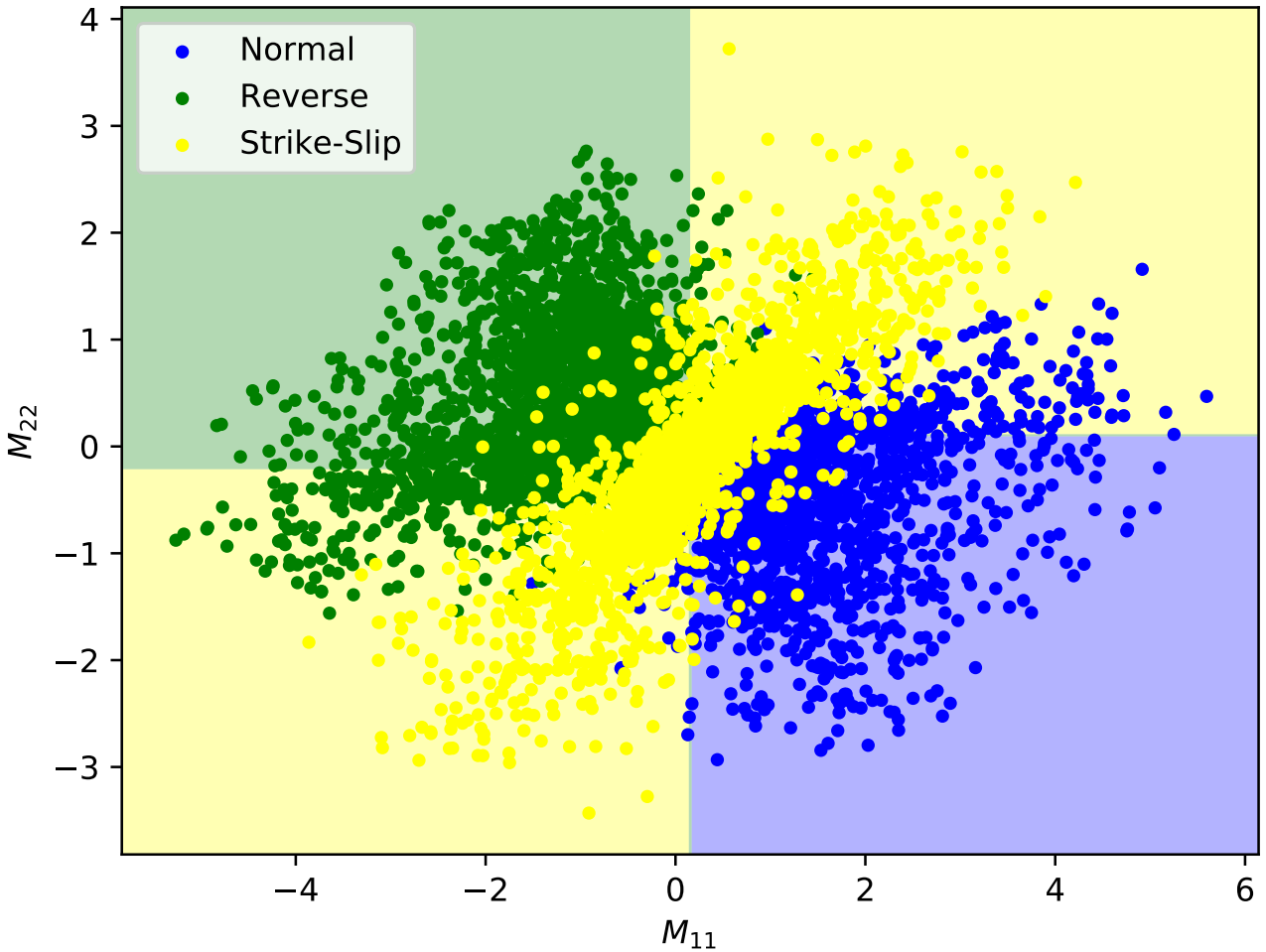


min_sample_leaf: 1000



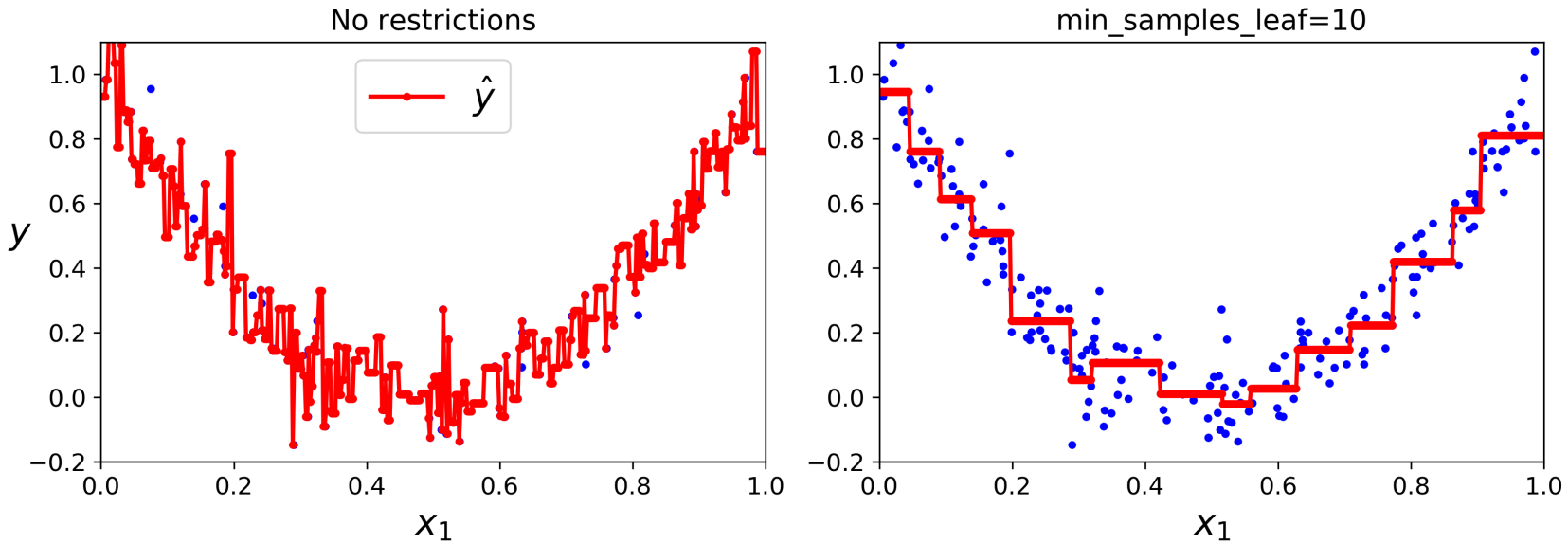
Original data

min_sample_leaf: 1000



With PCA (rotation of coordinate system)

Decision tree regression



We can use decision trees for regression, need to apply regularization, output stepwise constant

Basic idea

- ML often cannot capture all characteristics of training data
- Different settings/different methods reproduce different aspects
- Combine predictions from several methods/subsets for better predictions

Bagging

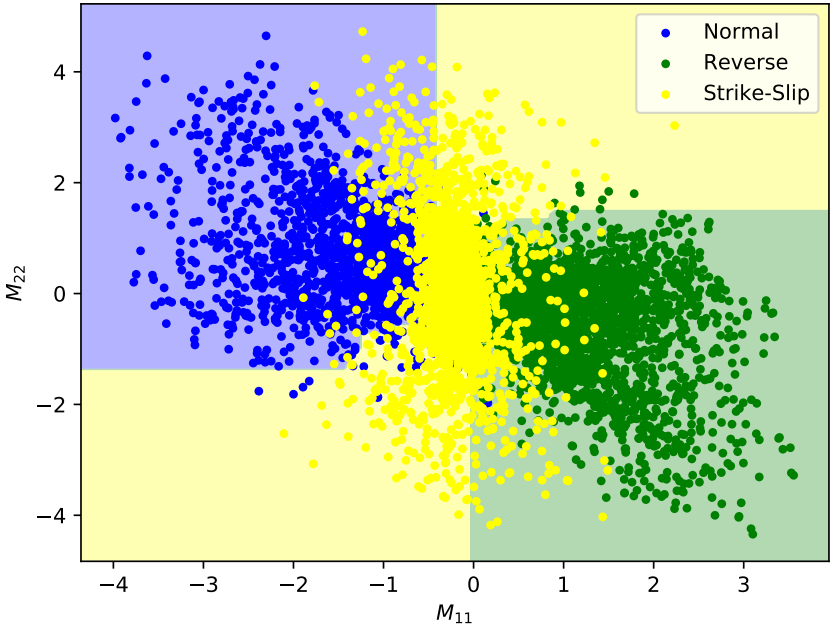
- Bootstrap-Aggregation (Bagging) creates training subsets by random sampling with replacement

$$(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5) \rightarrow (\boldsymbol{x}_2, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_3), (\boldsymbol{x}_1, \boldsymbol{x}_5, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_3)$$

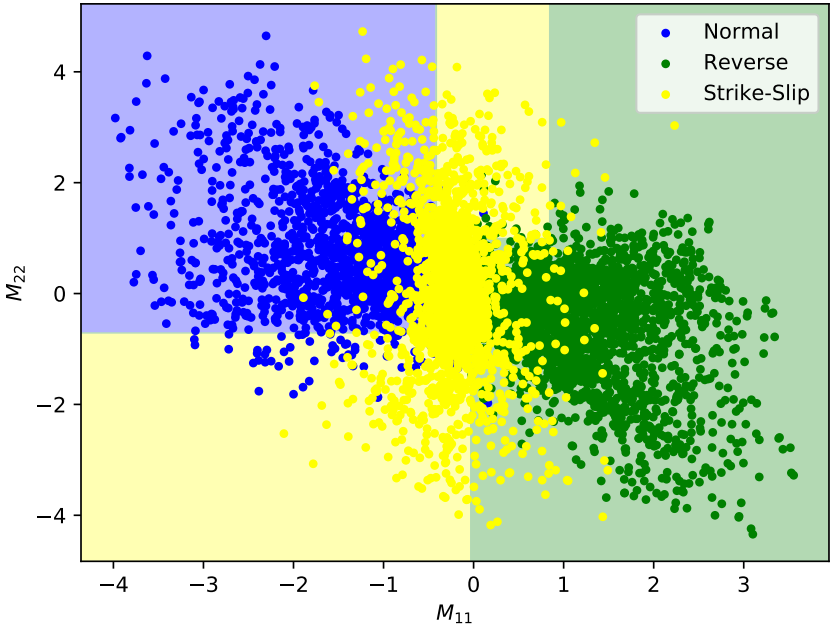
- Train classifier/regressor on each dataset
- Combine results by taking the most often occurring classification or average (regression)

Random Forest

Random Forest



Decision Tree



Ensemble of Decision Trees is called Random Forest

- Create training datasets through bagging
- Decide on split for individual trees based on subset of features
- Combine results by majority vote (hard voting)

Some properties

- Can get estimates of prediction probabilities (average of individual trees)
- Can get importance of features (gini importance), total decrease in node impurity
- Bagging means that not all training data points are used (Out-of-Bag), can be used to estimate accuracy without extra validation dataset (but still better to check).

Other ensemble methods

ExtraTrees: Extremely Randomized Trees, do not search for optimum threshold t_k, but use a random one

GradientBoosting: Chain classifiers to improve on error of previous one

Stacking: Train Classifier/Regressor on output of ensemble