Class core values

1. Be **respect**ful to yourself and others
2. Be **confident** and believe in yourself
3. Always do your **best**
4. Be **cooperative**
5. Be **creative**
6. Have **fun**
7. Be **patient** with yourself while you learn
8. Don’t be shy to **ask “stupid” questions**
9. Be **inclusive** and **accepting**

Week 2, Lecture 1

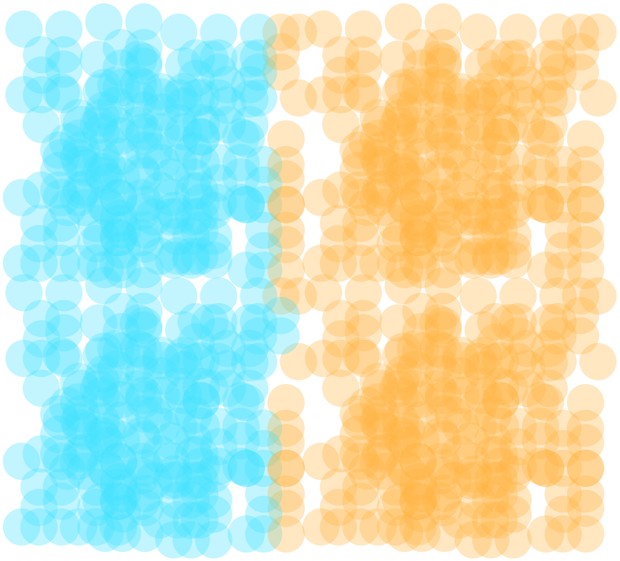
Trees, forests, and classification problems

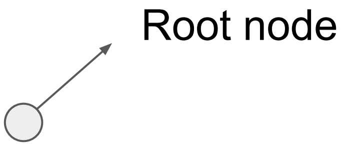
Learning Objectives

1. Describe the basic concept of a decision tree
2. Apply computational coding to write a simple decision tree
3. Describe the basic concept behind a random forest
4. Apply knowledge of depth and number of trees to tune a random forest
5. Apply computational tools to write and employ random forests to a real dataset
6. Practice knowledge of data preparation and performance measurement

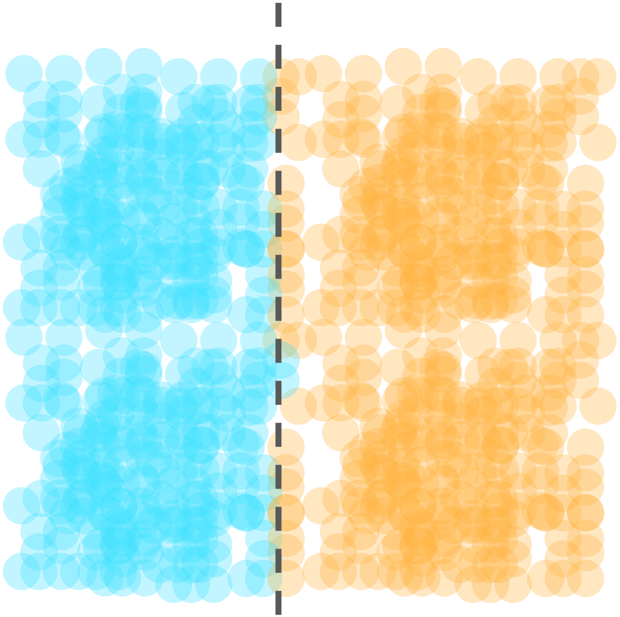
Decision trees are one of the simplest methods for binary classification

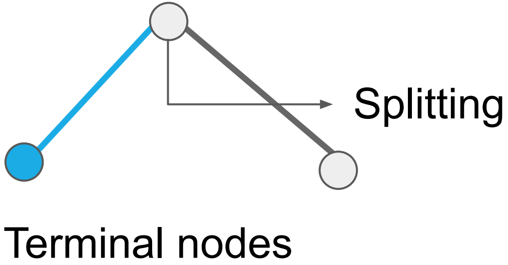
Decision trees are one of the simplest methods for binary classification



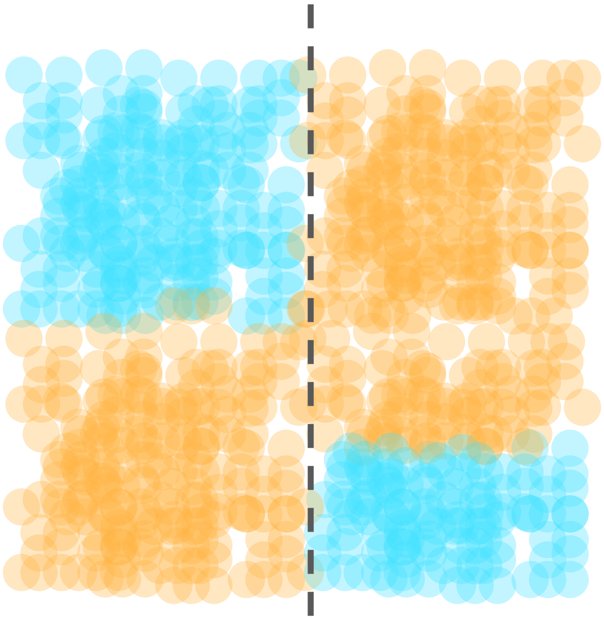


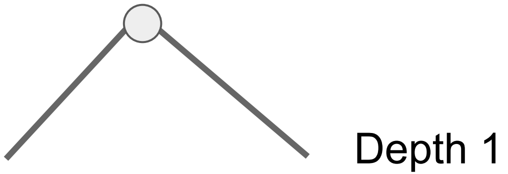
Decision trees are one of the simplest methods for binary classification



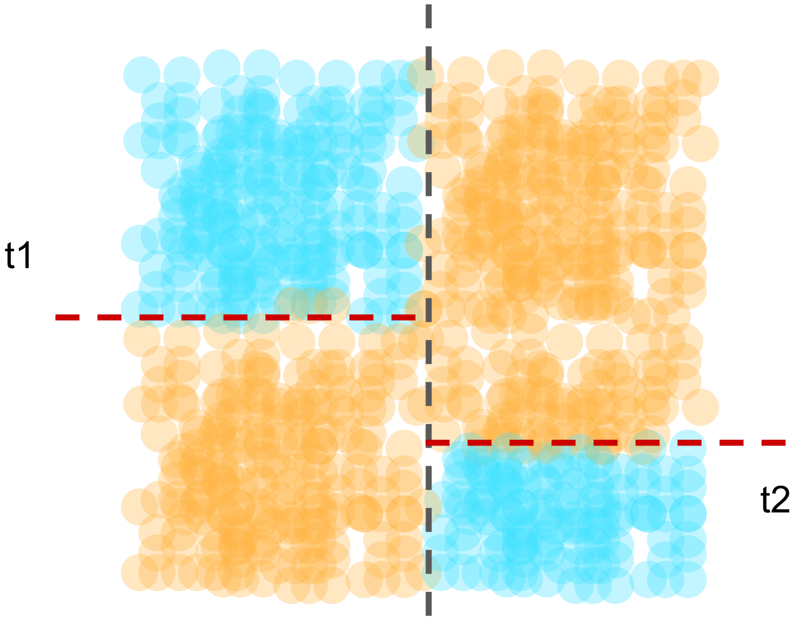


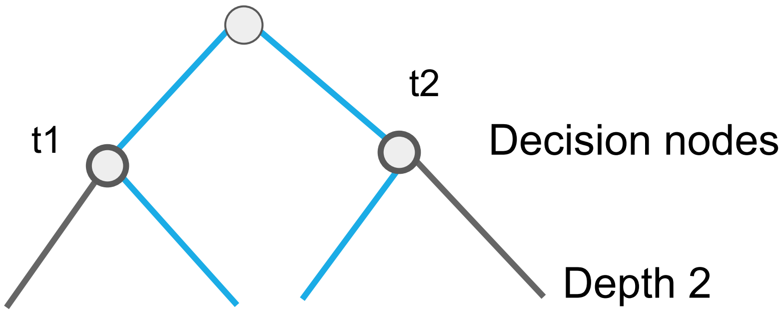
The depth of the decision tree can influence its performance



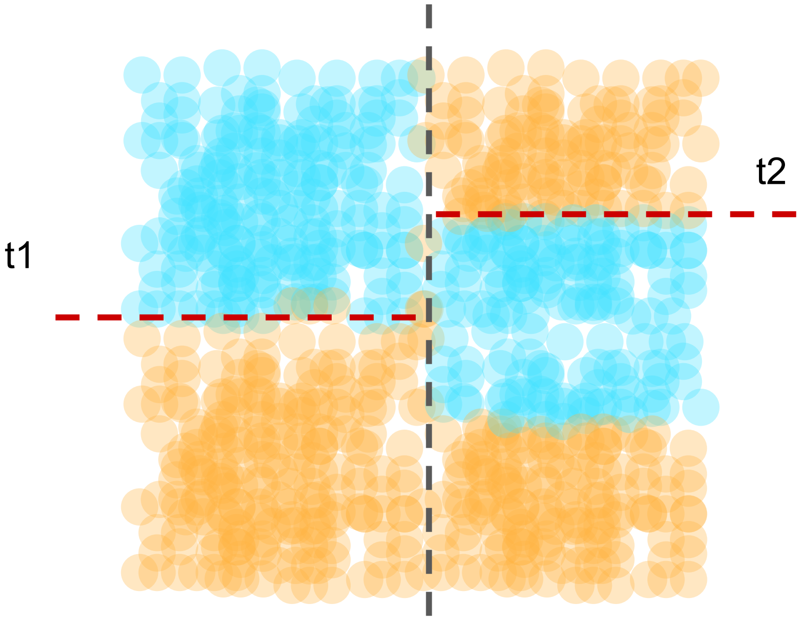


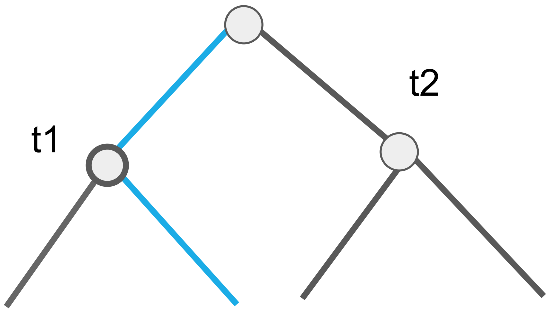
The depth of the decision tree can influence its performance



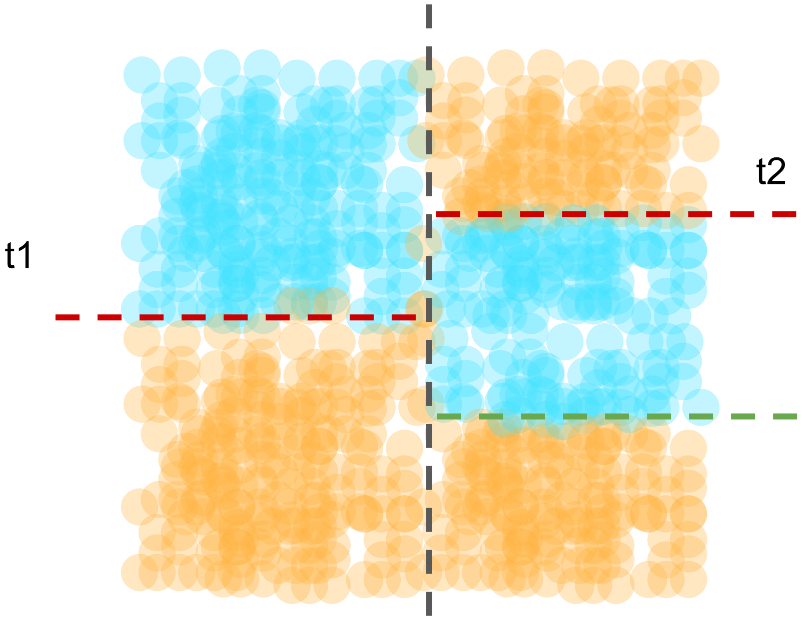


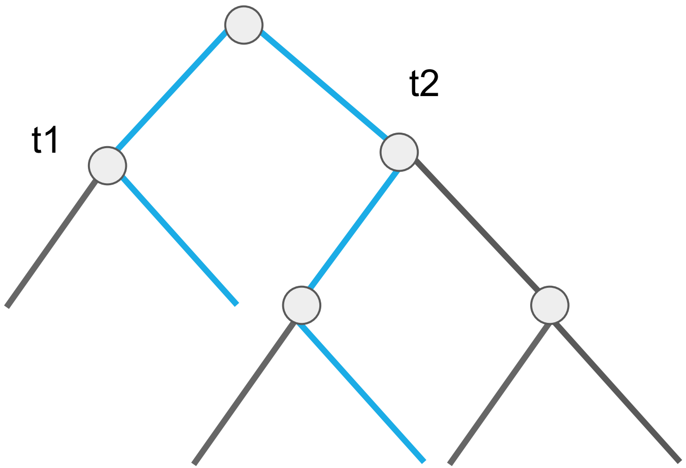
The depth of the decision tree can influence its performance



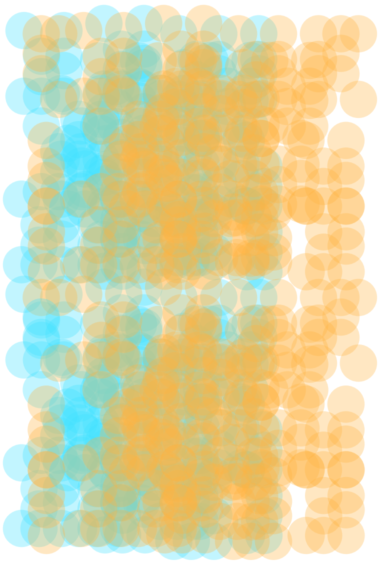


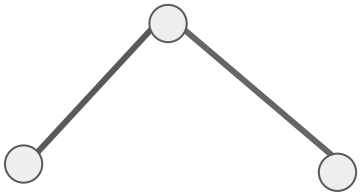
The depth of the decision tree can influence its performance



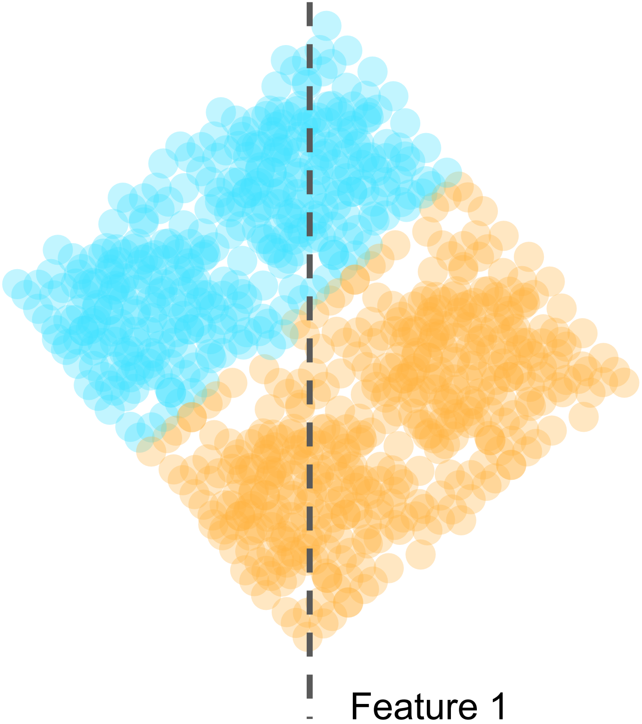


For decision trees to work, there needs to be an actual signal in the data

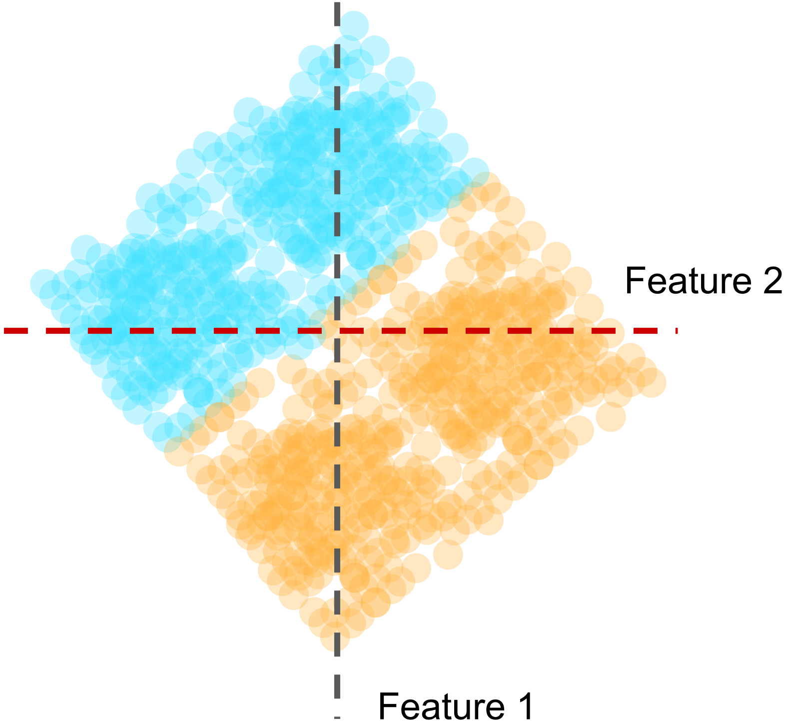




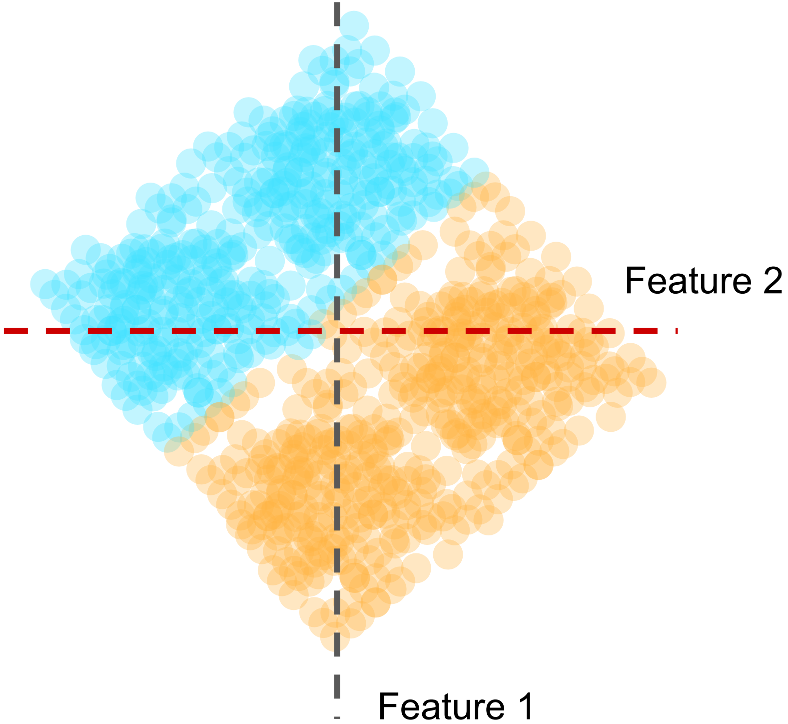
Feature engineering can help to find the best split

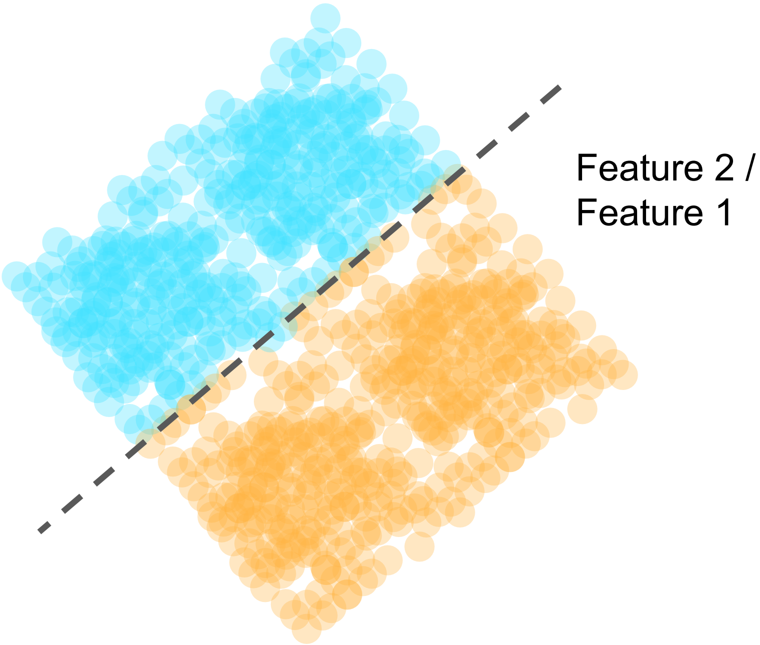


Feature engineering can help to find the best split

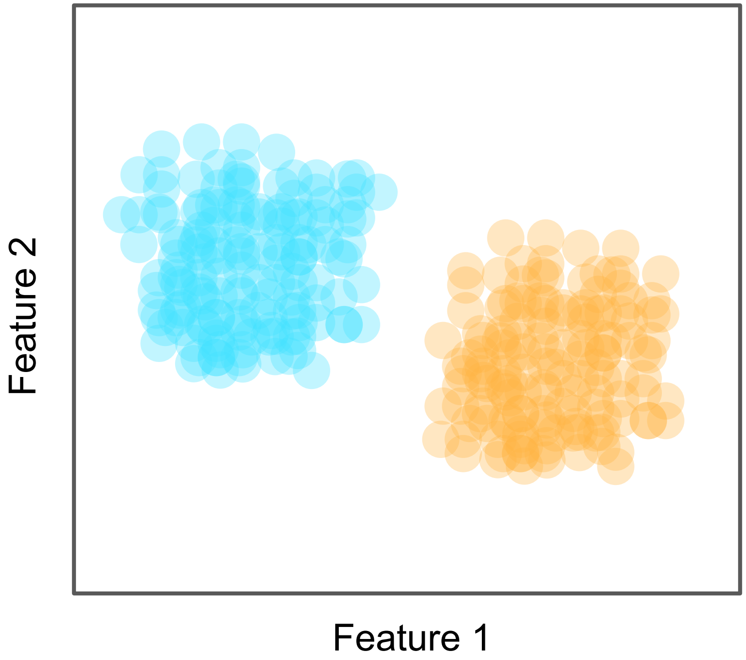


Feature engineering can help to find the best split

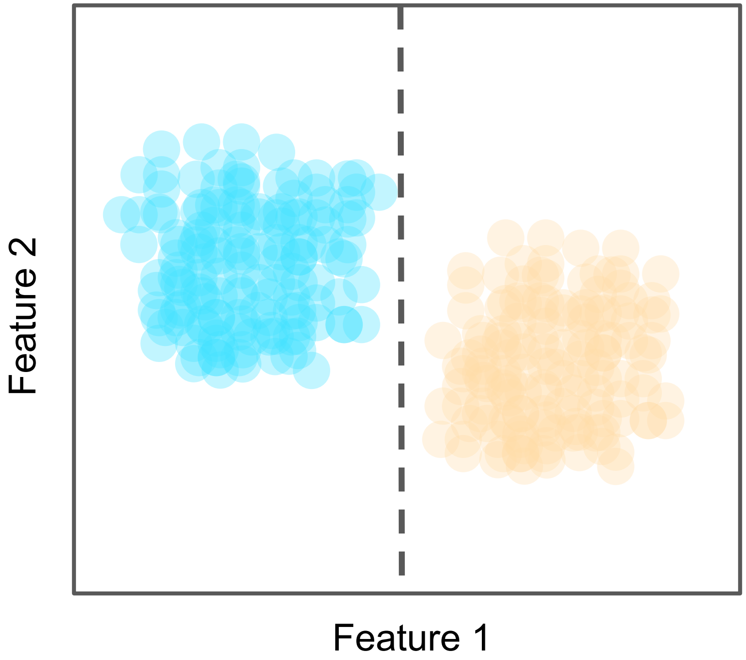




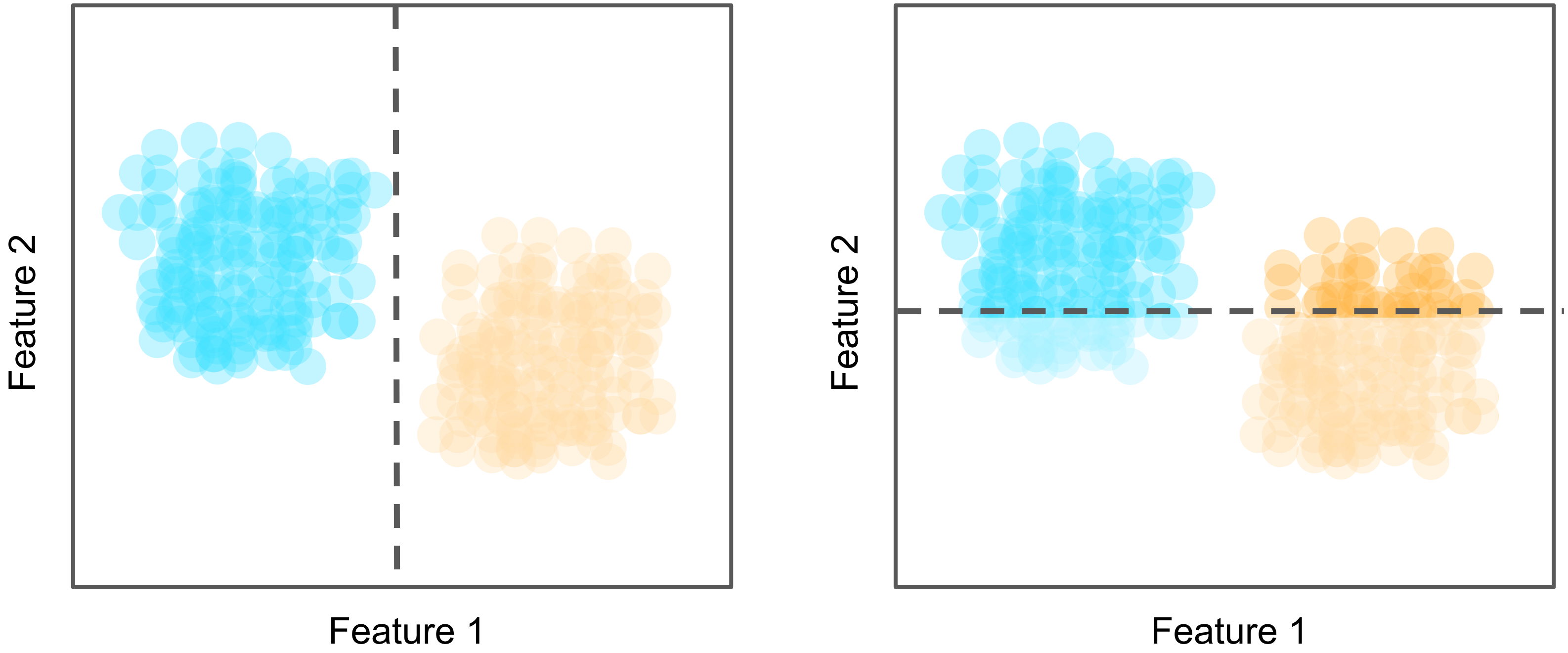
Decision trees optimize their decision to make the best split



Decision trees optimize their decision to make the best split

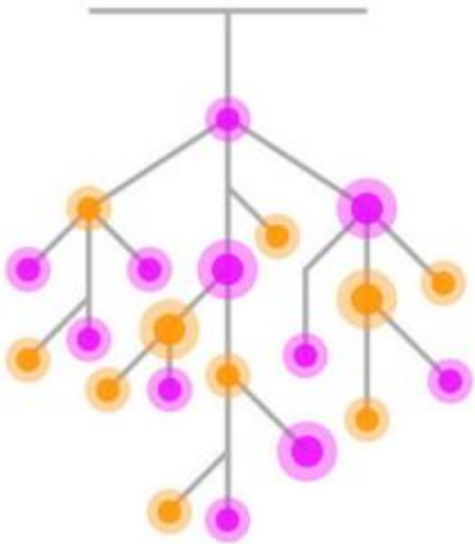


Decision trees optimize their decision to make the best split

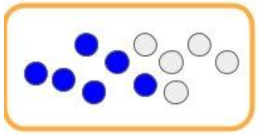


In-class activity

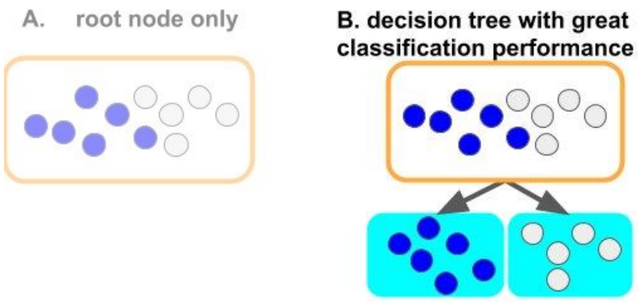
Writing a [very very] simple decision tree



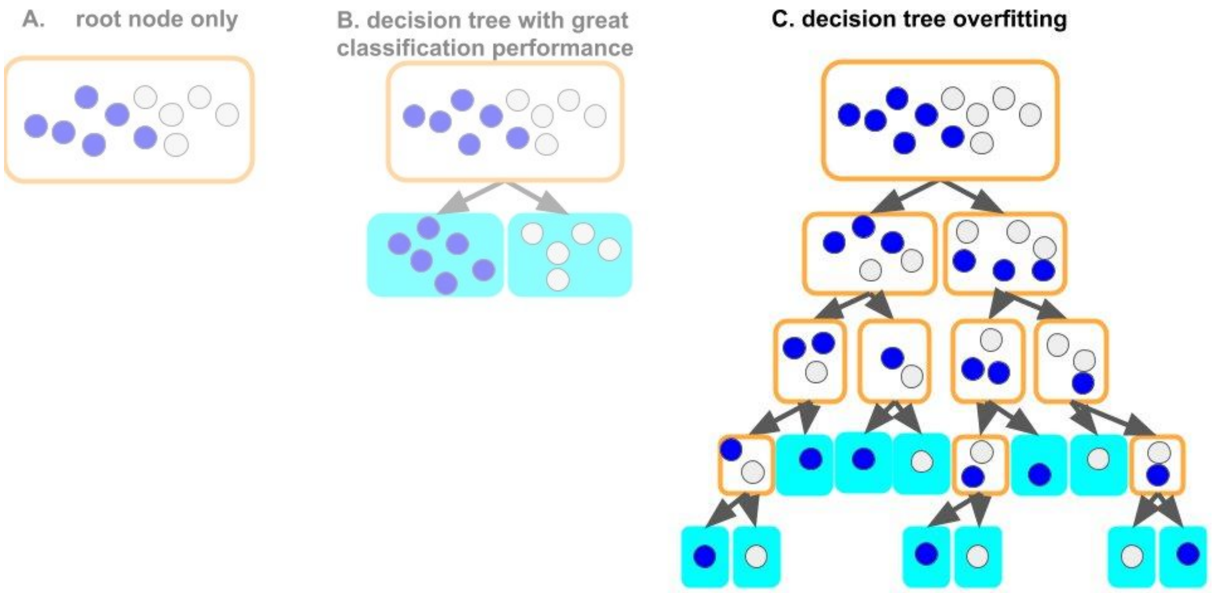
Due to their optimization procedure, decision trees are prone to overfitting



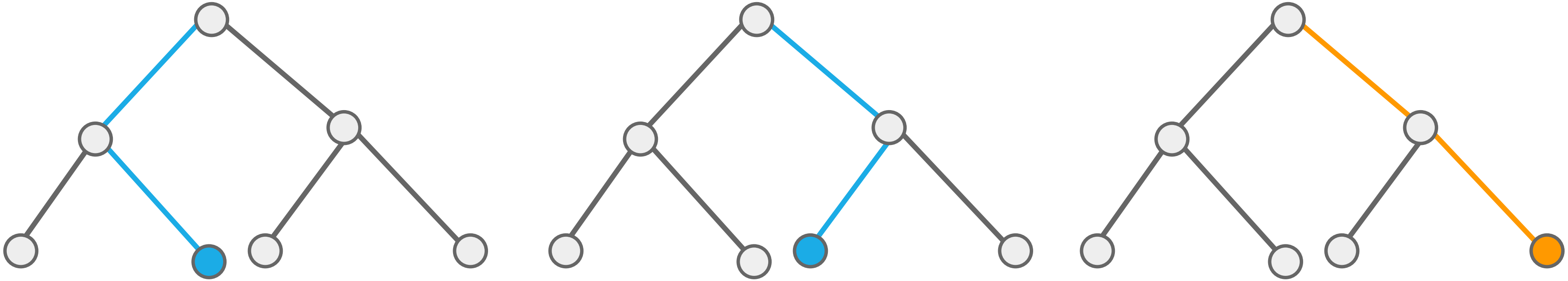
Due to their optimization procedure, decision trees are prone to overfitting



Due to their optimization procedure, decision trees are prone to overfitting



Combining different [shorter] trees (a forest) can result in a better performance

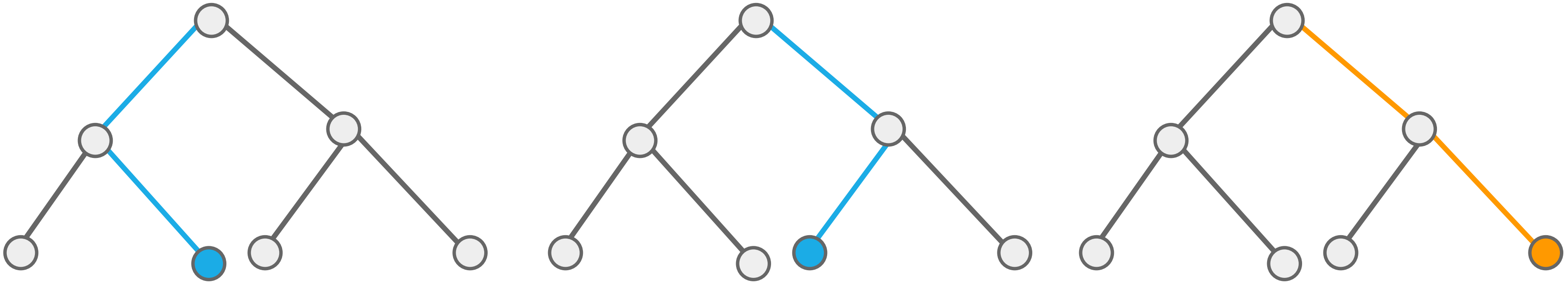


In-class activity

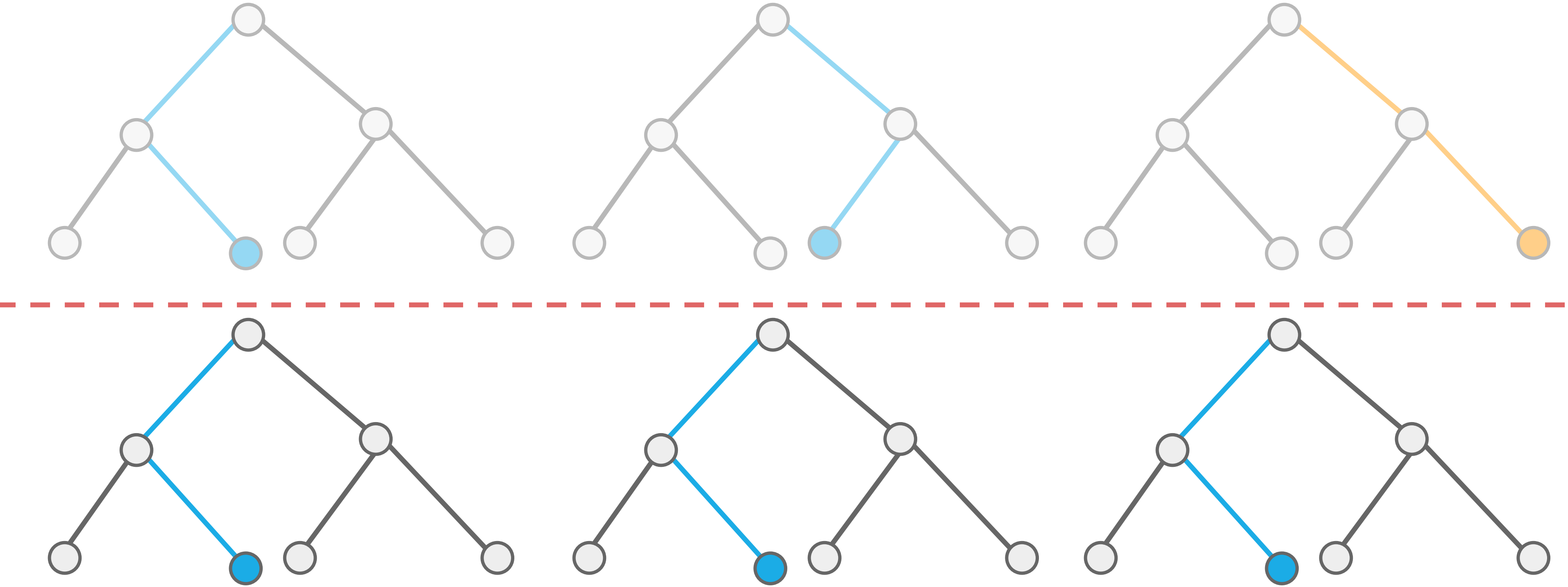
Voting

preg	plas	pres	skin	test	mass	pedi	age	Class
5	155	84	44	545	38.7	0.619	34	?
8	196	76	29	280	37.5	0.605	57	?
9	122	56	0	0	33.3	1.114	33	?
1	109	60	8	182	25.4	0.947	21	?
6	190	92	0	0	35.5	0.278	66	?

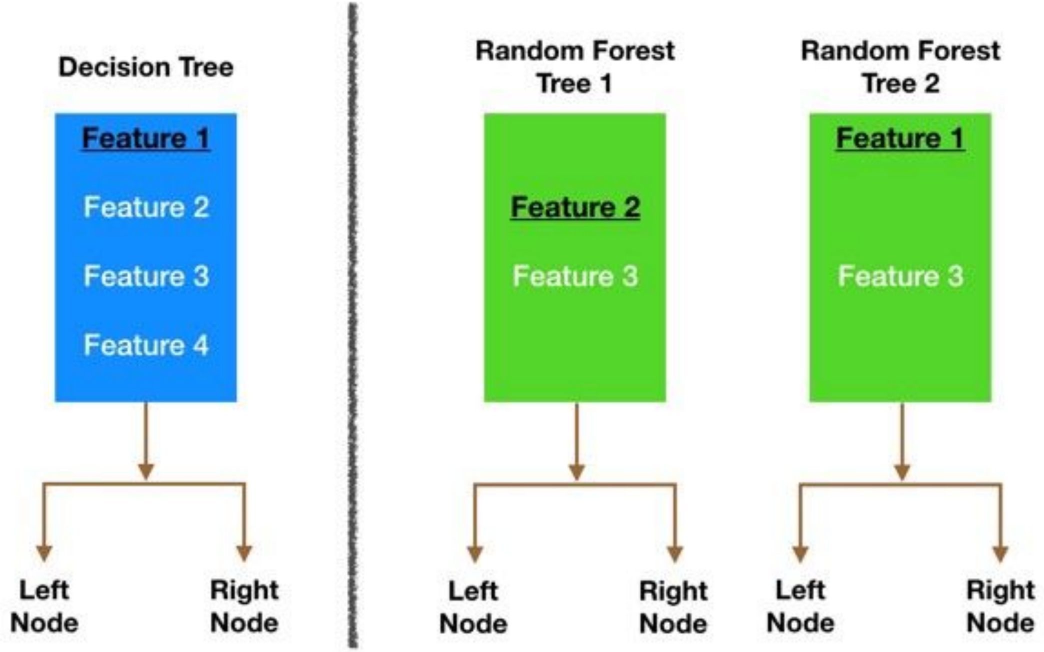
For random forests to work, the trees need to have low correlation



For random forests to work, the trees need to have low correlation



Feature randomness is one way to make sure trees diversify each other



Bagging (bootstrap aggregating) is another method to ensure diversity among trees

N

N-k

+ k

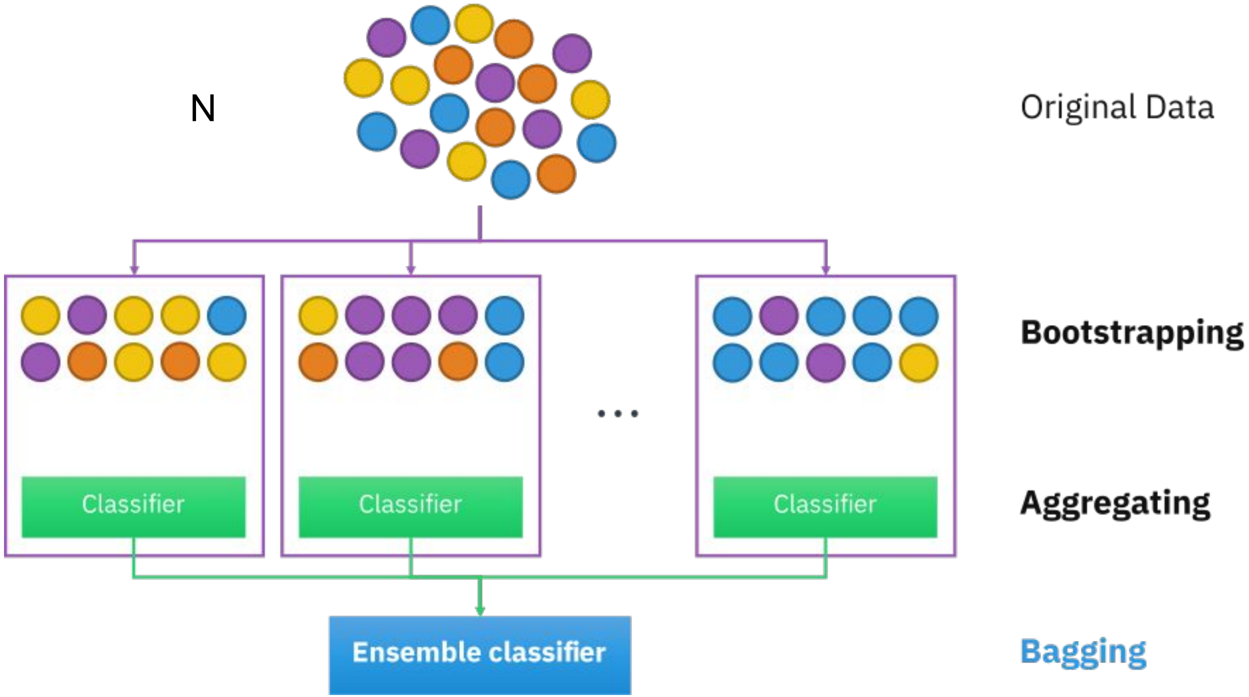
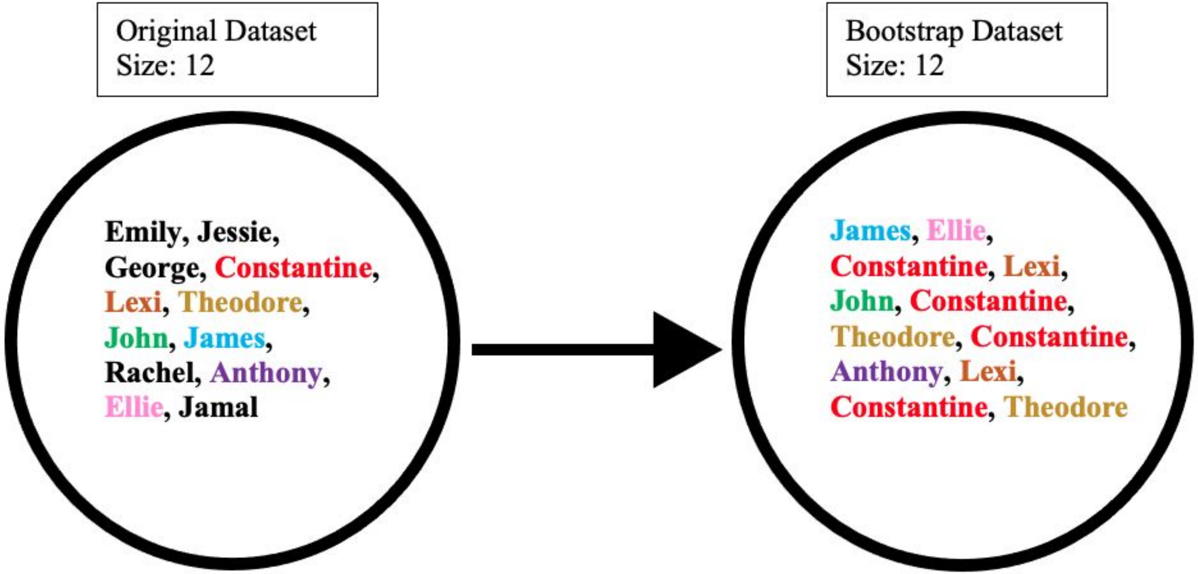


Image from Wikipedia: bootstrap aggregating

Bagging (bootstrap aggregating) is another method to ensure diversity among trees



Random forests are similar to a wise crowd

5 important features:

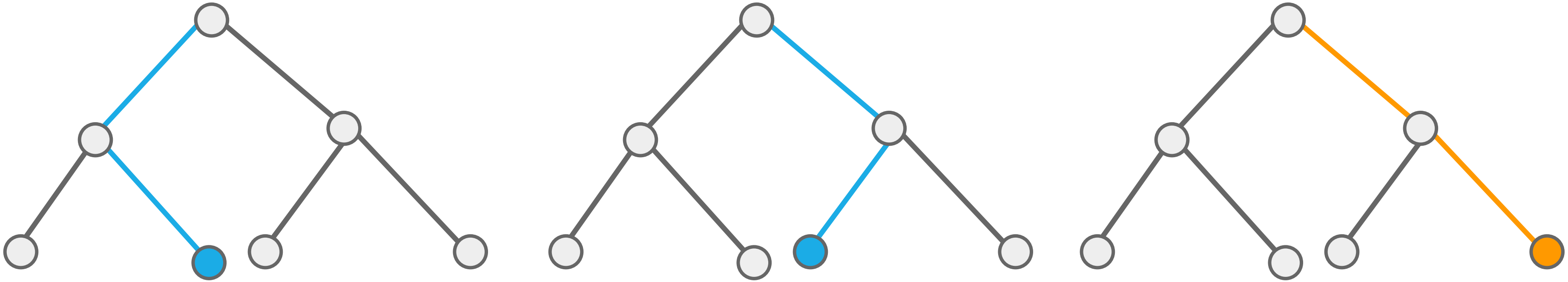
Random forests are similar to a wise crowd

5 important features:

1. Diversity of opinion
2. Independence
3. Decentralization
4. Aggregation
5. Trust

In-class activity

Building a random forest



Notes on random forests

Pros:

1. Lower variance than decision trees (slower risk of overfitting)
2. Easy to set-up and powerful
3. Works well with non-linear data
4. Easy preparation and deals well with missing/noisy data

Notes on random forests

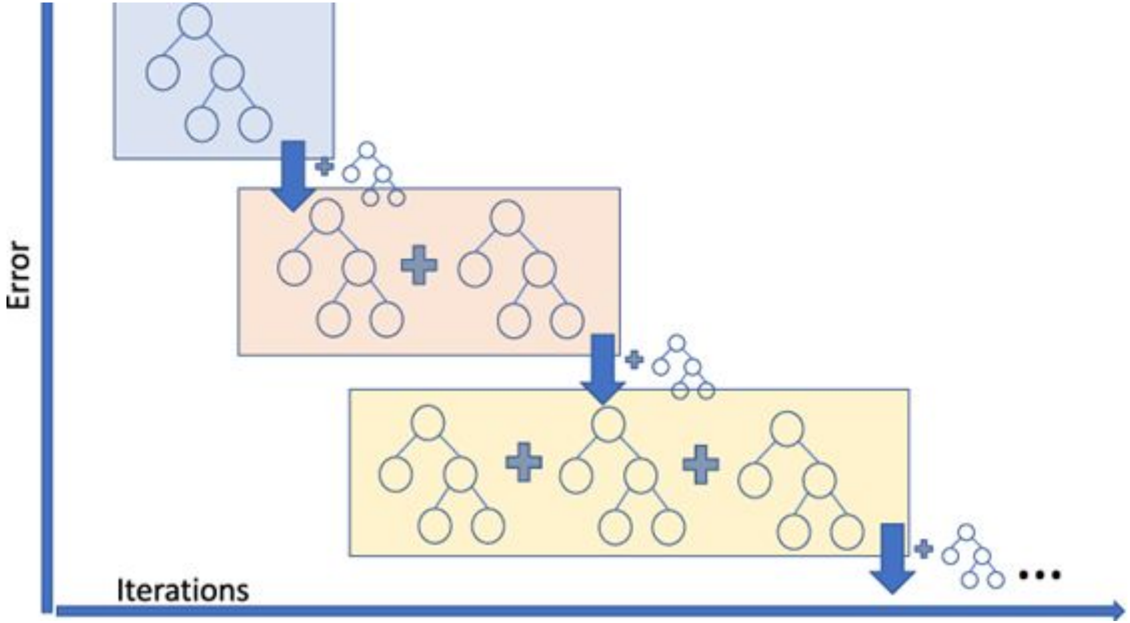
Pros:

1. Lower variance than decision trees (slower risk of overfitting)
2. Easy to set-up and powerful
3. Works well with non-linear data
4. Easy preparation and deals well with missing/noisy data

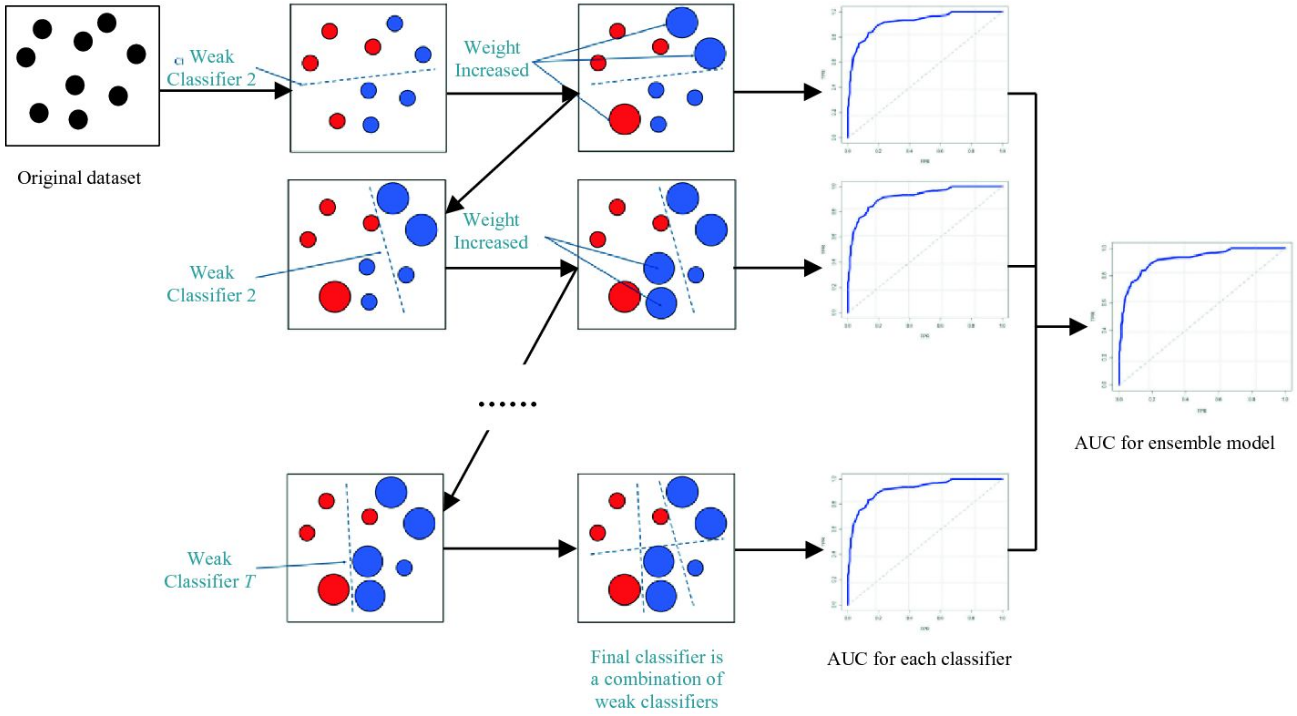
Cons:

1. Slow
2. Very dependent on datasets
3. More complex → not as interpretable
4. Can't predict beyond the range of training data

Gradient boosting is another way to connect decision trees



Gradient boosting is another way to connect decision trees



Some notes on gradient boosting

- Can perform better than RF if build properly
- Not appropriate for noisy data (overfitting)
- Harder to tune

Next lecture:

Simple methods can take you a long way!