GENERAL ASSEMBLY

Intro to Ensemble Methods & Bagging

Patrick Smith

LEARNING OBJECTIVES

- Explain the power of using ensemble classifiers
- Know the difference between a base classifier and an ensemble classifier
- Describe how bagging works
- Use the bagging classifier in scikit-learn

PRE-WORK

- Perform a classification
- Use label encoder
- Use cross validation to evaluate model performance

OPENING

What is an Ensemble Method?

What is an Ensemble Method?

What is Bagging?

What classifiers have we learned about thusfar? Which is your favorite?

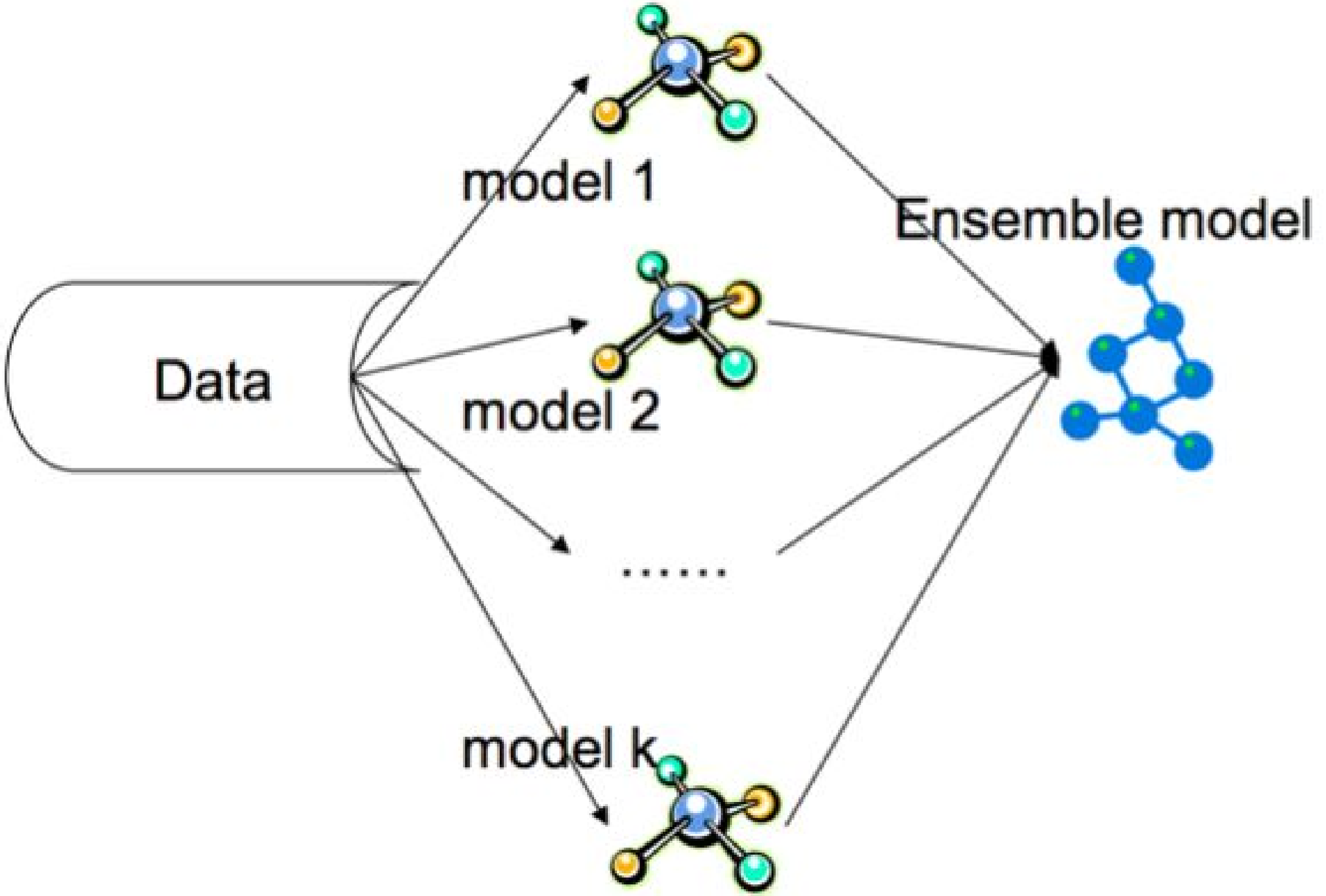
How do we assess the goodness of a model?

How could we improve the performance of a model?

Introduction: Ensemble Techniques

What is an Ensemble Method?

Ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions.


model 1
Ensemble model
Data
model 2
.....
model k

What is an Ensemble Method?

Ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions.

When might this be useful?

There are two families of Ensemble Methods

1. Averaging Methods

- Driving principle is to build several estimators independently and then to average their predictions.
- On average, the combined estimator is usually better than any of the single base estimator because its variance is reduced.

EX: Random Forest and Bagging

There are two families of Ensemble Methods

2. Boosting Methods

- Base estimators are built sequentially and one tries to reduce the bias of the combined estimator.
- The motivation is to combine several weak models to produce a powerful ensemble. We will discuss these in a future lecture.

EX: AdaBoost and Gradient Tree Boosting

The Hypothesis Space

In any supervised learning task, our goal is to make predictions of the true classification function f by learning the classifier h. In other words we are searching in a certain hypothesis space for the most appropriate function to describe the relationship between our features and the target.

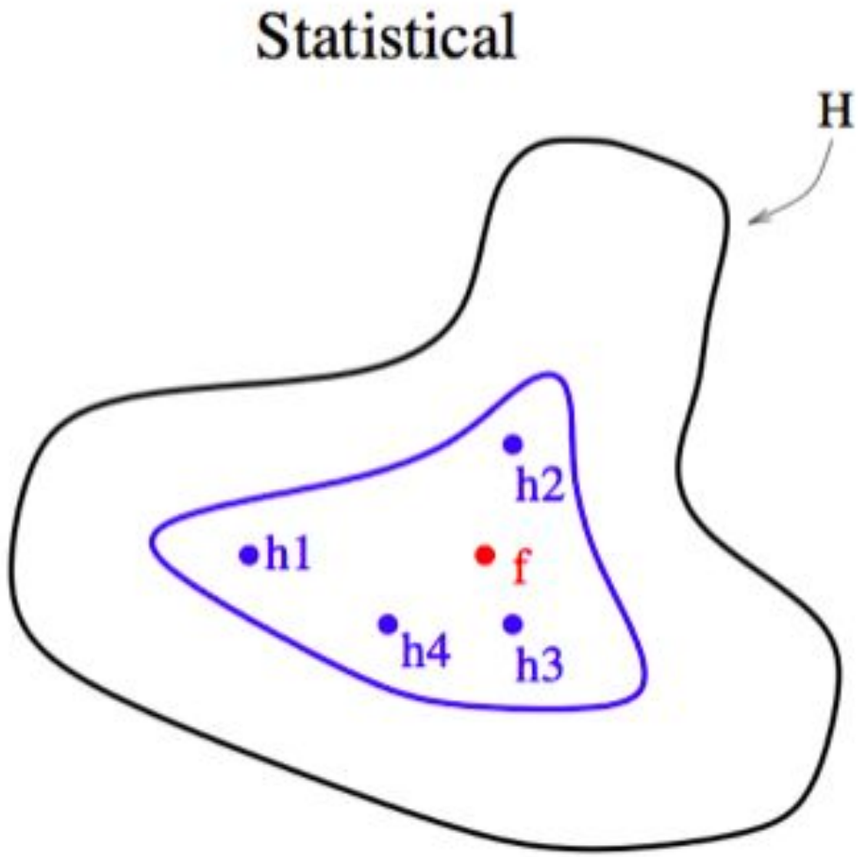
Can you give an example of how this search is performed using one of the classifiers you know?

What reasons could be preventing our hypothesis to reach perfect score?

Intro to Ensemble Methods

The Statistical Problem:

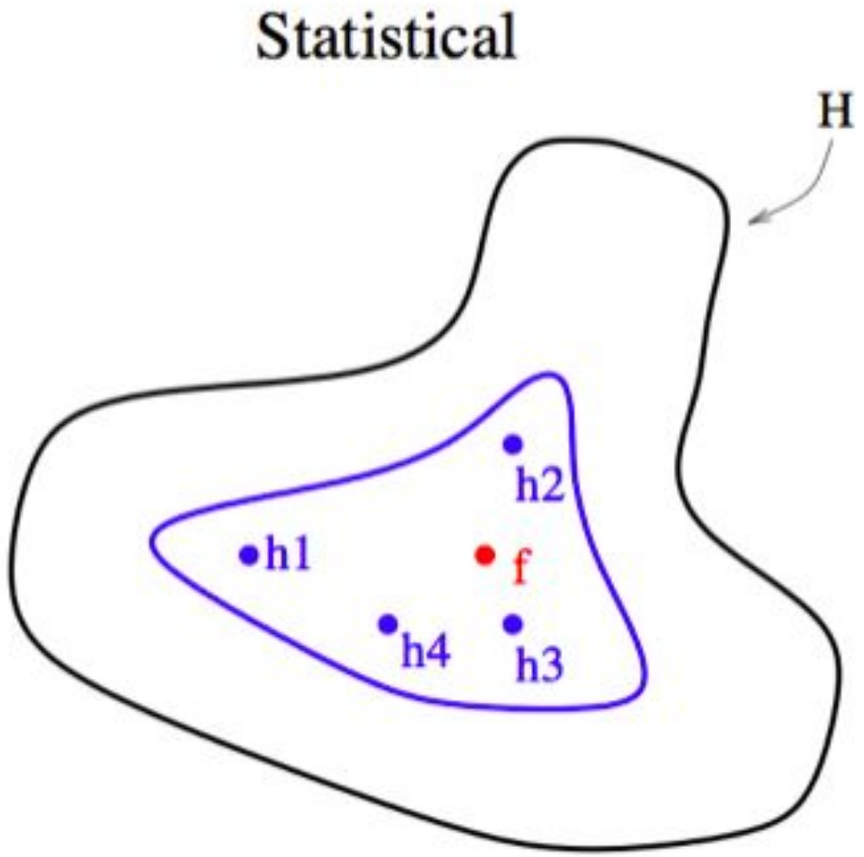
If the amount of training data available is small, the base classifier will have difficulty converging to ***h***.



Intro to Ensemble Methods

The Statistical Problem:

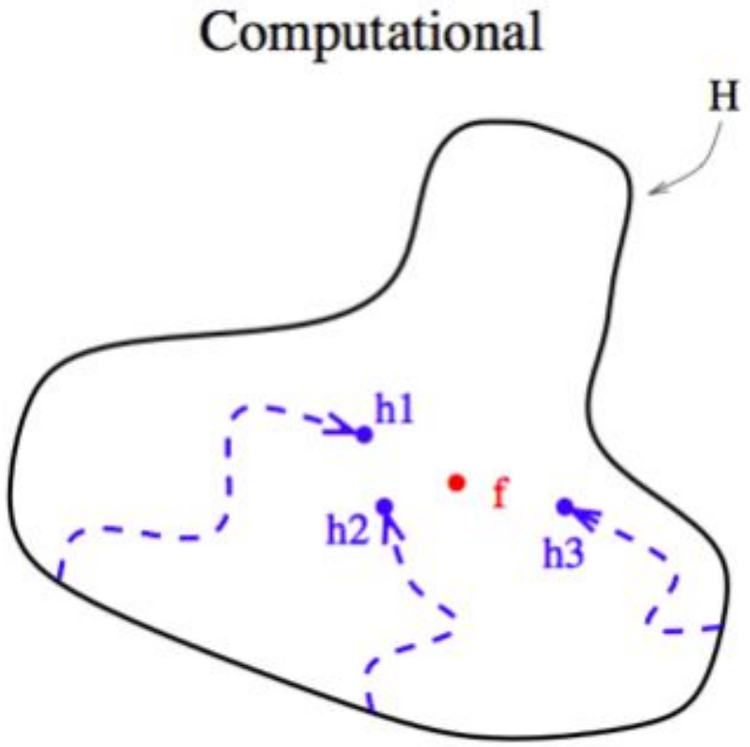
An ensemble classifier can mitigate this problem by "averaging out" base classifier predictions to improve convergence.



Intro to Ensemble Methods

The Computational Problem

Even with sufficient training data, it may still be computationally difficult to find the best classifier ***h***.

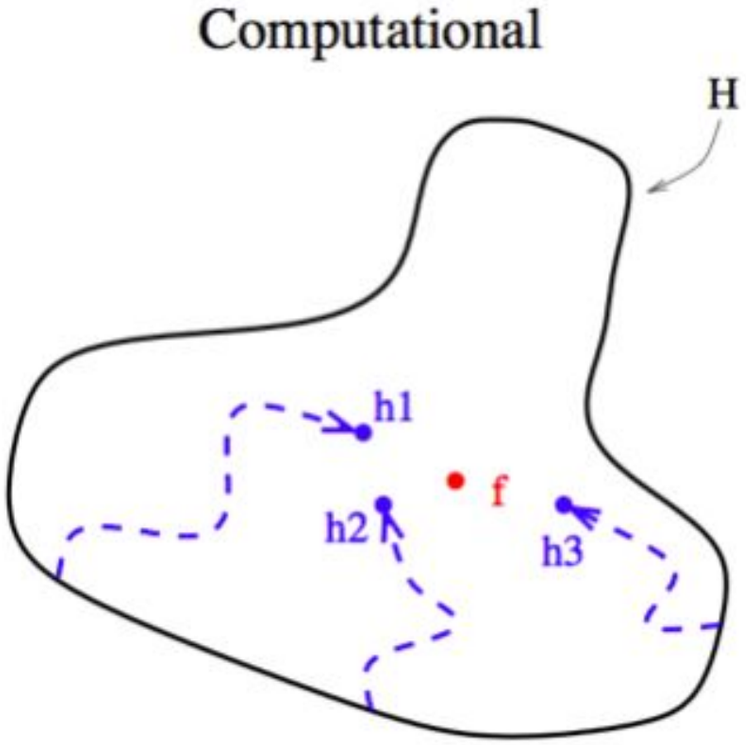


Intro to Ensemble Methods

The Computational Problem

Even with sufficient training data, it may still be computationally difficult to find the best classifier ***h***.

The true function ***f*** is often best approximated by using several starting points to explore the hypothesis space



Intro to Ensemble Methods

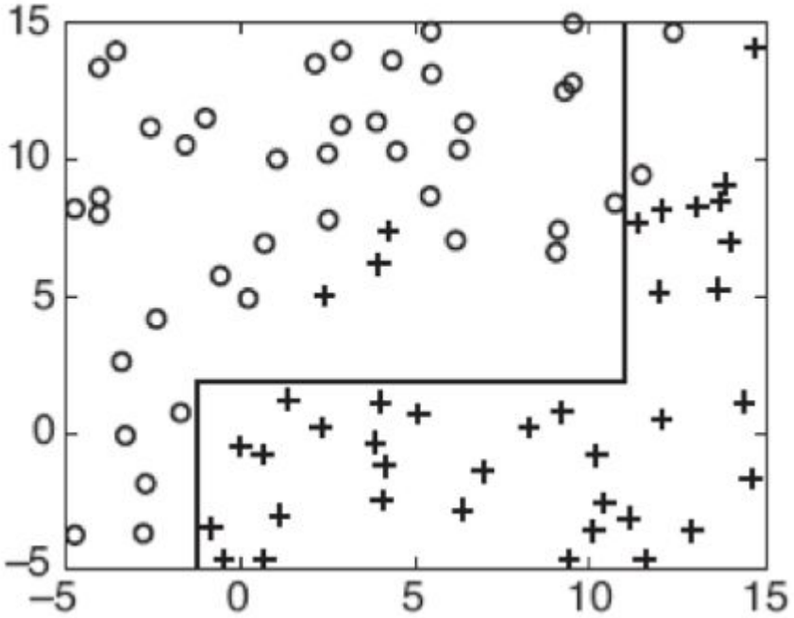
The Representational Problem

Sometimes f cannot be expressed in terms of our hypothesis at all. To illustrate this, suppose we use a decision tree as our base classifier. A decision tree works by forming a rectilinear partition of the feature space, i.e it always cuts at a fixed value along a feature.

Intro to Ensemble Methods

The Representational Problem

Sometimes f cannot be expressed in terms of our hypothesis at all. To illustrate this, suppose we use a decision tree as our base classifier. A decision tree works by forming a rectilinear partition of the feature space, i.e it always cuts at a fixed value along a feature.

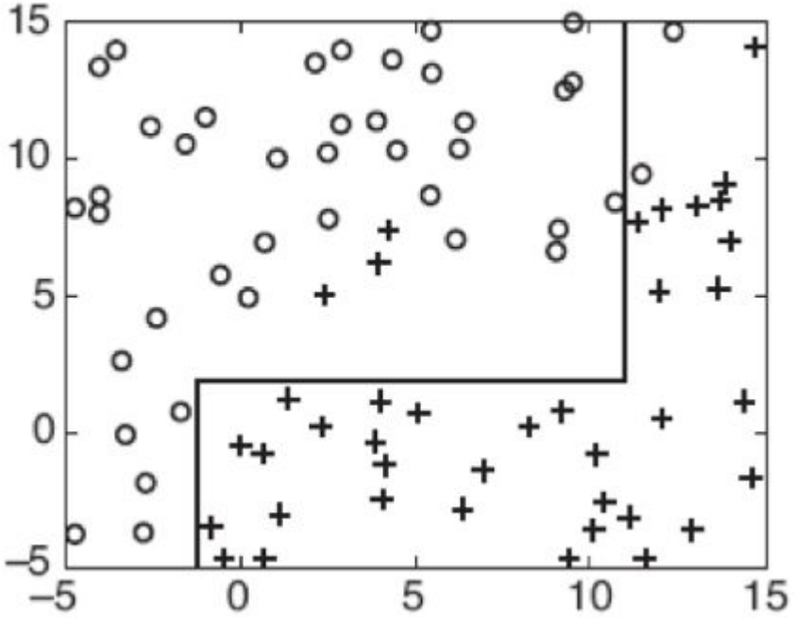


What is f is a diagonal line?

Intro to Ensemble Methods

The Representational Problem

Then it cannot be represented by finitely many rectilinear segments, and therefore the true decision boundary cannot be obtained by a decision tree classifier.

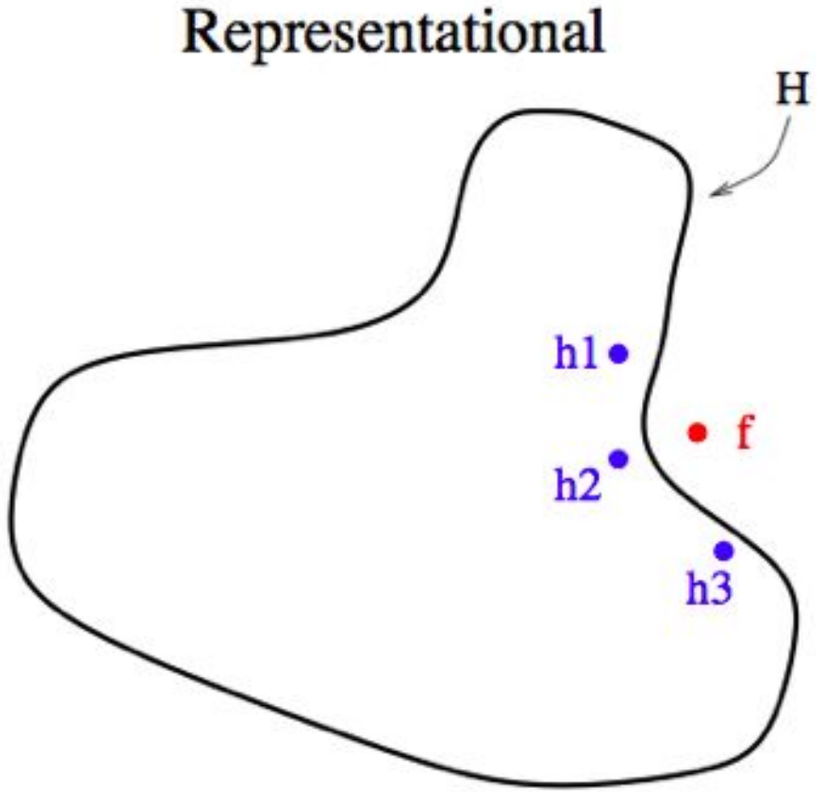


What is f is a diagonal line?

Intro to Ensemble Methods

The Representational Problem

However, it may be still be possible to approximate f or even to expand the space of representable functions using ensemble methods.



Intro to Ensemble Methods

Characteristics of Ensemble Methods

In order for an ensemble classifier to outperform a single base classifier, the following conditions must be met:

- **accuracy:** base classifiers outperform random guessing
- **diversity:** misclassifications must occur on different training examples

Practice: Real World Applications of Ensemble Methods

Introduction: Bagging

Intro to Ensemble Methods

Bagging (Bootstrap Aggregating)

Bagging or bootstrap aggregating is a method that involves manipulating the training set by resampling

Intro to Ensemble Methods

Bagging (Bootstrap Aggregating)

Bagging or bootstrap aggregating is a method that involves manipulating the training set by resampling

We learn ***k*** base classifiers on ***k*** different samples of training data

Intro to Ensemble Methods

Bagging (Bootstrap Aggregating)

Bagging or bootstrap aggregating is a method that involves manipulating the training set by resampling

We learn ***k*** base classifiers on ***k*** different samples of training data

Unlike Boosting, Bagging uses uniform weights (e.g. a uniform sampling distribution)

Demo: Bagging Classifier in Scikit Learn

Independent Practice

Conclusion

Q & A

Review

EXIT TICKET

DON’T FORGET TO FILL OUT YOUR EXIT TICKET