GENERAL ASSEMBLY

Intro to Ensemble & Bagging

Sam Stack

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

Ensemble

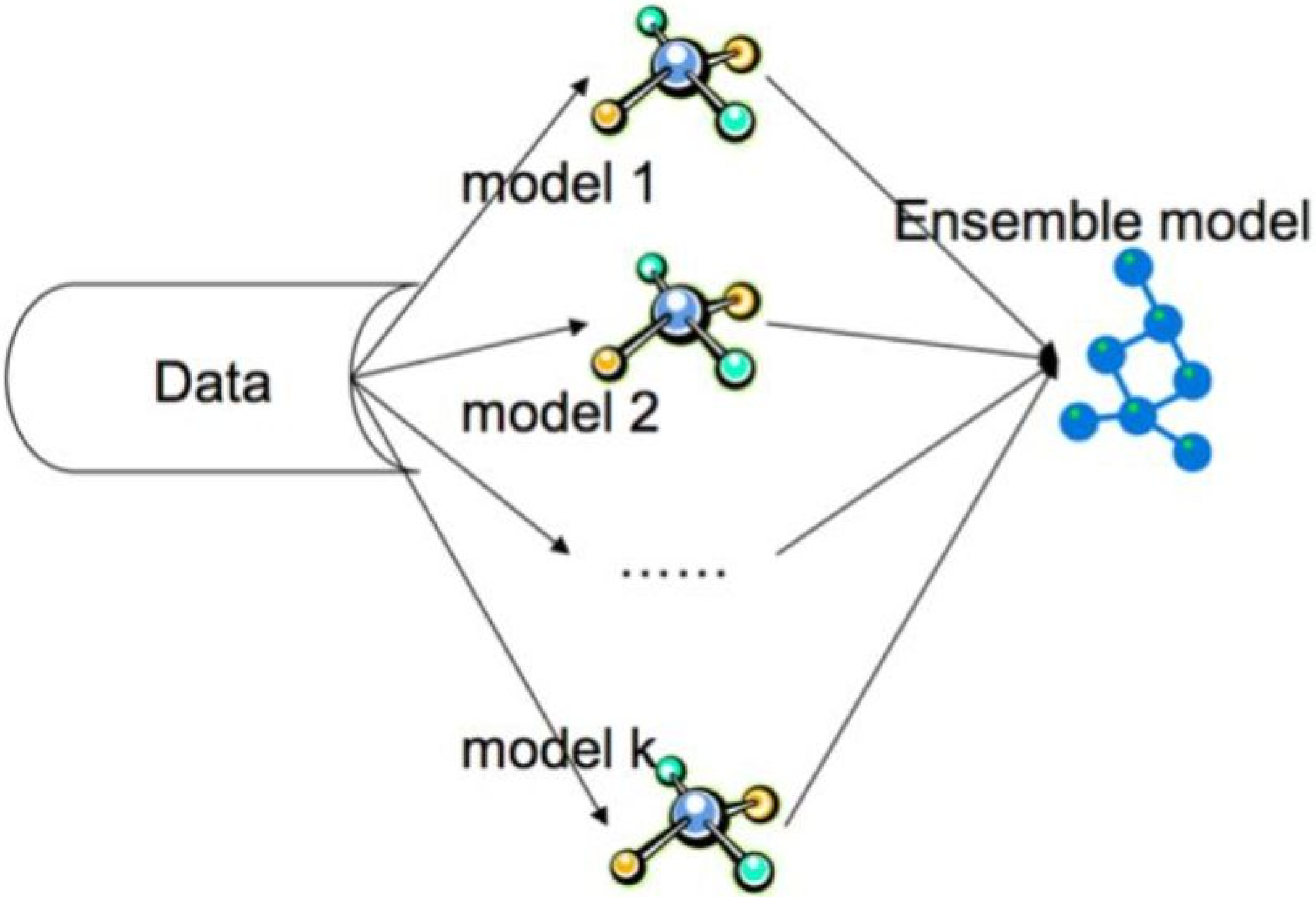
What is an Ensemble?

Intro to Ensemble & Bagging

Ensembles

- Ensemble techniques are supervised learning methods to improve predictive accuracy by combining several base models in order to enlarge the space of possible hypothesis to represent our data. Ensembles are often much more accurate than the base classifiers that compose them.

Intro to Ensemble & Bagging



Ensembles

- Ensemble techniques are supervised learning methods to improve predictive accuracy by combining several base models in order to enlarge the space of possible hypothesis to represent our data. Ensembles are often much more accurate than the base classifiers that compose them.

When might this be useful?

Intro to Ensemble & Bagging

Averaging Methods:

- Build several independent estimators and then average their predictors. This act of combining estimators using results in better models because it reduces variance. (This tactic sounds familiar)
  - Bagging
  - Random Forests

Intro to Ensemble & Bagging

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.)*
  - AdaBoost
  - Gradient Tree Boosting

Intro to Ensemble & Bagging

The Hypothesis Space

In any supervised learning task, our goal is to make predictions of the true classification function f by learning the classifier $\boldsymbol{h}$. 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.

The Hypothesis Space

There are a few reasons that could be prevent our hypothesis from reaching a perfect score?

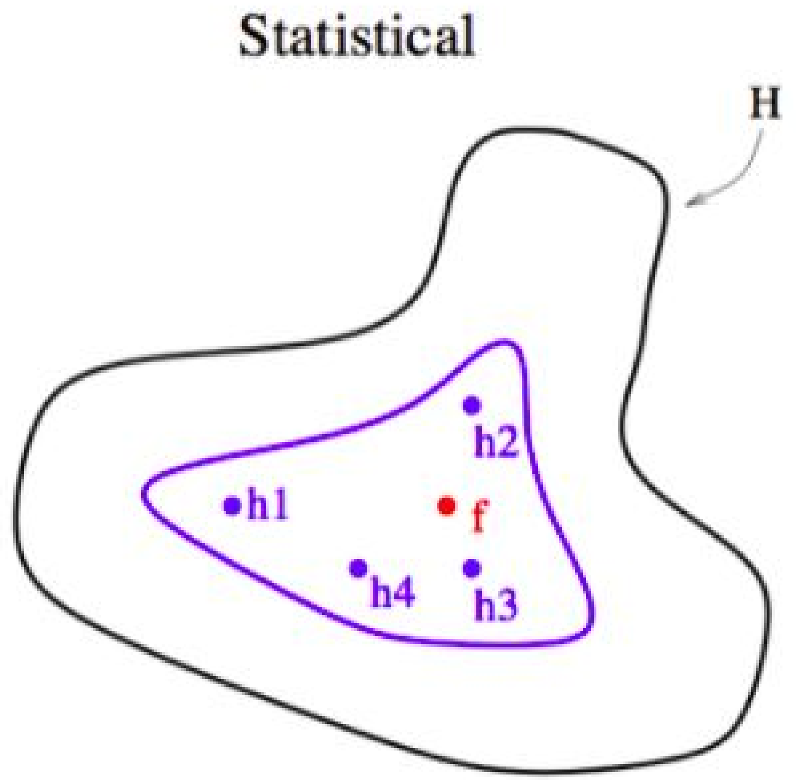
These are the three types:

- Statistical
- Computational
- Representational

The Statistical Problem

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

An ensemble classifier can mitigate this problem by "averaging out" base classifier predictions to improve convergence. This can be pictorially represented as a search in a space where multiple partial perspectives are combined to obtain a better picture of the goal.



- The true function f is best approximated as an average of the base classifiers.

The Computation Problem

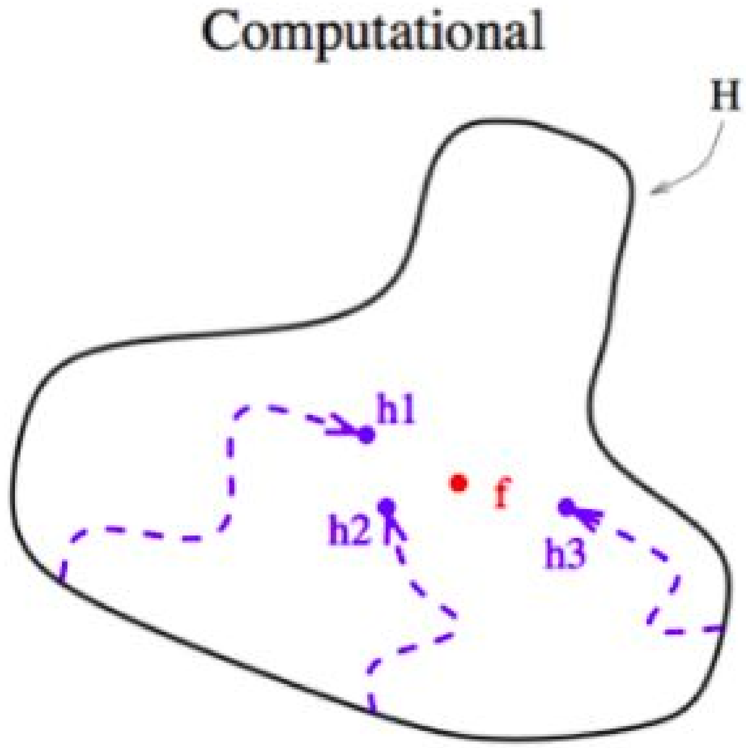
Even with sufficient training data, it may still be computationally difficult to find the best classifier ***h***.

For example, if our base classifier is a decision tree, an exhaustive search of the hypothesis space of all possible classifiers is extremely complex (NP-complete).

In fact this is why we used a heuristic algorithm (greedy search).

An ensemble composed of several *Base Classifiers* with different starting points can provide a better approximation to f than any individual *Base Classifiers*.

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



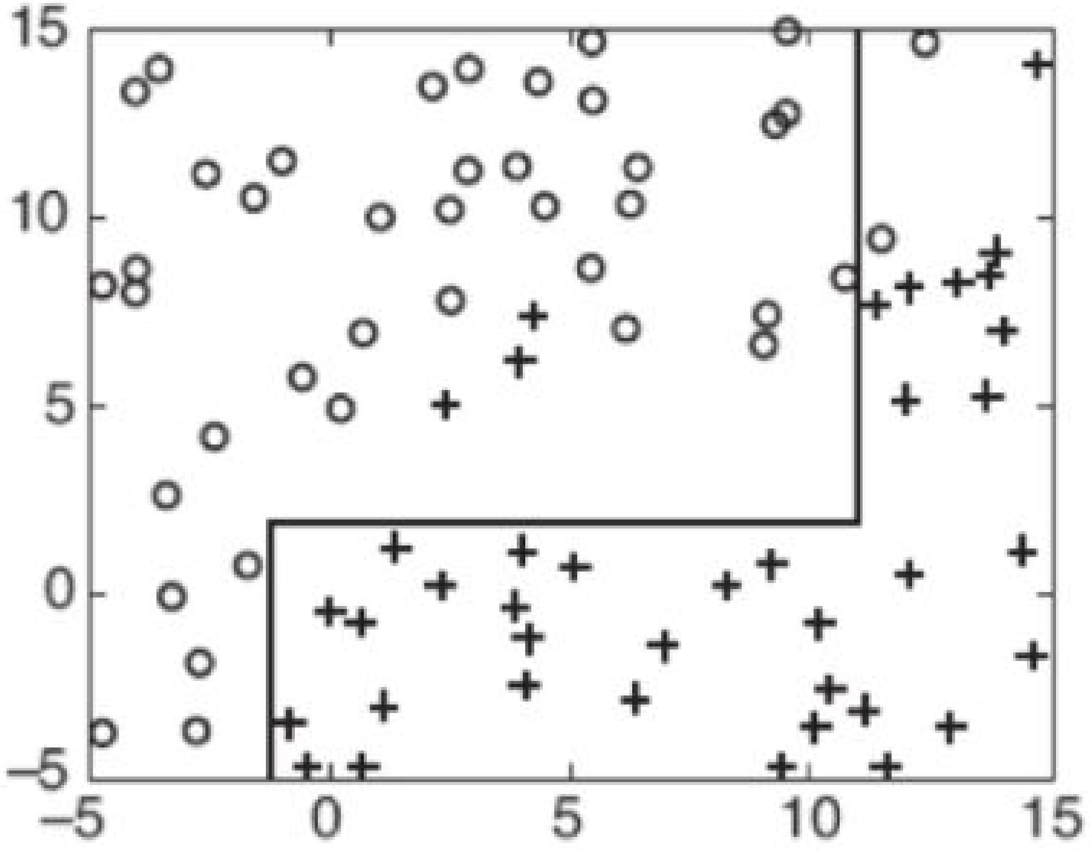
The Representational Problem

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.

But what if f is a diagonal line?

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

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

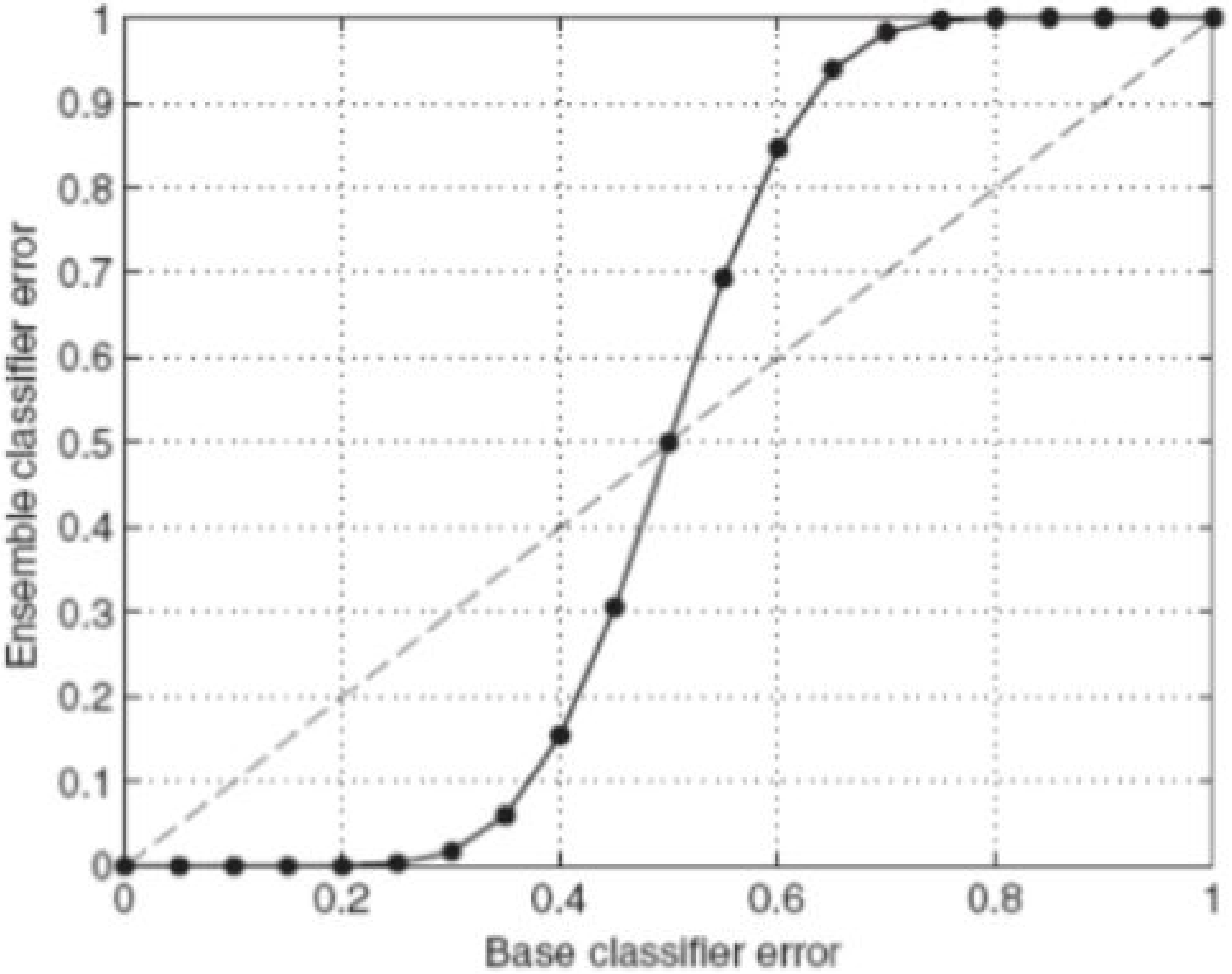


Characteristics of a Using Ensemble

For an Ensemble method to be able to outperform a single baseline classifier, a few conditions must be met.

- **Accuracy:** The baseline classifier must outperform random guessing.
- **Diversity:** Misclassification must occur on different training sets.

Characteristics of a Using Ensemble



Bagging
a.k.a.
Bootstrap Aggregating

Bootstrap Aggregating

- Ensemble method at involves manipulating the training set by resampling
- Samples are independently created with uniform weights.
- Models in the Bagged ensemble are individually trained on these randomly drawn subsets.

Fun Facts about Bagging!

- Reduces Variance
- Most Error from bagging is due to Bias.
- Not very susceptible to overfitting with noisy data.
- Work best with strong and complex models

Fun Facts about Bagging!

- Bagging reduces the variance in our generalization error by aggregating multiple base classifiers together (provided they satisfy our earlier requirements).
- If the base classifier is stable, then the ensemble error is primarily due to bias, and bagging may not be effective.
- Since each sample of training data is equally likely, bagging is not very susceptible to overfitting with noisy data.
- As they provide a way to reduce overfitting, bagging methods work best with strong and complex models (e.g., fully developed decision trees), in contrast with boosting methods which usually work best with weak models (e.g., shallow decision trees).

Bagging Code Along