Stacking

Ensemble learning

Many models

- Fit several predictive models to a data set
- Each model gives its “perspective” on the data
- Now have to get a consensus perspective

A simple solution

- One could take the average or weighted average of the predictions

Stacking

- Instead, use these predictions as **predictors** of the target in a new model
- “Regress” the target variable on these predictions

Stacking

- Fit Random Forest, boosted trees, linear regression, k-nearest neighbors on training set
- Grab the cross-validated predictions on the training set
- Grab the predictions on the test set
- Use these as training and test sets for new model predicting the target
- Maybe add on the original predictors

Stacking

```
base_algorithms = [logistic_regression, decision_tree_classification, ..

stacking_train_dataset = matrix(row_length=len(target), column_length=le
stacking_test_dataset = matrix(row_length=len(test), column_length=len(a


for i,base_algorithm in enumerate(base_algorithms):
    stacking_train_dataset[,i] = base_algorithm.fit(train, target).predi
    stacking_test_dataset[,i] = base_algorithm.predict(test)

final_predictions = combiner_algorithm.fit(stacking_train_dataset, targe
```