AASD 4000
Machine Learning - I

Applied AI Solutions Developer Program

Module 10
Ensemble Learning

Vejey Gandyer

Agenda

- Ensemble
- Voting
- Averaging
- Weighted Averaging
- Stacking
- Blending
- Bagging
- Boosting

Ensemble

What is it?

What is Ensemble?

Combine the decisions from multiple models to improve the overall performance

Voting

Averaging

Weighted Averaging

Stacking

Blending

Bagging

Boosting

Voting

Voting

Used for Classification problems

Multiple models are used to make predictions for each data point

Every prediction = 1 vote

The class label that got the maximum vote is the final prediction

Mode (Maximum occurrence of a class label)

Averaging

Averaging

Used for Regression problems

Multiple models are used to make predictions for each data point

Every prediction = 1 vote

The average of class predictions from models is the final prediction

Mean(Average of class labels from different models)

Weighted Averaging

Weighted Averaging

Used for Regression problems

Multiple models are used to make predictions for each data point

Every model prediction != 1 vote but **weighted vote**

The multiplication of class predictions with its model's weightage is the final prediction

Weighted Average of class labels from different models

Stacking

Stacking

Used for Classification and Regression problems

Layered level approach of models are stacked

Multiple models are used to make predictions in level 0

In the top level 1, a simple Logistic Regression is applied over the predictions of Level 0 models

The stacked model predictions is the final prediction

Stacking

1. Split the data into training and test sets. The training data is further split into K-folds just like K-fold cross-validation.
2. A base model(e.g k-NN) is fitted on the K-1 parts and predictions are made for the Kth part. (out-of-fold prediction)
3. This process is iterated until every fold has been predicted.
4. The base model is then fitted on the whole train data set to calculate its performance on the test set.
5. Repeat the last 3 steps for other base models.(e.g SVM,decision tree, KNN)
6. Predictions from the train set are used as features for the second level model.
7. Second level model (Meta-model) is used to make a prediction on the test set.

Blending

Blending

Used for Classification and Regression problems

Level-0 models (Base-models) fit on the training data and its predictions are compiled

Level-1 models (Meta-model) learns how to best combine the predictions of the base models

Blending

1. The train set is split into training and validation sets.
2. Base Model(s) are fit on the training set.
3. The predictions are made on the validation set and the test set.
4. The validation set and its predictions are used as features to build a meta-model.
5. This model is used to make final predictions on the test and meta-features.

Blending Vs Stacking

Blending: Meta-model is trained on predictions made on a holdout dataset (x_val)

Stacking: Meta-model is trained on out-of-fold predictions made during k-fold cross-validation

Bagging

Bagging

Used for Classification problems

Several decision trees are built with

- random subsets of data for building decision trees
- random subsets of features used as splitting criteria for decision trees

Several weak learners' predictions are averaged into a single final prediction

Example: RandomForest

Boosting

Boosting

Used for Classification problems

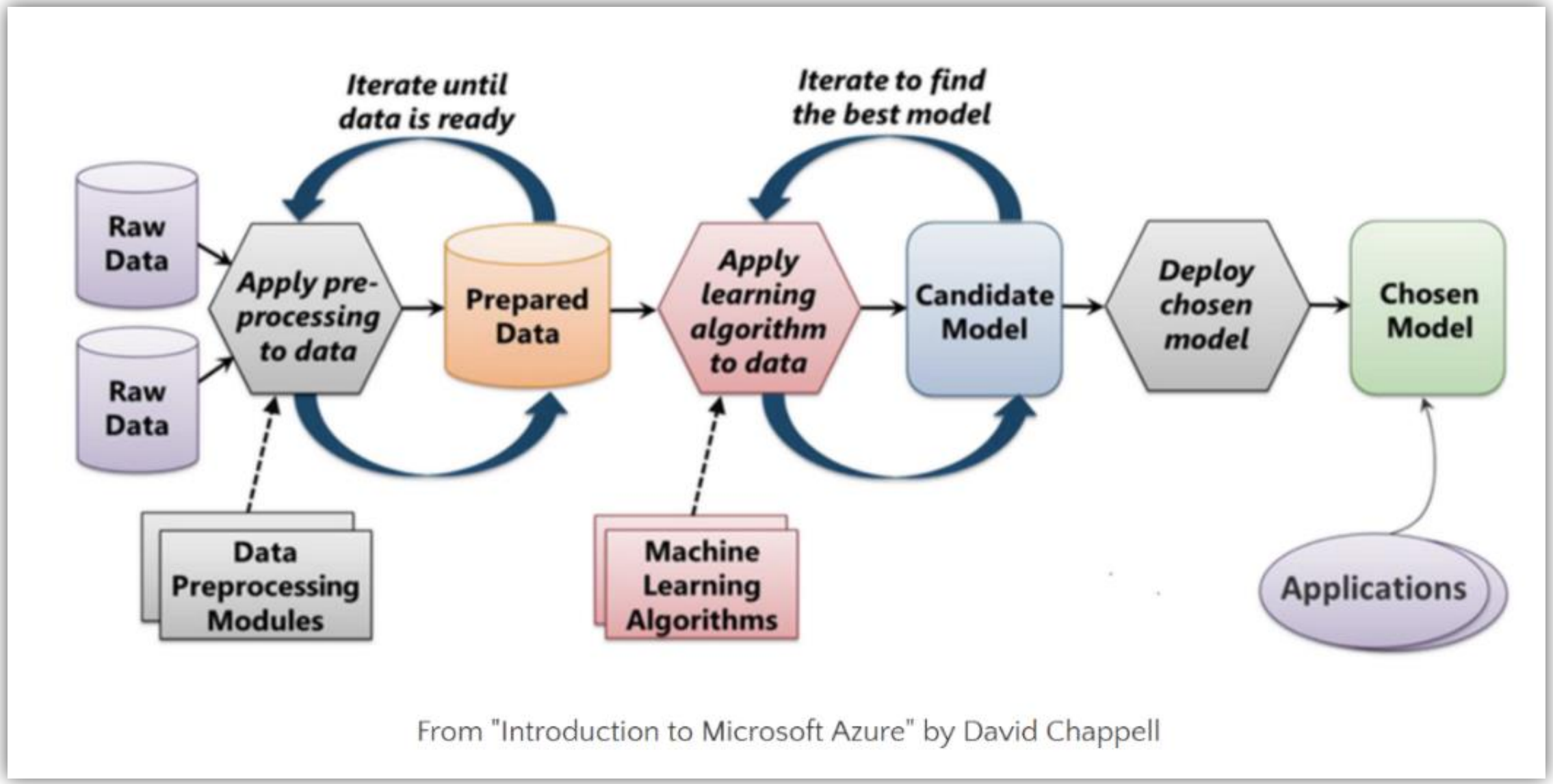
Prediction error in the previously built model is corrected in subsequent model-building exercise

At every iteration, errors from previous step are considered and used to learn new patterns at subsequent steps

After several iterations, a robust model with minimal errors is built

Example: XGBoost, LightGBM

Machine Learning Process



From "Introduction to Microsoft Azure" by David Chappell

Further Reading

Scikit-learn documentation

https://scikit-learn.org/stable/modules/ensemble.html

Python Data Science Handbook by *Jake Vanderplas*