NOVA IMS
Information Management School

7

ENSEMBLES OF METHODS

Machine Learning for Marketing

Acreditações e Certificações

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa

Summary

1. Introduction
2. Ensemble techniques
3. Considerations on the application
4. Application exercise

NOVA IMS Information Management School

7.1

Introduction

Ensembles of methods

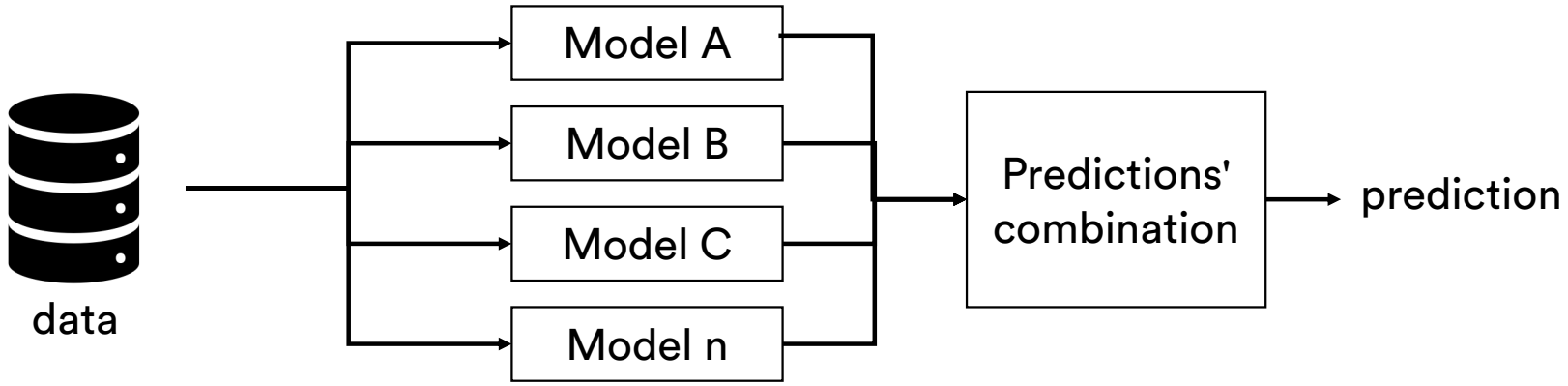
NOVA IMS Information Management School

Problems that can occur from using single models

- **High variance**: the model works very well with the training data, but badly when there are changes to the inputs
- **Low accuracy**: the model does not capture the entire training data patterns, as such does not achieve the required results
- **Features' noise and bias**: the measure relies heavily on specific features to make good predictions

Ensemble learning

Combination the decisions of multiple models to improve overall performance



NOVA IMS Information Management School

7.2

Ensemble techniques

Ensembles of methods

NOVA IMS Information Management School

Bagging

- Each model learns from the previous model, which used a different version of the dataset
- Reduces variance and minimizes overfitting
- Each subset has the same number of observations, allowing the models to be trained in parallel

NOVA IMS Information Management School

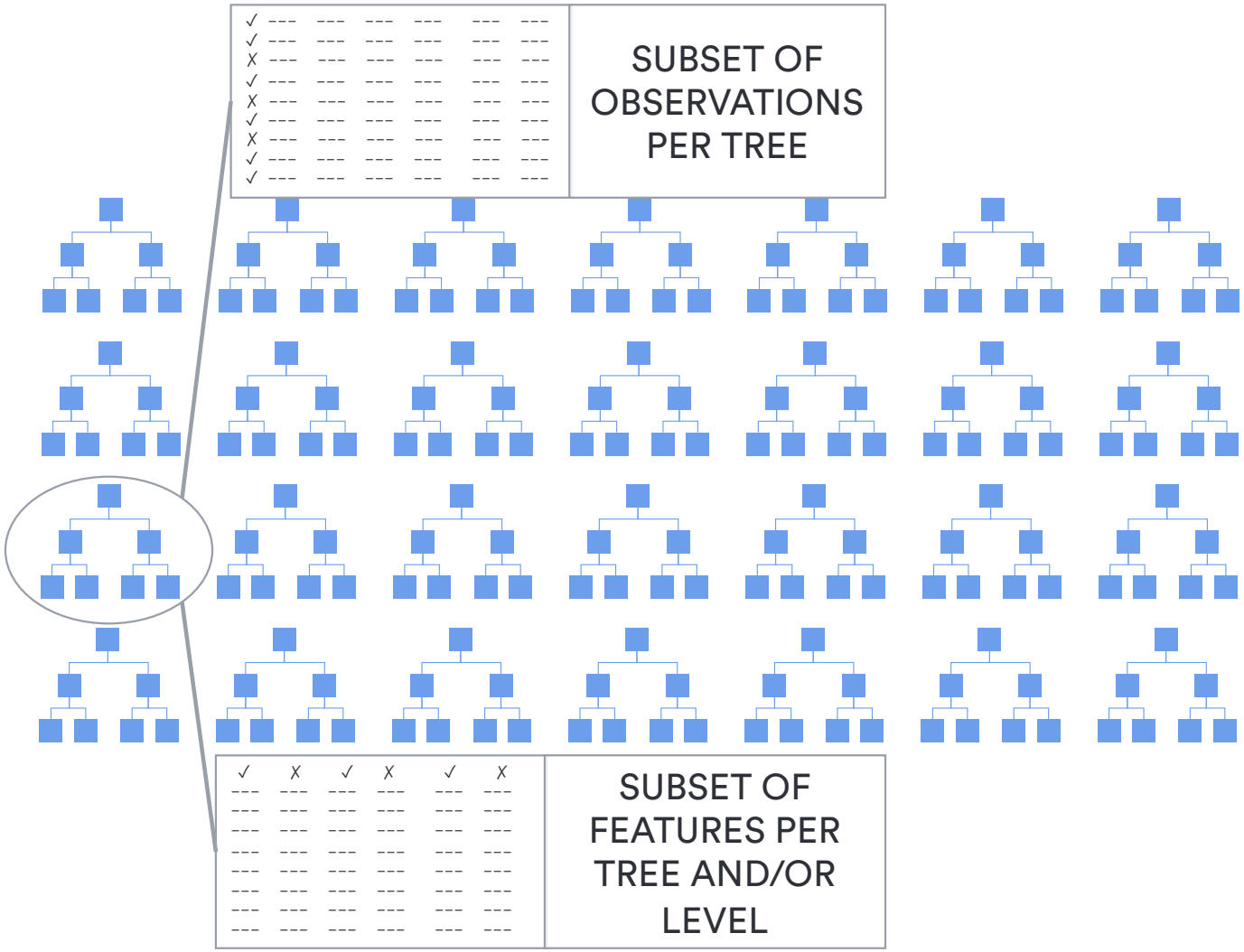
Bagging techniques

- Bootstrap aggregation:
  - Creates subsets of data with replacement
- Random forest:
  - Each model besides using only a subset of samples, only uses a subsample of features
  - Observations and features are typically selected randomly
- Extra-trees:
  - Similar to Random Forest, but uses all observations

Bootstrap aggregation (bagging)

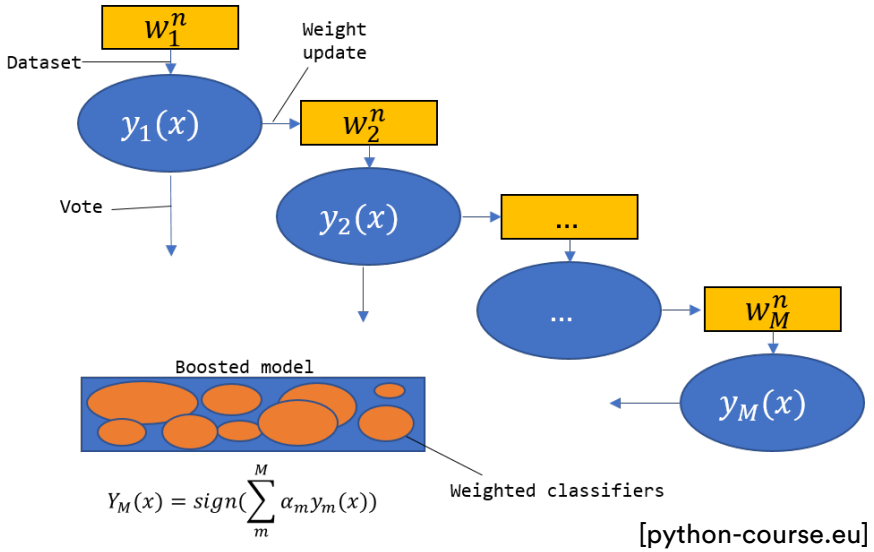
Example algorithms:

- XGBoost
- LightGBM
- CATBoost



Boosting

1. It starts by giving a weight of 1 to all observations and generates a first hypothesis
2. Incorrectly classified observations are given a greater weight. Correctly classified observations are given a smaller weight
3. Generates new hypothesis
4. Repeat 2 and 3 to generate K hypotheses
5. The end result is the result of the combined majority of the various hypotheses generated



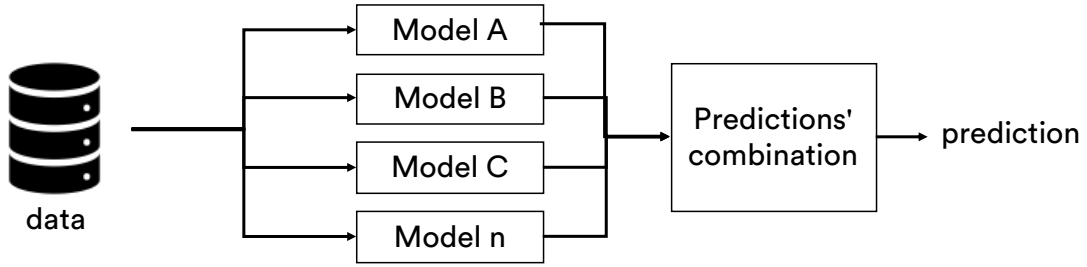
[python-course.eu]

Boosting techniques

- Adaptive boosting:
  - Adjust the model parameters to the training data based on the performance of the current iteration
  - Both the weights for re-weighting and the final aggregation weights are recomputed iteratively
  - Minimizes the exponential loss function, which makes the algorithm more sensitive to outliers
- Gradient boosting:
  - Combination of *gradient descent* and boosting
  - By using a *gradient* it means it uses two or more derivatives of the same function
  - It has three components: additive model, loss function, and a *weak learner*

Stacking

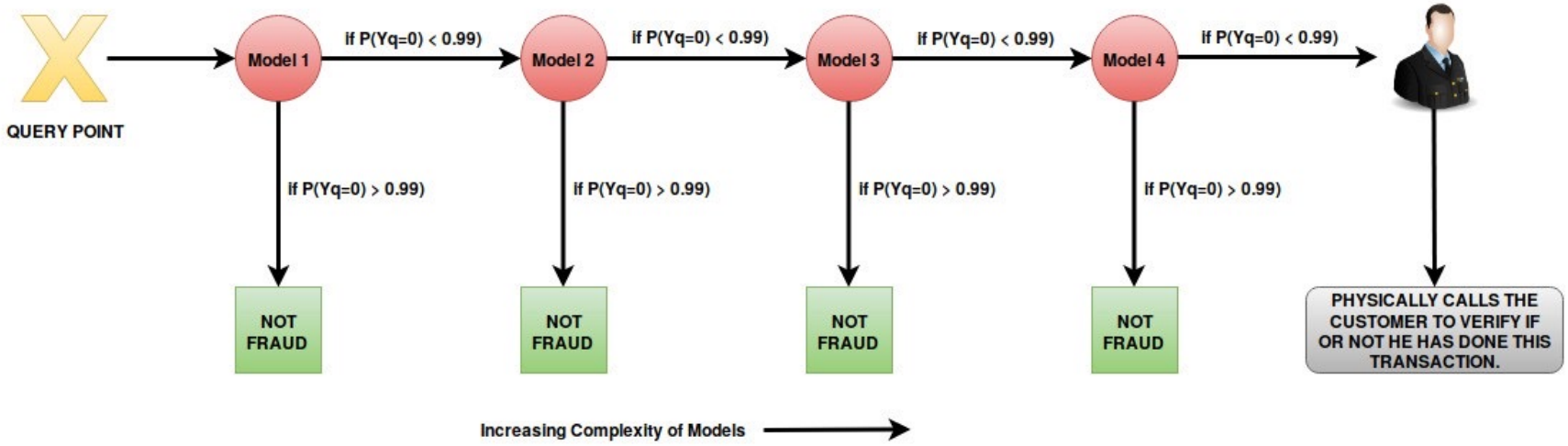
- Each model makes a prediction for each of the observation
- The results of each model are aggregated



- Aggregation methods:
  - Majority voting: used in classification problems. Based on what the majority of models predicted for each observation
  - Averaging: used in regression problems. Based on the average of the predictions for each observation
  - Weighted average: similar to Averaging, but with different weights given to different algorithms/models

Cascading

- Based on the concatenation of several models
- Mostly used when predictions need to ”absolutely” true
- Output of a model is the input in other model of high complexity



DIFFERENT STAGES OF QUERING CASCADE CLASSIFIERS IN A FOUR MODEL CASCADE SYSTEM

[medium.com]

NOVA IMS Information Management School

7.3

Considerations on the application

Ensembles of methods

NOVA IMS Information Management School

Considerations on ensemble of models

- **Simplicity and interpretability**: if simplicity and interpretability are a requirement, then using an ensemble of models may not be a good choice
- **Generalization**: if training parameters are not defined carefully, ensemble models can start working bad with unseen data (overfitting)
- **Inference time**: in many deployment situations, inference time is crucial. The more models are used, the the more time it will need to make the inference
- **Noise, bias, and variance**: if the use of ensemble models does not improve the performance of weak learners, its use should be questioned

NOVA IMS Information Management School

7.4

Application exercise

Ensembles of methods

Ensemble learning in Python with Scikit

- **Random Forest**: algorithm that builds multiple decision trees, each based on a dataset sample with replacement
- **Ada Boost**: algorithm that from small decision trees, with small variations, makes predictions based on majority vote (or average)
- **Gradient Tree Boosting**: powerful boosting algorithm to create classification and regression Models
- **Voting Classifier:** allows the use of various algorithms to make predictions and select the result based on most votes or average odds
- More info at https://scikit-learn.org/stable/modules/ensemble.html

Predicting customers who will leave the bank in the following 6 months

1. Copy from the datasets folder the dataset “Bank_Churn_Modelling.csv”
2. Copy and open the Jupyter notebook “PredictBankChurn_DF.ipynb”
3. Follow the presentation of the notebook, answer the questions and explore the challenges

Questions?

Machine Learning for Marketing

Acreditações e Certificações

Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa