Classical Machine Learning: Classification and Regression (III)

- (Continue) Learn the concept, theory, toy example, and scikit-learn usage of ensemble classifiers (rationale, parallel ensembles: bagging, random forest, and extra trees).
- Learn the concept, theory, toy example, and scikit-learn usage of ensemble classifiers (sequential ensembles: AdaBoost and Gradient Boost).

Review

“Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don’t think AI will transform in the next several years.”

— Andrew Ng

Game Changers

2025

Tech that could change the world

CBINSIGHTS

1 AI weather prediction

Models that can accurately predict extreme weather at the local and global scale

2 Ultra-deep drilling

Advanced drilling techniques that can go far deeper to unlock superhot rock energy

3 AI agent marketplaces

Enabling dynamic collaboration of specialized agents across software platforms

4 Advanced nuclear propulsion

Nuclear power approaches to open the door to deep space exploration

5 Biocomputing

Combining human neurons with chips to unlock the efficiency of the human brain for computers

6 Brain manipulation tech

AI brings forward personalized "brain pacemakers" and other devices to treat neurological disorders

7 Quantum-optimized portfolios

Using quantum computing to build higher-performing portfolios, faster

8 Cellular & epigenetic reprogramming

Altering the gene expression of cells to extend the healthy human lifespan

9 GPS-less navigation systems

Approaches that boost the resiliency of positioning services critical to global infrastructure

Game Changer 2025

Insurance

AI weather prediction

CB INSIGHTS

Game Changer 2025

Extreme weather is the #1 long-term global risk in the next 10 years

Global risks ranked by severity (negative impact on the world) over upcoming 10-year period*

Risk categories

- Environmental
- Societal
- Technological

1st	Extreme weather events
2nd	Critical change to Earth systems
3rd	Biodiversity loss and ecosystem collapse
4th	Natural resource shortages
5th	Misinformation and disinformation
6th	Adverse outcomes of AI technologies
7th	Involuntary migration
8th	Cyber insecurity
9th	Societal polarization
10th	Pollution

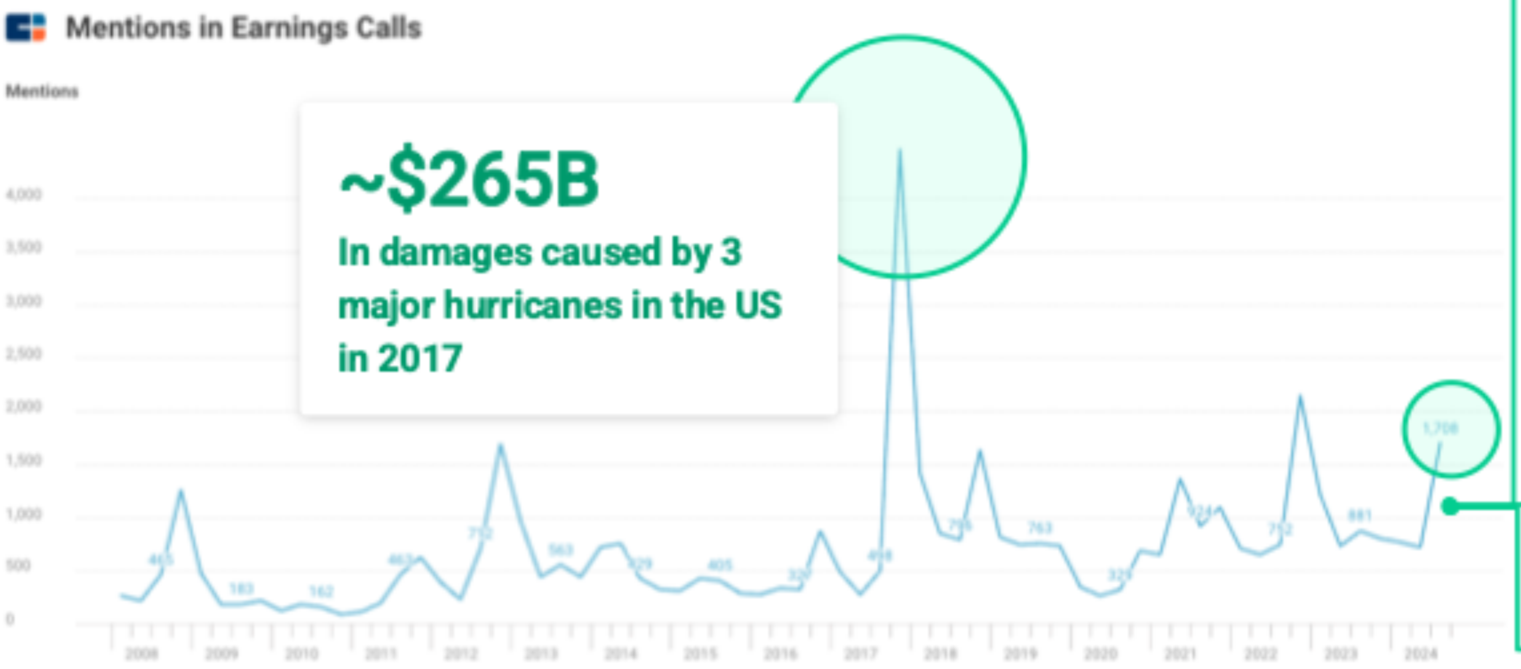
Source: Global Risks Report 2024, World Economic Forum

*Based on the Global Risks Perception Survey (GRPS) of 1,490 experts. "Global risk" is defined as the possibility of an event occurring that would negatively impact a significant portion of global GDP, population, or natural resources.

Game Changer 2025

Managing weather risk could protect $30T worth of global GDP impacted by weather

Extreme weather is top-of-mind for execs on earnings calls



"**Our insurance business was impacted by the severe weather events** in the Greater Toronto Area and the wildfires in Alberta in Q3 and by hailstorms in Calgary and floods in Montreal this month."

TD Bank, Q3'24

"But again, a bit of caution. We're well aware that we have a large part of the year still ahead of us. And especially in respect of natural catastrophes, **we're entering the hurricane season** as we speak."

Swiss Re, Q3'24

"The other issues were **supply chain challenges due to extreme weather patterns**."

Tiger Brands, Q2'24

Source: CB Insights earnings transcripts search — mentions of extreme weather events; CNBC
*Quarter reflects date call occurred.

Game Changer 2025

Deep learning is making it easier to accurately predict extreme weather events at the local and global scale

AI weather and climate models are rapidly improving, enabling faster and more accurate predictions at lower computation costs than traditional physics-based models, which rely on supercomputers to crunch through equations to simulate future conditions. New models use historical weather data to learn from weather patterns and generate future projections in minutes once trained.

February 2022	November 2023	May 2024	July 2024	August 2024
1	2	3	4	5
FourCastNet	**GraphCast**	**Aurora**	**NeuralGCM**	**StormCast**
Nvidia	Google DeepMind	Microsoft Research	Google Research	Nvidia Research
Physics-ML model that generates 7-day forecasts in <2 secs	AI model for 10-day forecasts at high accuracy in <1 min	1.3B-parameter foundation model of the atmosphere	Combines AI and physics models to project climate conditions into multiple decades	AI model that simulates extreme weather events at the km-scale

Source: Company releases
Note: Select models visualized.

Game Changer 2025

Outside of big tech, startups tackling better weather forecasting with AI are targeting industries like insurance, financial services, and defense

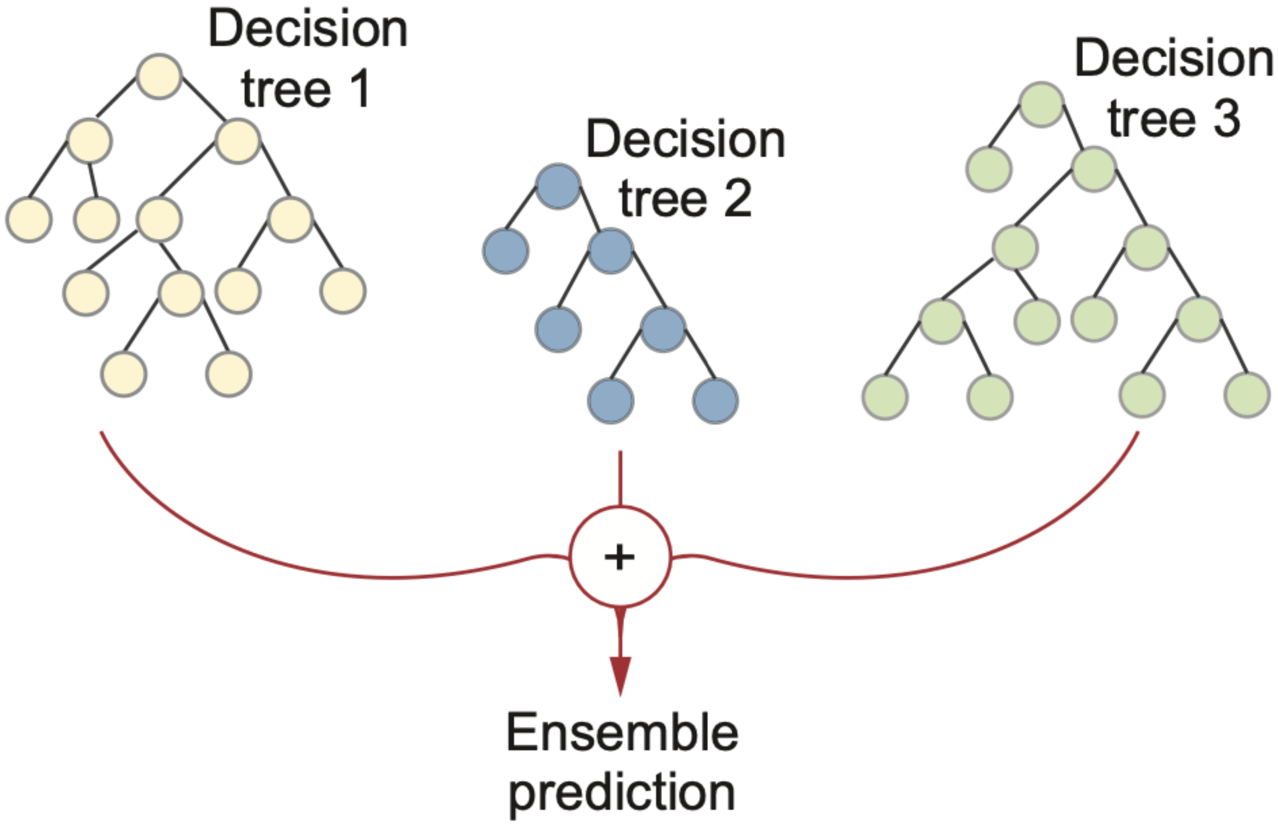
Startup	Description	Industries	Mosaic score*	Commercial Maturity**	Total funding	Select partners
Jua	AI model for 16-day weather forecasts	Energy trading	767	3 (Deploying)	$18M	Volue
Atmo	AI-powered precision weather forecasting	Government & defense, Aerospace, Energy, Agriculture	424	3 (Deploying)		US Air Force, Philippines government
WindBorne Systems	AI model WeatherMesh; long-duration balloons for atmospheric data collection	Government & defense	N/A	3 (Deploying)	$32M	NOAA, Department of Defense, US Air Force
reask	AI-powered natural hazard modeling	Insurance, Financial services, Government	595	3 (Deploying)	$4.6M	AXA Climate, Juniper Re

Source: CB Insights company profiles — Jua, Atmo, WindBorne Systems, Reask. Note: Select companies visualized. *Mosaic score (out of 1,000) measures a company's health (as of 9/24/2024). **Commercial Maturity (out of 5) measures a company's ability to acquire customers today.

Parallel Ensembles

Review

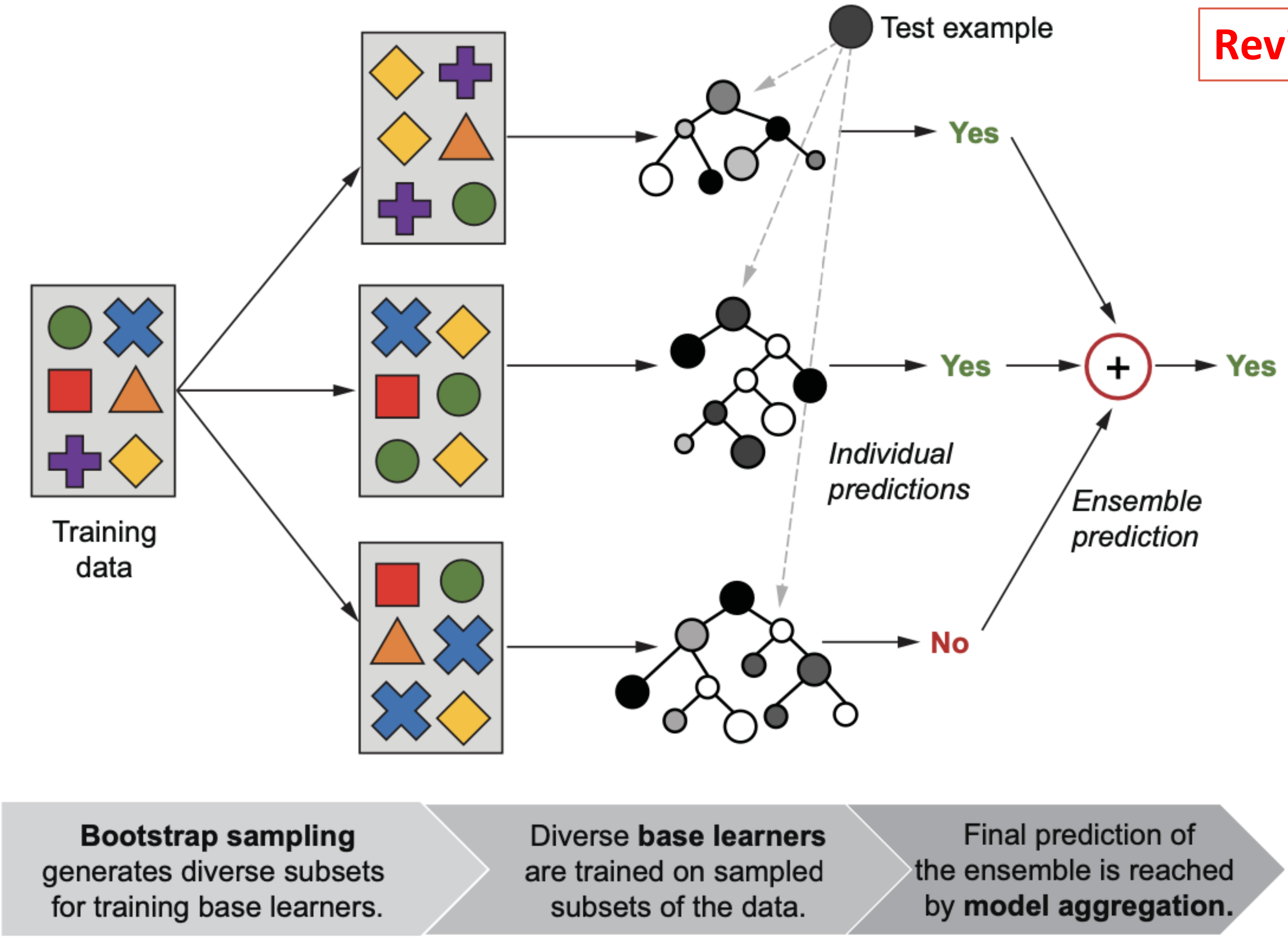
- Trained using the same base machine-learning algorithm.
- Ensemble diversity is created from a single algorithm with random data or feature sampling to train each base model.
- **Ensembles in this family**: bagging, random forest, extra trees etc.



G. Kunapuli (2023) Ensemble Methods for Machine Learning, Manning.

Bagging Illustration

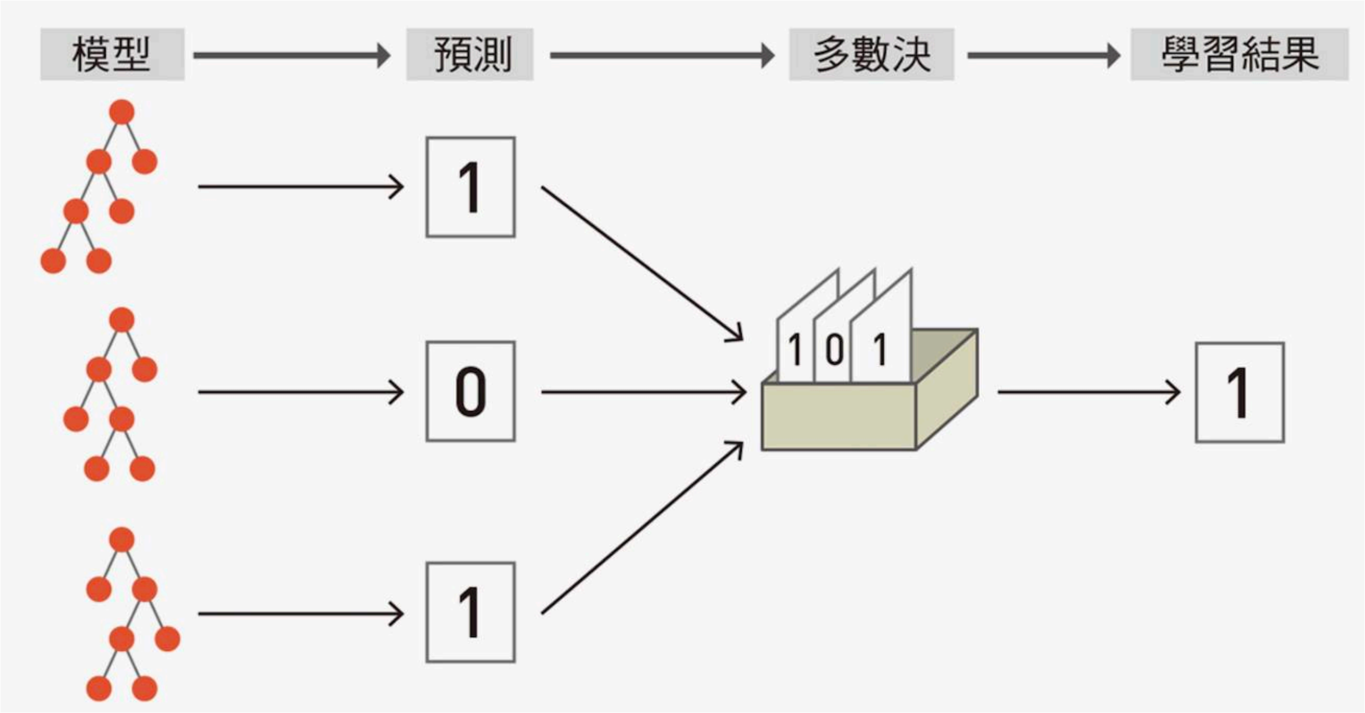
Review



G. Kunapuli (2023) Ensemble Methods for Machine Learning, Manning.

Random Forest Algorithm

- Construct an ensemble of decision trees by manipulating training set as well as features
  – Use bootstrap sample to train every decision tree (similar to Bagging)
  – Use the following tree induction algorithm:
    - ◆ At every internal node of the decision tree, randomly sample p attributes (p < d) for selecting split criterion

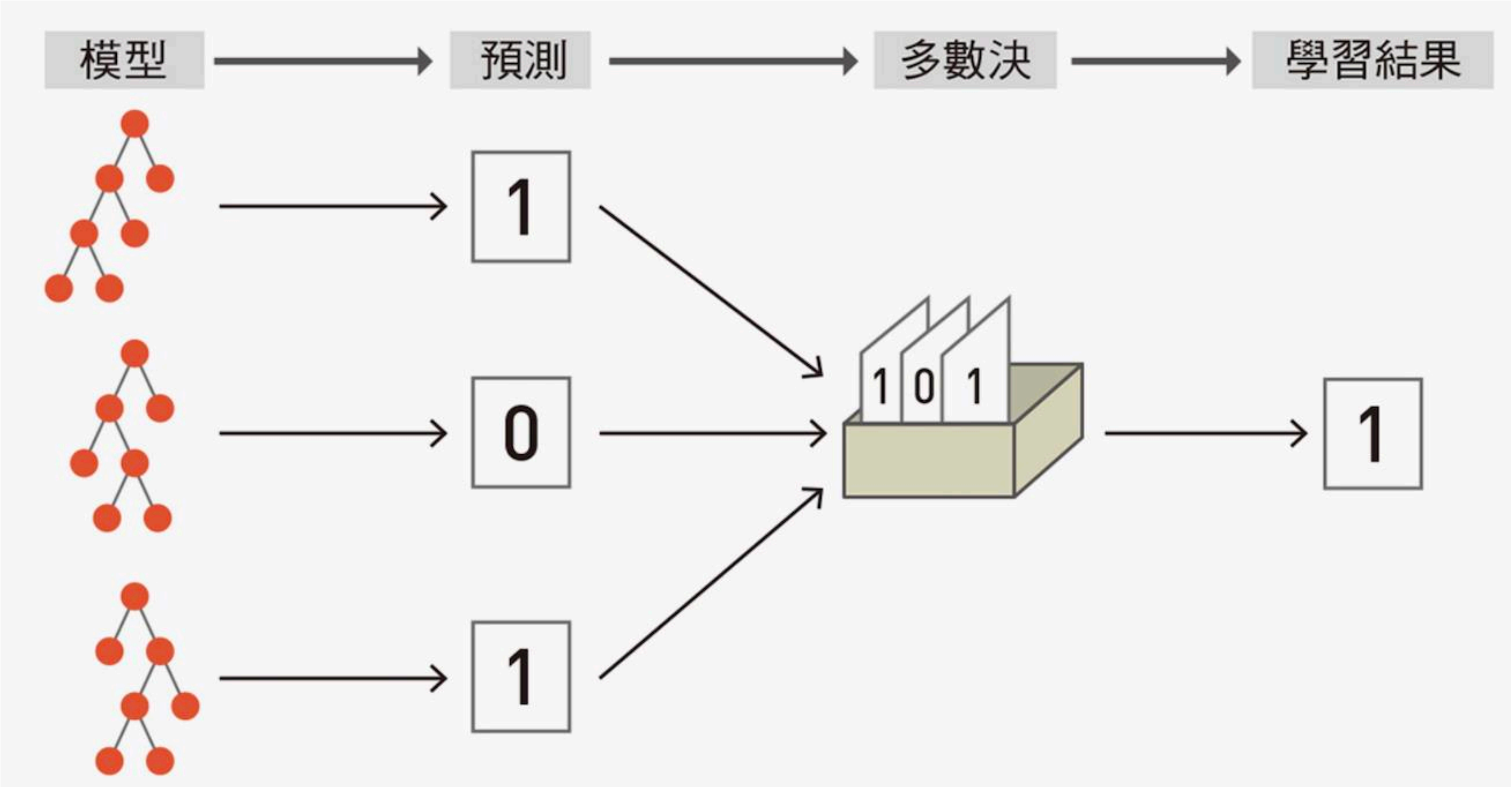


Random Forest: theoretical minimum and python example

- The phrase “theoretical minimum” is taken from a successful book series by Leonard Susskind, a great physicist at Stanford University.
- “Theoretical minimum” means just the minimum theories and equations you need to know to proceed to the next level.
- See Ensemble_RF_ET.pdf

Feature Importance: Extra Bonus of Random Forest

- Random forest measures a feature's importance by looking at how much the tree nodes that use that feature to reduce impurity on average (across all trees in the forest).
- The feature that can reduce more impurity, the more important.



Feature Importance: Extra Bonus of Random Forest

Automation in Construction 118 (2020) 103274

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

ELSEVIER

Machine learning-based seismic capability evaluation for school buildings

Nai-Wen Chi[a], Jyun-Ping Wang[b], Jia-Hsing Liao[c], Wei-Choung Cheng[d], Chuin-Shan Chen[b,*]

Check for updates

Fun Time: what is the most important feature of seismic capability for old school buildings in Taiwan?

1. Total floor area of the building
2. Spectral acceleration demand
3. Tensile strength of steel
4. Amount of walls in Y direction
5. The built year

Summary: Ensemble Rationale, Bagging, Random Forest and Extra Trees

- For the ensemble classifiers to outperform the base classifiers, two conditions must be met:
  - The base classifier should do better than random guessing. (This is easy in general)
  - The base classifiers should be independent of each other. (This is hard!)
- Three well-known **parallel** ensemble methods are Bagging, Random Forest, and Extra Trees.
- **Bagging** creates different subsets of data (this is called bootstrapping), trains one model per subset, and aggregates all predictions to get the final prediction.

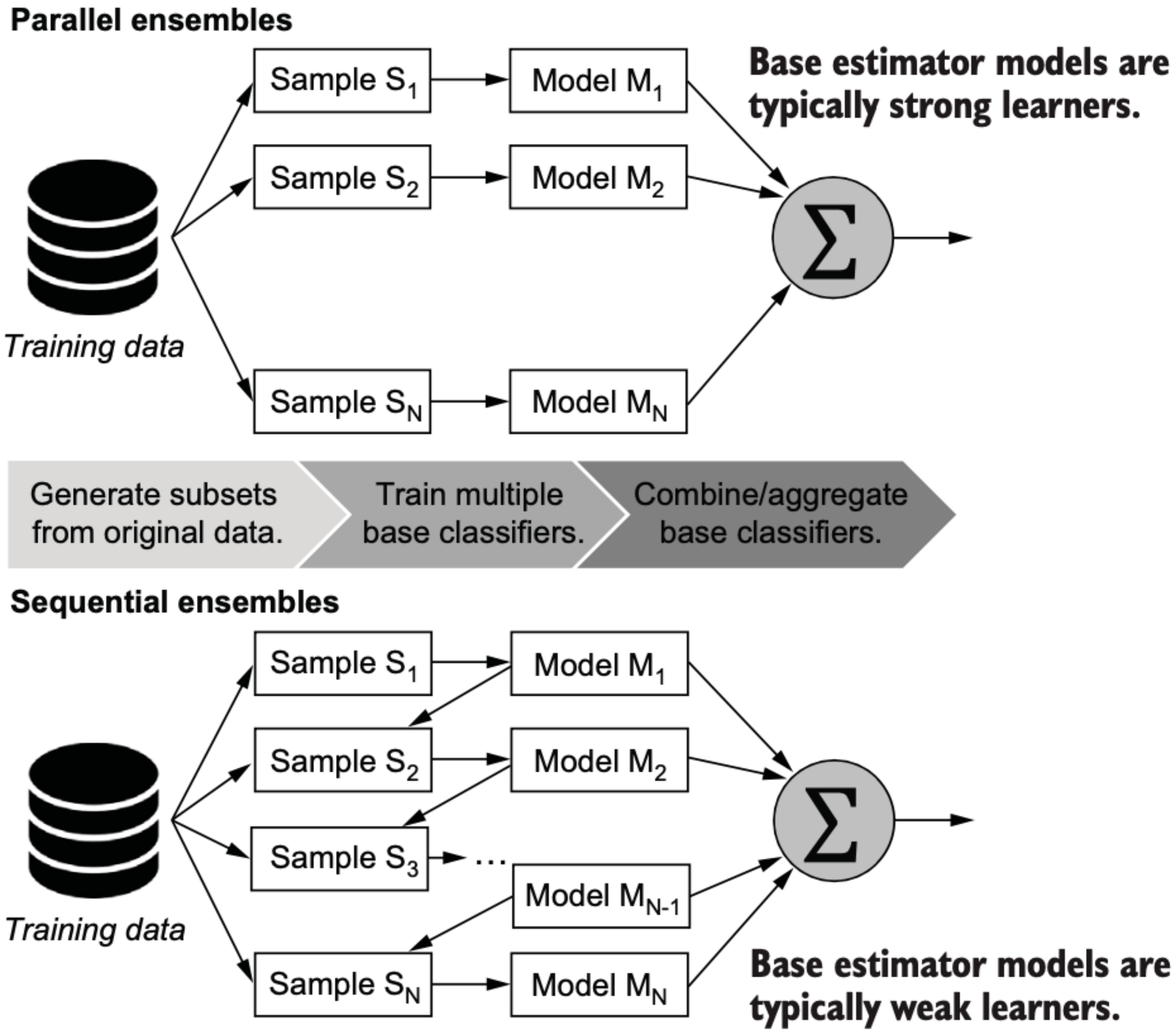
Summary: Ensemble Rationale, Bagging, Random Forest and Extra Trees

- **Random Forest** is similar to Bagging. Random Forest differs from Bagging by further randomly choosing candidate features to decide a node's split criteria.
- One benefit of using Random Forest is that it provides a natural mechanism for scoring features based on their importance.
- **Extra Trees** is similar to Random Forest, which randomly chooses candidate features. Extra Trees differ from Random Forest by further randomly deciding the split threshold.

Classification Algorithm Walkthrough: Ensemble Classifiers – Sequential Ensembles: AdaBoost,

C-S David Chen, Department of Civil Engineering, National Taiwan University

Parallel vs. Sequential Ensembles

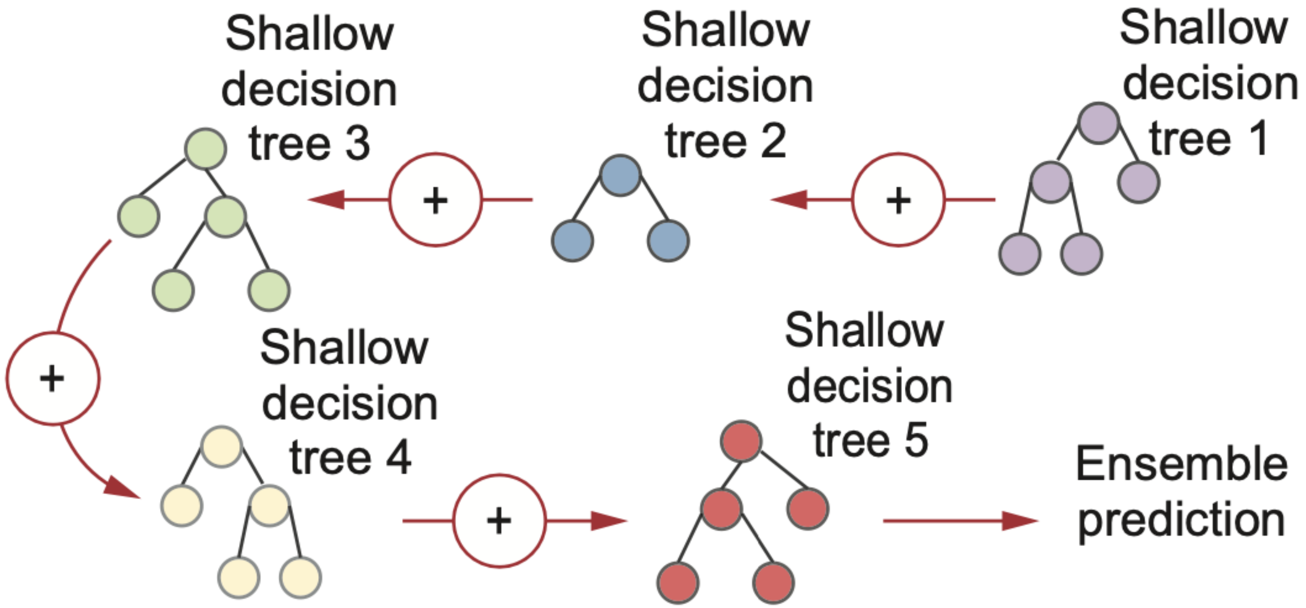


Differences between parallel and sequential ensembles

1. base estimators in parallel ensembles are trained independently of each other, while in sequential ensembles, they are trained to improve on the predictions of the previous base estimator
2. sequential ensembles typically use weak learners as base estimators.

Sequential Adaptive Boosting Ensembles

- Trained in a stage-wise, sequential manner.
- Each successive model is trained to fix the mistakes made by the previously trained model, allowing the ensemble to adapt during training.
- The predictions of a large number of weak models are boosted into a strong model!
- **Ensembles in this family**: AdaBoost, LogitBoost



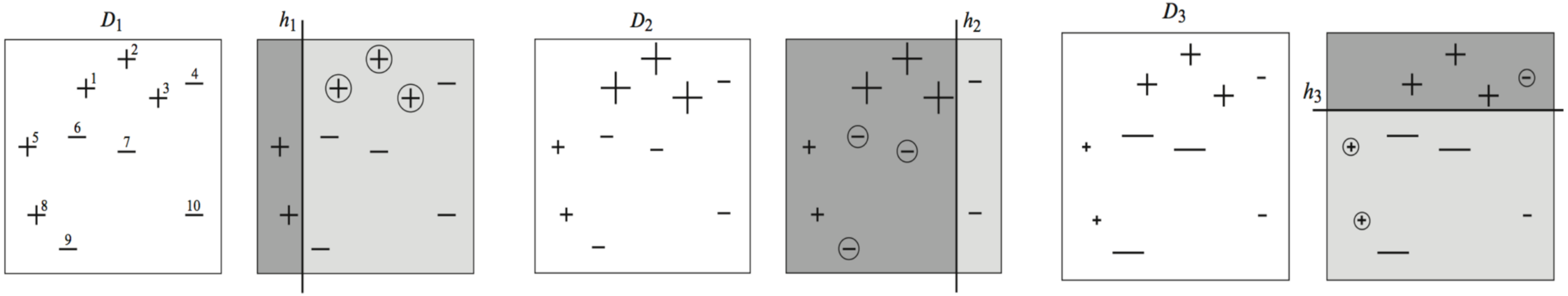
AdaBoost: theoretical minimum and example

- The phrase “theoretical minimum” is taken from a successful book series by Leonard Susskind, a great physicist at Stanford University.

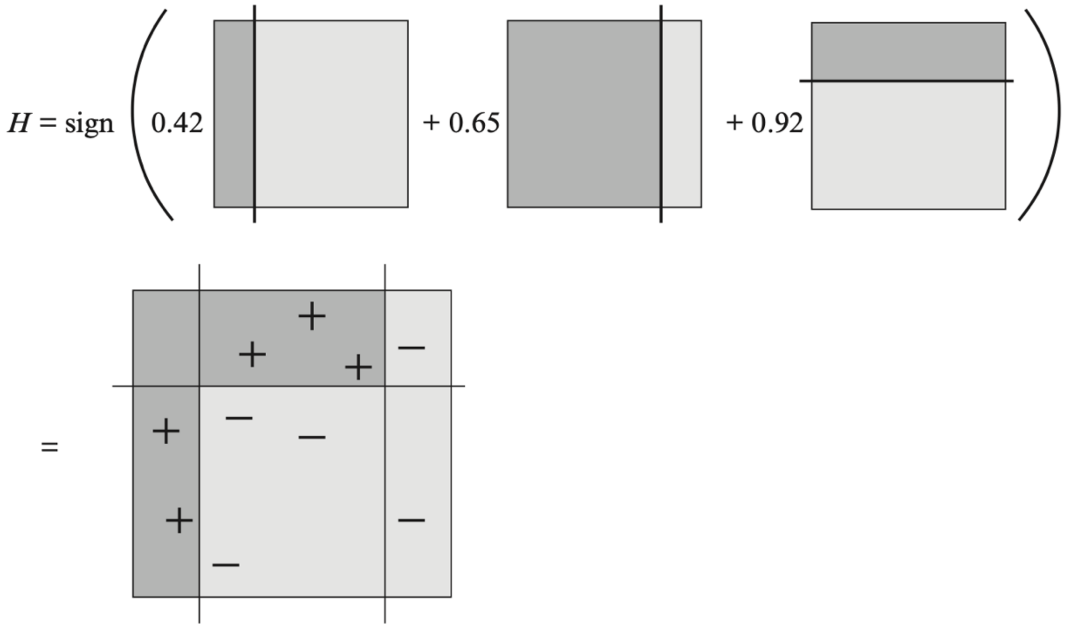
- “Theoretical minimum” means just the minimum theories and equations you need to know to proceed to the next level.

- See Ensemble_AdaBoost.pdf

AdaBoost: Summary



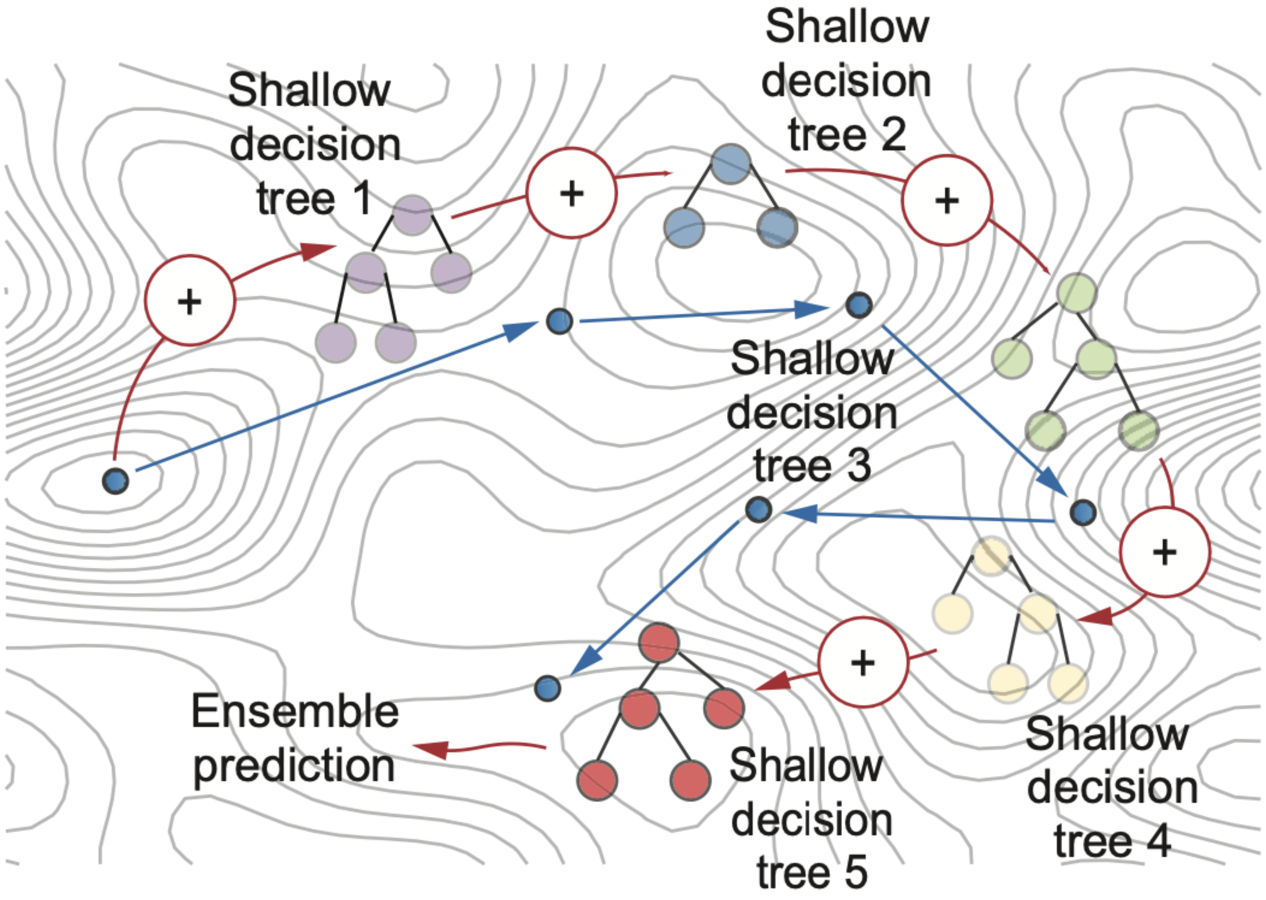
- Fit an additive model (ensemble) $\sum_t \alpha_t h_t(\mathbf{x})$ in a forward stage-wise manner.
- In each stage, introduce a weak learner to compensate the shortcomings of existing weak learners.
- In Adaboost, "shortcomings" are identified by high-weight data points.



credit: Cheng Li, A Gentle Introduction to Gradient Boosting.

Sequential Gradient Boosting Ensembles

- Trained stage-wise to emulate gradient descent over the task-specific loss function.
- Each successive model is trained to fit the residuals, or example-wise losses, of the previously trained model.
- Thus, each ensemble component is both an approximate gradient and a weak learner!
- **Ensembles in this family**: gradient boosting and LightGBM, Newton boosting and XGBoost, ordered boosting and CatBoost.



G. Kunapuli (2023) Ensemble Methods for Machine Learning, Manning.

Gradient Boosting: theoretical minimum and example

- The phrase “theoretical minimum” is taken from a successful book series by Leonard Susskind, a great physicist at Stanford University.
- “Theoretical minimum” means <u>just the minimum theories and equations you need to know to proceed to the next level.</u>
- See Ensemble_GradBoost.pdf

Summary: Gradient Boosting = Gradient Descent + Boosting

- Fit an additive model (ensemble) $\sum_t \alpha_t h_t(\mathbf{x})$ in a forward stage-wise manner.
- In each stage, introduce a weak learner to compensate the shortcomings of existing weak learners.
- In Gradient Boosting, "shortcomings" are identified by gradients.
- Recall that, in Adaboost, "shortcomings" are identified by high-weight data points.
- Both high-weight data points and gradients tell us how to improve our model.

credit: Cheng Li, A Gentle Introduction to Gradient Boosting.

Ensemble classifiers

Ensemble_compare_02.ipynb

- Bagging
- Random Forest
- AdaBoost
- XGBoost

Battle Between XGBoost, LightGBM, CatBoost

https://lavanya.ai/2019/06/27/battle-of-the-boosting-algorithms

- XGBoost（eXtreme Gradient Boosting） 特點是計算速度快，模型表現好，可以用於分類和回歸問題中，**號稱「比賽奪冠的必備殺器」**。
- LightGBM（Light Gradient Boosting Machine）的**訓練速度和效率更快**、使用的內存更低、**準確率更高**、並且支持並行化學習與處理大規模數據。
- Catboost（ Categorical Features+Gradient Boosting）採用的策略在降低過擬合的同時保證所有數據集都可用於學習。性能卓越、通用性更好、易於使用而且更實用。

原文網址：https://kknews.cc/tech/vlxj8n2.html