



# ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

---

## Lecture 4a – Decision Trees, Ensembles, and Evaluation



**Chizhi Chris ZHANG**

zhangchizhi@ciomp.ac.cn

Advanced Computing and Digital Technology Research Center

University of Chinese Academy of Sciences

---

Spring 2026

# Today's Question

与数字工程研究中心

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## What we are trying to answer

Why do decision trees feel so understandable, yet still need careful checking before we trust them in real decisions?

## Where this matters

A tree can help route a student case, predict delivery delay, flag fraud, or estimate demand, but every one of those uses carries a different cost when the model is wrong.

## What changes from AI3

AI3 used one fitted rule to estimate a number. AI4 moves to models that keep asking smaller questions and then combine those answers into a decision or an estimate.

# From AI3 to AI4 算与数字工程研究中心

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Last time

We learned to read prediction through residuals, loss, validation, and the difference between useful estimates and false certainty.

## Today

We keep that same evaluation habit, but the model is no longer one smooth formula. It is a sequence of branching questions.

## One simple sentence

Regression says, “fit one rule.” A tree often says, “first decide what kind of case this is.”

# Why Tree Decisions Feel Natural

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## How people already think

Many ordinary judgments already have a tree shape: urgent or not, expensive or not, serious or not, worth calling someone or not.

## A campus example

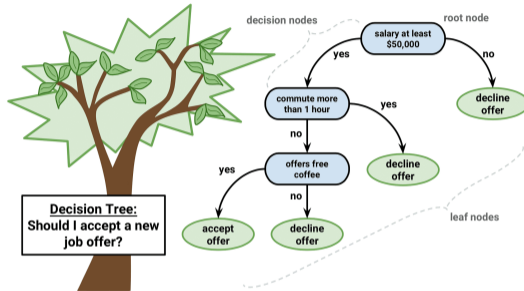
A clinic may first ask whether symptoms are urgent, then ask about fever, then ask how long the symptoms have lasted.

That is why trees often feel easier to read than a single formula. The model sounds like a short decision story.



# A Practical Decision Story

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER



## What this picture gets right

Each branch asks a question that a person can explain immediately.

## What this picture hides

The tree did not appear because the questions were wise. The questions were chosen from data, and data can reward shortcuts as easily as wisdom.

# Why Lines Are Not Always Enough

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Some problems feel threshold-based

If the waiting time is long, respond now. If the payment is large and the location is unusual, review the transaction. If the temperature is high and vibration is rising, stop the machine.

## Why trees feel natural here

These situations often sound more like a sequence of conditions than one smooth formula that covers everyone equally well.

# Why Trees Work Well on Tables

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Why tabular data fits

If each row is a person, an order, a patient, or an event, then a tree can keep asking yes-or-no or smaller-or-larger questions about the columns.

## A concrete example

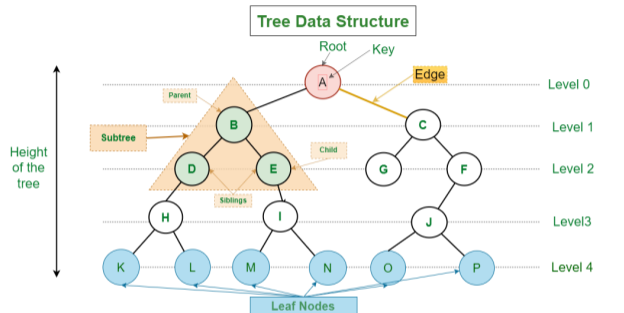
A scholarship model might ask about attendance first, then family income, then extracurricular load, not because these are the only things that matter, but because they reduce uncertainty quickly.

## The quiet limitation

Trees are strongest when the world really can be sliced into useful regions. They are less elegant when the signal is spread smoothly across many variables.

# Tree Anatomy

计算与数字工程研究中心  
ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER



## Read the parts

The root asks the first question, internal nodes ask follow-up questions, and leaves give the final class, score, or estimate.

## Why this vocabulary matters

Once root, split, branch, depth, and leaf are clear, later ideas like pruning, forests, and boosting stop sounding mysterious.

# What a Split Is Trying to Do

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Intuition first

At each node, the tree looks for a question that makes the resulting groups more internally consistent than the original mixed group.

## One useful formula

$$\text{Gini}(D) = 1 - \sum_k p_k^2$$

Lower impurity means the node has become less mixed.

## What matters more than the formula

A good split is simply a good question. After you ask it, the situation should become clearer.



# A Small Split Story

先进计算与数字工程研究中心

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Imagine a library help desk

The first question may be whether the request is about borrowing, research support, or a technical problem.

## Then the path narrows

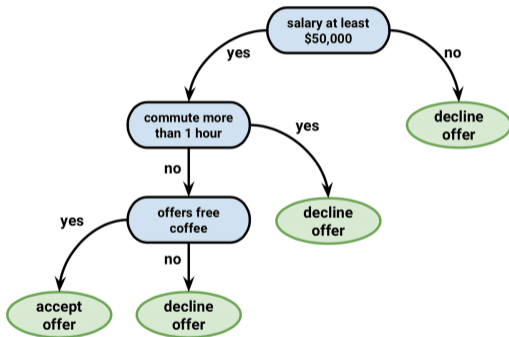
Inside “technical problem,” the next split may ask whether the user cannot log in or whether the file format is unreadable.

The power comes from question order, not from one giant rule written in advance.



# Reading One Path as a Decision

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER



## Inference is one route

For a new case, the model starts at the root and follows only the branch answers that apply to that case.

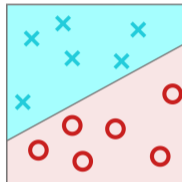
## The teaching sentence

Training builds the map. Inference walks one path through it.

# Classification Tree and Regression Tree

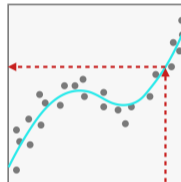
ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

**Classification** Groups observations into "classes"



Here, the line classifies the observations into X's and O's

**Regression** predicts a numeric value



Here, the fitted line provides a predicted output, if we give it an input



## Classification tree

The leaf gives a label such as approve or reject, spam or safe, high risk or low risk.

## Regression tree

The leaf gives a number such as waiting time, demand, or expected cost.

# When Growth Must Stop

## If the tree stops too soon

The model remains too coarse and misses useful structure that is really present in the data.

## If the tree grows too far

It starts memorizing small quirks, rare rows, and fragile rules that look sensible only on the training set.

## Why pruning exists

Pruning removes branches whose extra detail does not improve trustworthy performance enough to justify the added complexity.

# Why One Tree Is Unstable

## A small data change can reshape the tree

Because each split is chosen greedily, a slightly different sample can send the whole structure down another route.

## What this feels like

Two classes looking at slightly different survey samples may produce noticeably different first questions even when both are trying to predict the same thing.

## The practical consequence

A single tree is often easy to explain but less stable than students first expect.



# Bagging in Plain Language

## Core idea

Train many trees on slightly different resampled versions of the data and then combine their outputs.

## How to picture it

If one tree is moody, a crowd of trees can calm the decision down by averaging or voting.

## What it mainly buys

Bagging mainly reduces variance. It makes the answer less sensitive to one lucky or unlucky sample.

# Random Forest Intuition

工程研究中心  
ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## What random forest adds

It keeps the bagging idea but also limits the features each split can consider, so the trees do not all become too similar.

## A classroom analogy

If every student can only look at a different subset of clues before voting, the discussion becomes more diverse and the group is less likely to share one blind spot.

## What to remember

Diversity among trees is part of the strength. If every tree behaves the same way, the forest is not doing much extra work.

# Boosting as Repeated Repair

## A different spirit

Boosting does not mainly average many independent trees. It adds trees one after another, letting later trees focus on what earlier ones handled poorly.

## How to picture it

The first tree gives a rough answer. The next tree studies the leftover error. Another tree corrects what still remains.

## Why people like it

It can become very strong on tabular problems because it keeps repairing difficult cases instead of treating every case equally.

# Bagging and Boosting Feel Different

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Bagging mood

Many separate trees vote together so that no one tree dominates the final answer too much.

## Boosting mood

Each new tree is less a voter and more a critic of the current model's weakness.

Bagging says, “let many voices stabilize the answer.” Boosting says, “keep correcting the answer until the error shrinks enough.”



# Choosing Between Them

## When random forests often feel comfortable

They are robust, relatively forgiving, and useful when we want strong tabular performance without tuning every corner too aggressively.

## When boosting often shines

It can extract more performance when the problem rewards careful sequential correction and the pipeline is managed well.

## The general-ed takeaway

Do not memorize brand names first. Understand the tradeoff between stability, sensitivity, and complexity.

# Why Accuracy Alone Misleads

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## The easy trap

If most cases belong to one class, a model can look strong on overall accuracy while quietly failing on the minority cases that matter most.

## A familiar example

In rare-disease screening, saying “healthy” to almost everyone may score well on paper and still miss the cases the system was built to catch.

## Why this belongs here

Trees and ensembles can produce very convincing headline numbers. The harder question is whether the metric matches the real cost of error.

# Confusion Matrix Means Four Outcomes

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## What the matrix does

It separates correct positives, correct negatives, missed positives, and false alarms instead of hiding them inside one overall score.

## Why teachers like it

The confusion matrix forces us to ask, “Which kind of mistake did the model make?” instead of only, “How often was it right?”

Two models with the same accuracy can imply very different lived experiences if one misses urgent cases while the other mainly creates extra review work.



# Precision and Recall Depend on Harm

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Precision asks

When the model says “positive,” how often should we believe it?

## Use case

Spam filters care about precision because users become angry if important emails are falsely thrown away.

## Recall asks

Among the true positive cases, how many did we actually catch?

## Use case

Safety screening often cares about recall because missing a dangerous case can be much worse than reviewing an extra false alarm.

# Thresholds Turn Scores into Action

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## What threshold means

Many classifiers produce a score before they produce a final yes-or-no decision. The threshold tells us where to place that cut.

## Why this matters

Lowering the threshold usually catches more positives but also raises false alarms. Raising it does the opposite.

The model gives evidence. The threshold turns evidence into action.



# Calibration Means Honest Confidence

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## A calibrated model

If the system says “70% chance,” then cases like that should really be positive about seven times out of ten.

## Why this matters beyond ranking

In medicine, finance, or policy, people often need probabilities they can reason with, not just a sorted list.

## Important warning

A model can rank cases well and still give misleading confidence values.



# A Screening Story Makes This Real

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Imagine a mental-health check-in tool

The system helps flag students who may need human follow-up, but the goal is not to replace counseling staff.

## What matters here

Missing a struggling student, overwhelming staff with false alarms, and assigning false confidence all carry different human costs.

This is where evaluation stops being only math and becomes a question of purpose.



# One Refund Model Can Fail Three Ways

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Mistake one

Approve too many suspicious refunds, and the platform loses money.

## Mistake two

Reject too many genuine refunds, and customers lose trust quickly.

## Mistake three

Give overconfident scores, and human reviewers may stop checking borderline cases carefully.

## Why this example helps

One model can create financial cost, fairness cost, and trust cost at the same time.

# Common Tree Failure Modes

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Failure pattern one

The tree keeps splitting on accidental quirks that look important only because the dataset is limited.

## Failure pattern two

The tree locks onto a proxy variable that is easy to measure but only loosely connected to what we really care about.

## Failure pattern three

The environment changes after deployment, so yesterday's useful threshold becomes today's misleading shortcut.

Every one of these failures can still produce a clean-looking diagram.

# Leakage and Drift Still Matter

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Leakage

The model accidentally sees clues during training that would not be available at real decision time.

## Drift

The relationship between variables and outcomes changes because people, institutions, prices, or habits change.

## Why trees are not immune

Threshold-based models can become especially brittle when the meaning of those thresholds shifts in the real world.

# Interpretability Has Limits

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## What is genuinely useful

A small tree lets people inspect the path and ask whether the sequence of questions is reasonable.

## Where that comfort ends

A very large tree, or an ensemble of hundreds of trees, quickly stops feeling like a simple conversation and starts feeling like machinery again.

## Important nuance

“More interpretable than a deep net” is not the same as “transparent enough for every serious use.”

# Trees Can Freeze Old Bias

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## How bias can enter

If past decisions reflected unequal opportunity or uneven observation, the tree may discover those patterns and repeat them faithfully.

## Why this matters in a general course

The issue is not only technical. A model can encode institutional history even when every split looks clean and objective.

## What interpretability can and cannot do

Interpretability helps us inspect bias. It does not automatically remove bias.



# Tree or Neural Network

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## Why trees remain important

On many tabular problems, especially when data is structured and not enormous, tree ensembles are still extremely competitive.

## A simple rule of thumb

Loan tables, customer records, and operational dashboards often suit trees well. Images, speech, long text, and open-ended language tasks usually push us toward neural models.

Model choice is not a ranking of “old” versus “new.” It is a judgment about data shape, task demands, explanation needs, and deployment risk.



# Why AI4 and NN4 Belong Together

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

## What AI4 adds to the course

We now have one more model family that thinks in local questions instead of one global formula.

## What NN4 adds next

NN4 returns to neural models and asks how many simple units compute and learn together once the network becomes deeper.

# Summary

## 先进计算与数字工程研究中心

ADVANCED COMPUTING AND DIGITAL TECHNOLOGY RESEARCH CENTER

- Decision trees turn prediction into a sequence of questions that split the data into smaller, clearer groups.
- One tree is readable but unstable; ensembles improve stability and power in different ways.
- Evaluation is about the meaning of errors, not only about headline accuracy.
- Thresholds, calibration, and confusion matrices matter because different mistakes create different real costs.
- A readable model still needs skepticism about drift, leakage, bias, and false confidence.



### Where this story goes next

NN4 returns to neural computation and asks what happens when many simple units are stacked, activated, and trained together.

### What to watch for

Tree models organize questions over rows and columns. Neural models transform signals through layers. The evaluation habit from AI3 and AI4 still stays with us.



# Thank You

