ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

Lecture 3a – Regression, Error, and Practical Prediction

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

What we are trying to answer

How can a machine predict a number well enough to help, without pretending the future is fully knowable?

Where this shows up

Price, waiting time, travel time, energy use, demand, score, and risk are all numerical targets.

What changes from AI2

AI2 asked whether the data story is trustworthy. AI3 keeps that question, but now the answer is not a label. It is an amount.

From AI2 to AI3

Last time

We spent our time asking whether the rows, labels, and splits were honest enough for learning.

Today

We ask what happens when the model must estimate “how much” instead of choosing “which class”.

One simple way to say it

Classification chooses a side. Regression estimates an amount.

Why Numerical Prediction Matters

Numbers quietly shape ordinary decisions

- a map estimates arrival time
- a platform estimates demand later tonight
- a bank estimates risk before approving credit
- a power grid estimates tomorrow's load

A familiar scene

Even a weather app is not handing you certainty. It is handing you the best numerical estimate it can make from past patterns and current signals.

Regression Means Predicting an Amount

Classification asks which bucket

- spam or not spam
- cat or dog
- approve or reject

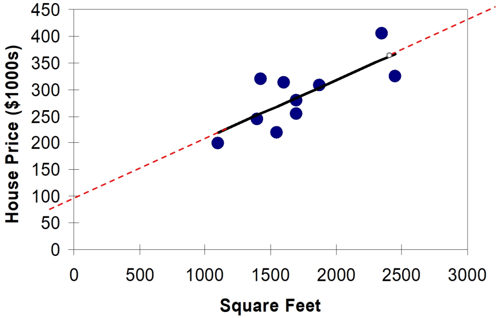
Regression asks how much

- how long will the ride take
- how much electricity will be used
- what price is reasonable today

The practical difference

Both tasks learn from old examples, but regression returns a number that people often use for planning, budgeting, or risk judgment.

A Running Example: Housing Price



What the picture suggests

Larger homes often cost more, so a line is a reasonable first guess.

What the picture also warns us

Area is not the whole story. Location, age, renovation, school district, and timing still matter.

Prediction Is Not Prophecy

What the model really gives

It gives a disciplined estimate based on the patterns it has seen before. It does not promise that the world will obey the estimate next week.

Why decimals can mislead

Students often see a number like 528.37 and feel that the model must know something exact. Usually it only means the software printed many digits.

The caution

More digits do not mean more truth.

From Table to Estimate

A tiny housing table

Area	Age	Price
78	12	168
92	7	215
105	5	246

What the data table is doing

Each row records a situation on the left and the outcome that really happened on the right.

What learning means here

The model tries to turn the left side into a usable estimate of the right side.

The Simplest Regression Form

Core formula

$$\hat{y} = w^\top x + b$$

How to read it

Take the input values, weight them, add them together, shift with a bias, and produce one estimate.

Housing version

If size, age, and distance to downtown are our features, the model gives each one a role in the final price estimate.

What this formula does not prove

It describes association inside the dataset. It does not automatically prove cause in the real world.

What the Parameters Mean

Weights

Weights tell the model how strongly to react when one feature changes while the others stay fixed.

Bias

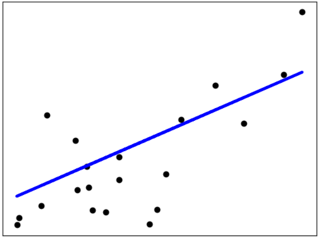
The bias gives the model a starting level. Without it, the estimate would be forced to pass through zero in a way the data may not support.

A common over-reading

Students often jump from “large weight” to “important cause.” The dataset usually cannot support that jump by itself.

Parameters tell us how the model is using the recorded variables. They do not tell us the full hidden mechanism of the world.

Residuals Show Where the Model Misses



Residual

$$e_i = y_i - \hat{y}_i$$

It is the gap between what happened and what the model predicted.

Why teachers care about this

If the misses form a pattern, the model is still missing structure.

Why Start with a Line

Why teachers start here

A straight line is easy to read, easy to compare, and good enough to reveal whether the data has any usable signal at all.

Why practitioners still start here

If a simple baseline already works reasonably well, a more complex model should earn its extra complexity instead of being taken on faith.

The right attitude

A line is a starting point. It is not a belief that the world must be linear.

Loss Says Which Errors Hurt

Absolute error

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Every miss counts in direct proportion to its size.

Squared error

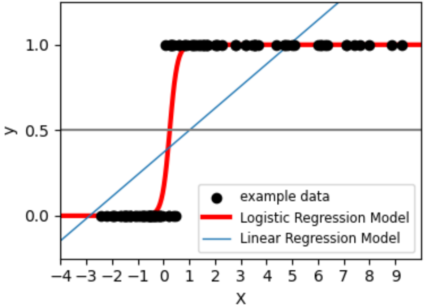
$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Large misses are punished much more sharply.

Why this matters

Choosing a loss function means deciding which kinds of mistakes the model should fear most.

Why Squared Error Changes the Fit



What changes

Missing by 2 units becomes 4 after squaring. Missing by 10 units becomes 100. Big errors start dominating the objective very quickly.

When that is useful

If large misses are especially costly, squared error pushes the fit to take them seriously.

Training Means Reducing Error

What training is doing

The algorithm keeps adjusting the parameters so the total loss on the training set becomes smaller.

A plain metaphor

It is closer to tuning a recipe by repeated tasting than to proving a theorem in one perfect step.

Why this matters

We are not hand-drawing a line. We are repeatedly checking error and moving toward a better setting.

Gradient Descent Means Repeated Correction

One compact formula

$$\theta \leftarrow \theta - \eta \nabla J(\theta)$$

How to read it

Look at the current error, see which direction makes it rise fastest, and then step in the opposite direction.

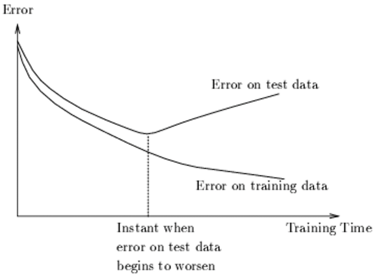
A physical picture

It is like walking down a hill in the dark: you feel which way is steepest, take a step downward, and repeat.

What the learning rate controls

Too large and you overshoot. Too small and you crawl.

Validation Checks Tomorrow, Not Yesterday



Why we hold data back

A model can improve on familiar rows while becoming worse on new cases.

What validation is really asking

Not “Did you memorize?” but “Do you still help when the next case arrives?”

Generalization Is the Real Goal

What really matters

In class we may admire a neat fitted curve. Outside class, the real test is whether the estimate still helps on tomorrow's cases.

A housing example

If a model only memorizes last year's unusual market, it may look strong on old records and fail badly after an interest-rate shock or policy change.

One sentence to keep

Good fit on old data is evidence. It is not the same thing as trustworthy performance in a changing world.

Regression Already Runs Daily Life

Where students already meet it

Travel time estimates, ride-hailing prices, waiting-time forecasts, recommendation scores, and electricity-demand forecasts are all numerical prediction systems.

The modern AI connection

Even very large AI systems still depend on internal scoring and estimation. Not every important model output is a sentence or an image.

A Number Should Support Judgment

What a good estimate does

It narrows uncertainty enough to help a person plan, compare options, or notice risk earlier.

What it should not do

It should not silence human judgment or pretend that one output number settles the whole decision.

Models help people decide under uncertainty. They do not remove uncertainty from the world.

A Coin Story and a Lottery Story

What the comparison teaches

Regression can estimate regular patterns such as average demand or average delay. It cannot squeeze personal certainty out of events that are close to random.

The deeper lesson

Some uncertainty is not a failure of the algorithm. It is a property of the world.

Prediction Is Not Explanation

Prediction asks

Can this pattern help me estimate the next case well enough to be useful?

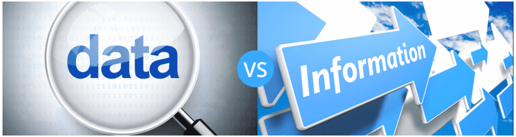
Explanation asks

What mechanism in the world produced that pattern in the first place?

Why people mix them up

When a model predicts well, it is tempting to assume it has discovered the true cause. That conclusion needs stronger evidence than fit alone.

Features Decide What the Model Can Notice



What the model sees

The model only sees the variables that survived the measurement process. Anything never recorded is invisible to it.

Why this matters in regression

If price depends heavily on school district, renovation quality, or seasonality and we never record them, the fitted line will always be trying to explain too much with too little.

Why Simple Baselines Still Matter

Why they stay valuable

They are quick to train, easy to read, and strong enough to show whether the project has any usable signal at all.

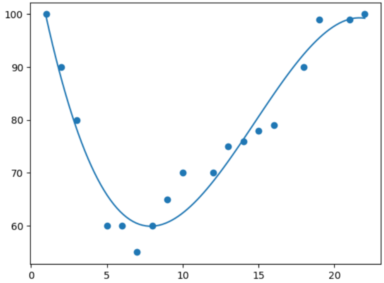
Why this is still current

A good baseline remains one of the best ways to detect hype, leakage, or meaningless complexity.

A practical use

If a transparent linear baseline already works well, a more complex model should justify itself with a real gain, not with fashion.

When One Line Is Too Rigid



What the picture says

The relation bends. One straight rule misses the shape because the pattern changes across the input range.

What this usually means

Thresholds, saturation, interactions, and mixed regimes are often signs that one global line is too simple.

Extrapolation Is a Quiet Risk

Interpolation

Estimating between examples we have already seen is usually safer than estimating far beyond them.

Extrapolation

Using the same fitted trend far outside the observed range can look confident on the slide and fail badly in real life.

A housing warning

A model trained mostly on ordinary apartments should not be trusted automatically for luxury villas or unfinished rural houses.

Some Uncertainty Will Not Go Away

What remains unpredictable

Human choices, sudden policy changes, weather shocks, missing variables, and plain randomness can all leave uncertainty that the model cannot remove.

The sober reading

Better models can reduce avoidable error. They cannot remove uncertainty that was never observable in the first place.

Two Common Misreadings

Misreading 1

"A precise-looking number must be trustworthy."

Better reading

Trust comes from data quality, residual analysis, validation, and sensible use, not from the number of digits after the decimal point.

Misreading 2

"If the error is not zero, the model failed."

Better reading

Many real tasks become useful long before prediction becomes perfect, because the real goal is support under uncertainty.

When Error Is Asking for a New Model

What patterned error can mean

Important variables are missing, the relation is nonlinear, the target changed over time, or several different situations were mixed into one dataset.

What to do first

Before rushing to a fancier method, look at the residuals, the data source, and the use case. Sometimes the next fix belongs outside the model.

Why this is the bridge forward

The best reason to move beyond linear regression is not fashion. It is a clear pattern in the error that the current model cannot express.

Why AI3 and NN3 Belong Together

What AI3 gives us

We now have a working language for prediction, residuals, loss, validation, and the difference between useful estimates and empty certainty.

What NN3 adds

NN3 asks how a model can build richer internal representations so that harder patterns become easier to predict.

Why AI4 Follows Naturally

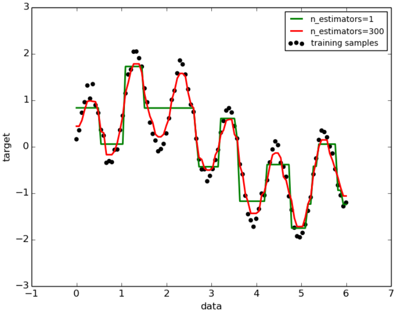
What linear regression taught us

One global rule is elegant and interpretable, but it can become too rigid when the data bends, branches, or changes regime.

What decision trees will offer

Instead of one line for everyone, a tree can split the space into smaller regions and use different local rules in different cases.

Where Trees Enter



Why this picture matters

The green steps adapt locally. They do not insist on one smooth global line.

What to remember

Different model families are different ways of organizing approximation. Each one trades off simplicity, flexibility, and interpretability.

Summary

- Regression is about estimating an amount, not choosing a class.
- A linear model turns recorded features into a numerical estimate through weights and a bias.
- Residuals, loss, and validation matter more than a pretty fitted line.
- Good prediction helps judgment under uncertainty; it does not erase uncertainty.
- We move beyond one line when the error pattern shows that one global rule is too rigid.

Next Time

Where this story goes next

AI4 moves from one global regression rule to model families that split the world into smaller pieces and then combine many weak rules into stronger predictors.

Carry this question with you

If one line is too rigid, what is a sensible way to divide the problem without losing the habit of checking error honestly?

Thank You