# Introduction to Interpolation and Extrapolation

## Lecture 21 | Estimating values from discrete data

| | |
|---|---|
| **Core idea** | Construct a function that agrees with known data points. |
| **Key tools** | Linear interpolation, polynomial models and graphical interpretation. |
| **Learning focus** | Distinguish interpolation from extrapolation and assess reliability. |

# Learning outcomes

**By the end of this lecture, students should be able to:**

| | |
|---|---|
| **Define** | Distinguish interpolation, extrapolation and curve fitting. |
| **Calculate** | Estimate unknown values using linear and quadratic interpolation. |
| **Interpret** | Recognize error, node spacing and risks outside the data range. |

# Why interpolate?

**Measurements are often available only at discrete points.**

| | |
|---|---|
| **Known data** | A table contains temperature at selected times or distance at selected positions. |
| **Unknown value** | Estimate a value between observations without repeating the experiment. |
| **Model** | Build a mathematical function through the available data points. |

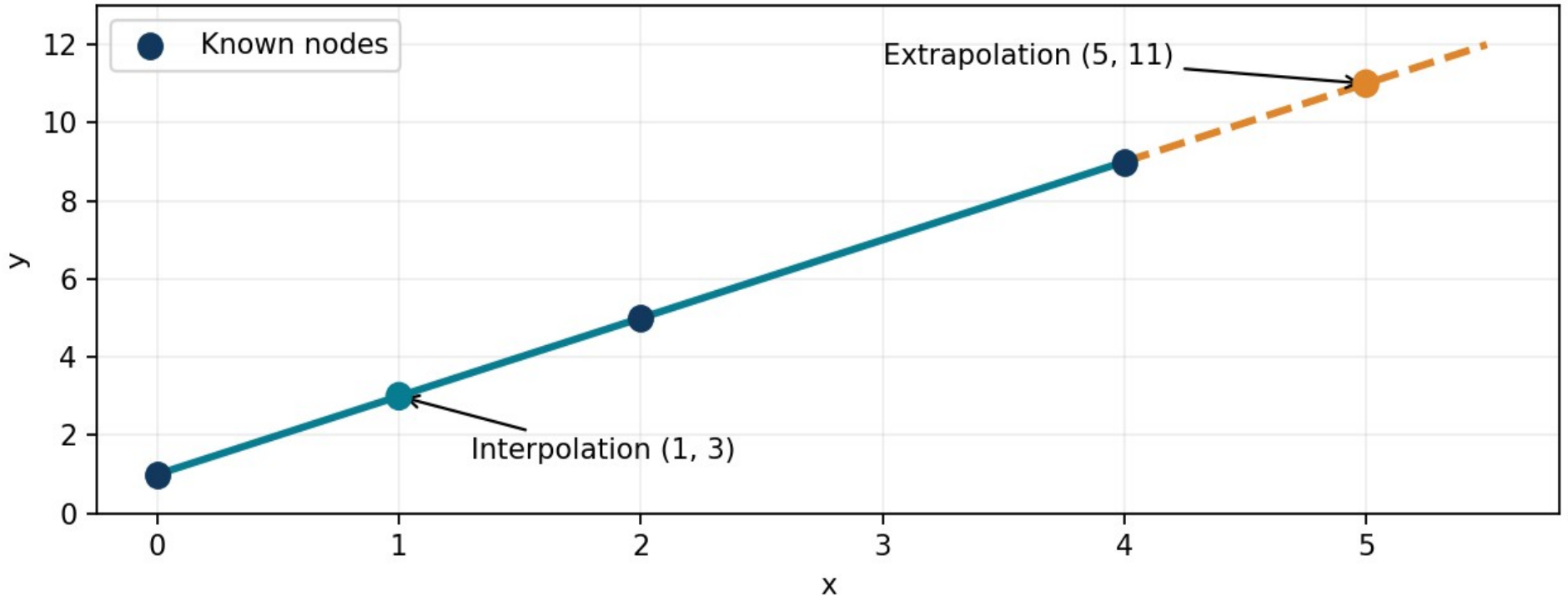
# Interpolation versus extrapolation

**The location of the requested input determines the task.**

| | |
|---|---|
| **Interpolation** | Estimate f(x) for x within the range of observed inputs. |
| **Extrapolation** | Estimate f(x) for x outside the range of observed inputs. |
| **Caution** | Outside the observed range, assumptions about the trend become more important. |

# Graphical interpretation

## Example: data points (0, 1), (2, 5), (4, 9).



# Linear interpolation

## Use the straight line through two distinct nodes.

**Slope**

$m = (y_1 - y_0)/(x_1 - x_0)$, with $x_1 \neq x_0$.

**Formula**

$P_1(x) = y_0 + (x - x_0)(y_1 - y_0)/(x_1 - x_0)$.

**Weights**

With $t = (x - x_0)/(x_1 - x_0)$, $P_1(x) = (1 - t)y_0 + ty_1$.

# Example 1 — temperature

**At 10:00 the temperature is 20°C; at 12:00 it is 26°C. Estimate at 11:30.**

**Step 1** $x_0 = 10$, $x_1 = 12$, $x = 11.5$; $t = 1.5/2 = 0.75$.

**Step 2** $P_1(11.5) = 20 + 0.75(26 - 20)$.

**Result** 24.5°C, assuming a linear change between measurements.

# Example 2 — table lookup

**Data: x = 1, 3, 5; corresponding y = 2, 8, 18. Estimate y at x = 4.**

| | |
|---|---|
| **Select nodes** | Use the adjacent points (3, 8) and (5, 18). |
| **Calculate** | $P_1(4) = 8 + (4 - 3)(18 - 8)/(5 - 3)$. |
| **Result** | $P_1(4) = 13$. The estimate is between the two observed values. |

# Example 3 — extrapolation

**Use the last two nodes (3, 8) and (5, 18) to estimate y at x = 6.**

| | |
|---|---|
| **Calculate** | $P_1(6) = 8 + (6 - 3)(18 - 8)/(5 - 3)$. |
| **Result** | $P_1(6) = 23$; $t = 1.5$ lies outside $[0, 1]$. |
| **Risk** | The linear extension need not match the actual function beyond $x = 5$. |

# Polynomial interpolation

**For n + 1 distinct nodes, a unique polynomial of degree at most n exists.**

| | |
|---|---|
| **Two nodes** | Degree at most 1: a straight line. |
| **Three nodes** | Degree at most 2: a quadratic polynomial. |
| **More nodes** | Higher-degree polynomials can fit the data exactly but may oscillate. |

# Example 4 — quadratic model

**Fit $P_2(x) = a + bx + cx^2$ through (0, 1), (1, 3), (2, 7).**

| | |
|---|---|
| **Equations** | $a = 1$; $a + b + c = 3$; $a + 2b + 4c = 7$. |
| **Solve** | $b + c = 2$; $b + 2c = 3$; therefore $c = 1$ and $b = 1$. |
| **Result** | $P_2(x) = 1 + x + x^2$; $P_2(1.5) = 4.75$. |

# Interpolation versus curve fitting

**Both construct models from data, but their objectives differ.**

| | |
|---|---|
| **Interpolation** | Passes through every distinct data node exactly (in exact arithmetic). |
| **Curve fitting** | Usually minimizes a measure of discrepancy rather than passing through all points. |
| **Choice** | Use fitting for noisy observations; interpolation is useful for trusted tabulated values. |

# Linear interpolation error

**For f twice continuously differentiable on the interval:**

| | |
|---|---|
| **Remainder** | $f(x) - P_1(x) = f''(\xi)(x - x_0)(x - x_1)/2$ for some $\xi$ in the interval. |
| **Bound** | If $\lvert f''\rvert \le M$, then $\lvert f(x) - P_1(x)\rvert \le M\, h^2/8$, where $h = x_1 - x_0$. |
| **Meaning** | Halving the spacing reduces this bound by a factor of four. |

# Example 5 — error bound

**For $f(x) = x^2$, interpolate at $x = 1.5$ using nodes 1 and 2.**

| | |
|---|---|
| **Estimate** | $P_1(1.5) = 1 + 0.5(4 - 1) = 2.5$. Exact $f(1.5) = 2.25$. |
| **Actual error** | Absolute error $= \lvert 2.25 - 2.5 \rvert = 0.25$. |
| **Bound** | $M = 2$, $h = 1$; $M h^2/8 = 0.25$. The bound is attained. |

# Algorithm — piecewise linear lookup

**Assume sorted distinct nodes and a target within the data range.**

| | |
|---|---|
| **Locate** | Find adjacent nodes $x_i \le x \le x_{i+1}$. Return known y if x equals a node. |
| **Compute** | $t = (x - x_i)/(x_{i+1} - x_i)$; return $(1 - t)y_i + ty_{i+1}$. |
| **Validate** | Reject outside-range targets unless extrapolation is explicitly intended. |

# Python implementation

## NumPy example for interpolation within the observed range.

**Data**

```
import numpy as np
xs = [1, 3, 5]; ys = [2, 8, 18]
```

**Estimate**

```
x = 4
assert xs[0] <= x <= xs[-1]
print(np.interp(x, xs, ys))  # 13.0
```

**Note**

np.interp defaults to endpoint values outside the range, not linear extrapolation.

# Practice questions

**Apply the formulas and identify the assumptions.**

**Q1** Given (2, 4) and (6, 12), estimate y at x = 3.

**Q2** Using the same nodes, estimate y at x = 8 and classify the task.

**Q3** For $f(x) = x^2$ on [0, 2], find the linear estimate at 1 and its absolute error.

# Solutions and recap

**Interpolation is a local estimation tool; extrapolation needs additional caution.**

| | |
|---|---|
| **Q1 / Q2** | $P_1(3) = 6$ (interpolation); $P_1(8) = 16$ (extrapolation). |
| **Q3** | $P_1(1) = 2$; $f(1) = 1$; absolute error = 1. |
| **Next lecture** | NC_22 — Finite Differences and Difference Tables. |