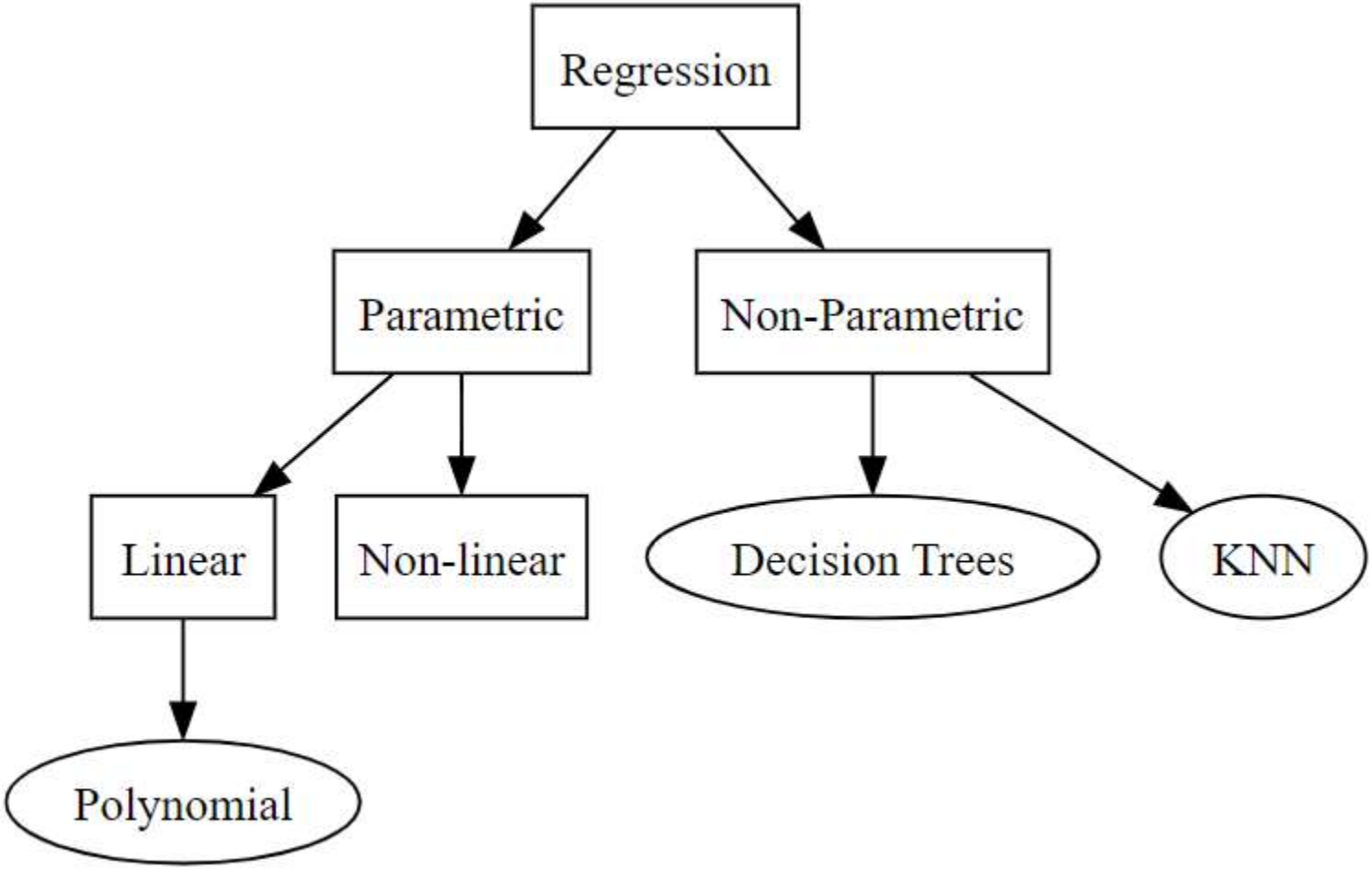
# Regression

Introduction to
Parametric and Non-Parametric Models


Regression
Parametric
Non-Parametric
Linear
Non-linear
Decision Trees
KNN
Polynomial

# Purpose

**Regression aims to predict continuous output variables based on the input variables.**

- The output variables are the dependent variables
- The input variable are the independent variables, also called features

## Parametric Model

This is a form of regression in which the relationship between the dependent and independent variables is assumed to have a specific functional form. For instance, a polynomial relationship.

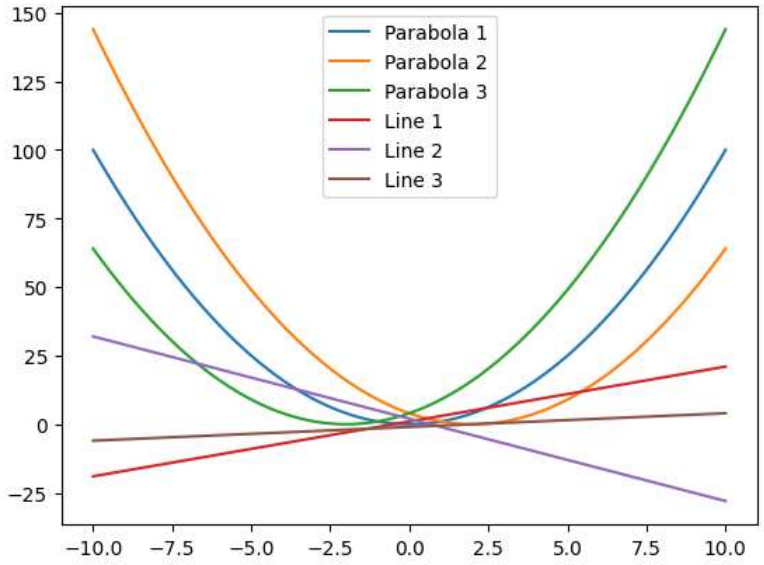
## Non-Parametric Model

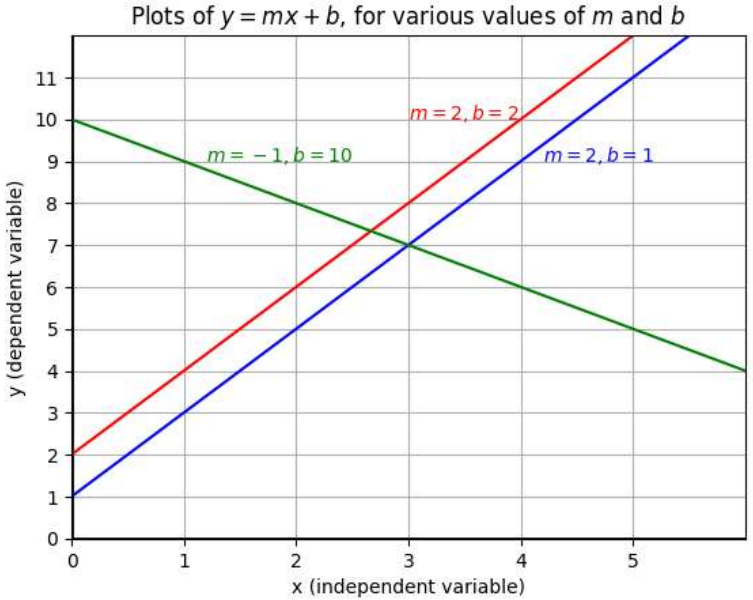
This form of regression makes no assumptions about the functional form of the relationship between predictors and response variable. Instead, it tries to model the data according to its distribution and inherent structure. Decision trees and k-nearest neighbors (k-NN) algorithms are examples of non-parametric regression techniques.

# Parametric Models

Polynomial relationship:

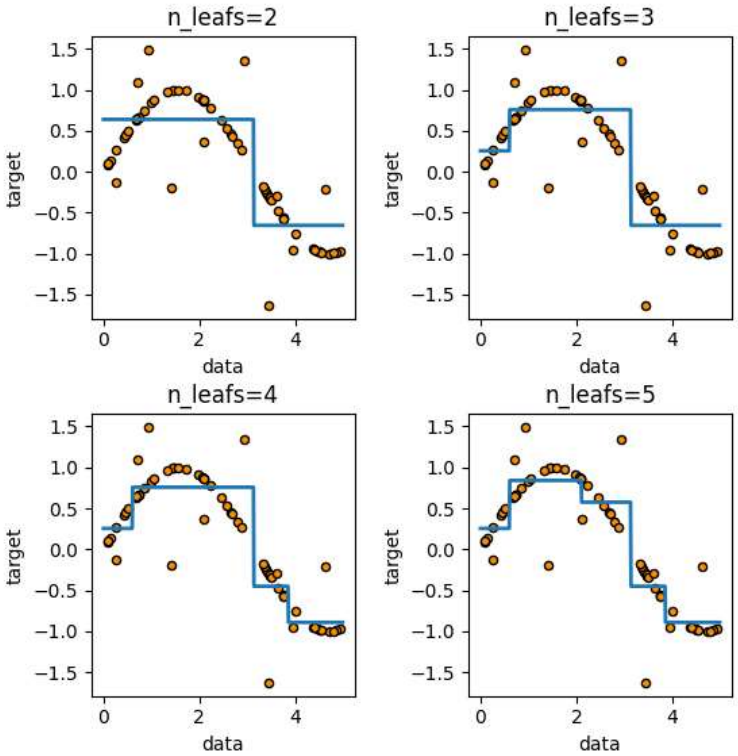
$$f(X, \beta) = \beta_0 + \beta_1 * X_1 + \beta_2 * X_1^2 + ... + \beta_n * X_1^n$$



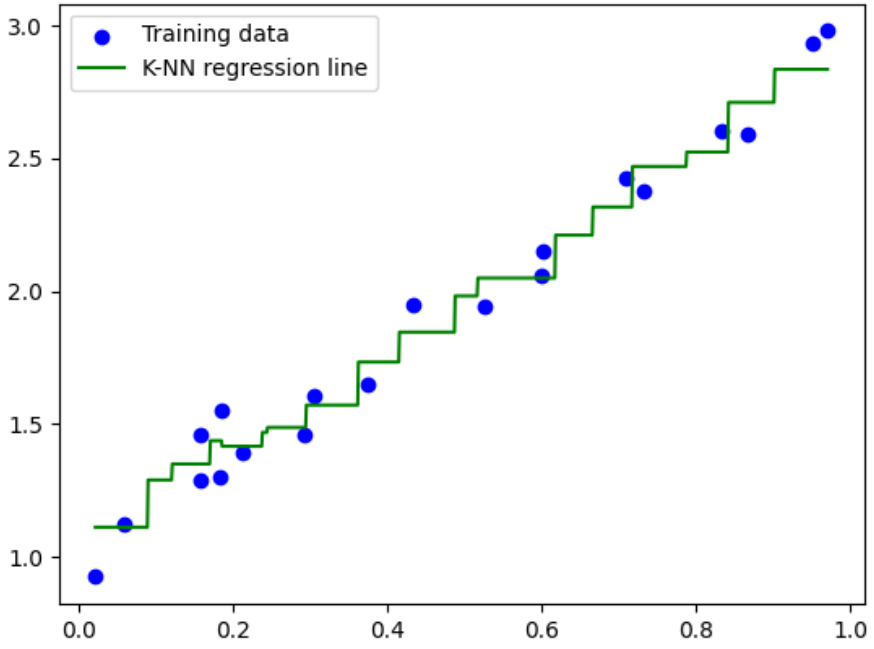



# Non Parametric Models

## Decision Trees

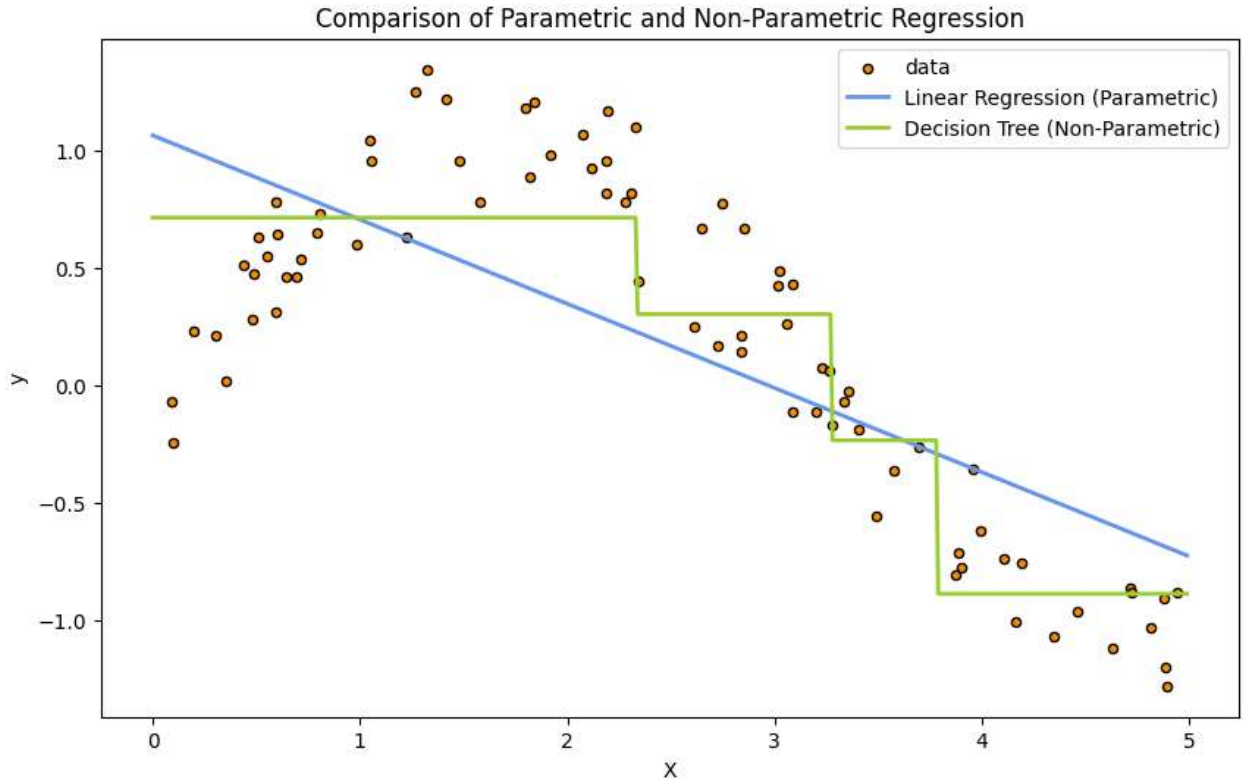



## K Nearest Neighbors




k=3

# Example


Comparison of Parametric and Non-Parametric Regression
data
Linear Regression (Parametric)
Decision Tree (Non-Parametric)
y
X
1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
4
5

# Pros and Cons

| | Parametric Regression | Non-Parametric Regression |
|---|---|---|
| **Pros** | - Simplicity<br>- Efficiency<br>- Interpretability<br>- Good for Extrapolation | - Flexibility<br>- Robustness |
| **Cons** | - Potential Bias | - Less Efficiency<br>- Less Interpretability<br>- Poor Extrapolation |

# Parametric Model - Pros and Cons

| | Parametric Regression | Non-Parametric Regression |
|---|---|---|
| **Pros** | - Simplicity<br>- Efficiency<br>- Interpretability<br>- Good for Extrapolation | - Flexibility<br>- Robustness |
| **Cons** | - Potential Bias | - Less Efficiency<br>- Less Interpretability<br>- Poor Extrapolation |

**Parametric Regression:**

Pros:

1. **Simplicity:** Parametric models are generally simpler, as they make specific assumptions about the relationship between variables.
2. **Efficiency:** Because they use a predetermined form, parametric models usually require less data and less computational resources to produce an accurate model.
3. **Interpretability:** The parameters of the model have a clear interpretation, which can provide insights into the underlying process that generated the data.
4. **Extrapolation:** If the parametric form of the model is correct, these models can extrapolate predictions outside the range of observed data.

Cons:

1. **Bias:** Parametric models can be biased if the chosen form of the function doesn't match the true relationship between variables. This can lead to underfitting.

# Non-Parametric Models - Pros and Cons

| | Parametric Regression | Non-Parametric Regression |
|---|---|---|
| **Pros** | - Simplicity<br>- Efficiency<br>- Interpretability<br>- Good for Extrapolation | - Flexibility<br>- Robustness |
| **Cons** | - Potential Bias | - Less Efficiency<br>- Less Interpretability<br>- Poor Extrapolation |

**Non-parametric Regression:**

Pros:

1. **Flexibility:** Non-parametric models make fewer assumptions about the form of the relationship, allowing them to adapt to the complexity of the data. This makes them useful for non standard or complex functional relationships.
2. **Robustness:** These models can often handle outliers and non-standard distributions of errors better than parametric models.

Cons:

1. **Efficiency:** Non-parametric models usually require more data and more computational resources to produce an accurate model.
2. **Interpretability:** Non-parametric models typically don't provide as much insight into the underlying process that generated the data.
3. **Extrapolation:** Non-parametric models often don't extrapolate well outside the range of the observed data.

# Use-cases for parametric and non-parametric regression

Each type of regression serves its own set of use cases and contexts.

**Parametric regression:**

Parametric methods are often favored in contexts where we have substantial prior knowledge about the functional form that the relationship between inputs and outputs should take.

- **Scientific Research:** In fields such as physics, engineering, and social sciences, the relationships between variables are often hypothesized based on theory. Parametric regression, such as linear regression, allows these theories to be formalized and tested.
- **Simplicity and Interpretability:** If you need clear interpretability and explanation of each feature's effect on the outcome, parametric models like linear or logistic regression are useful. The coefficients in these models directly reflect the change in the output variable for a one-unit change in the input variable.

**Non-parametric regression:**

Non-parametric methods, on the other hand, are used when the underlying data distribution does not meet the assumptions required for parametric techniques, or when we have little prior knowledge about the functional form.

- **Complex and High-Dimensional Data:** Non-parametric methods like decision trees, random forests, or KNN can handle a wide variety of data structures that may not be easily adapted to a standard parametric form. They are particularly useful when dealing with high-dimensional data.
- **Robustness:** They are more robust to outliers and the presence of irrelevant features, since they don't make specific assumptions about the underlying distribution of data.

Remember that each type of regression has its strengths and weaknesses, and the choice between using a parametric or non-parametric model often depends on the specific context and the trade-off between bias and variance that an analyst is willing to make.

| Strategy | Description | How It Adapts to LLMs |
|---|---|---|
| **Pair Programming** | Pairs switch roles between coder and reviewer, or “driver” and “navigator.” | Even if using LLMs, students must *discuss* and *justify* choices—builds teamwork and accountability |


1. Study the algorithm to implement the regression using Decision Trees
2. Commit and push your solution to your GitHub
3. Prepare three talking points about the most important sections of your code

**When you are ready:**

**You**

4a. Clone your peer’s code, review it, and ask questions about the code

**Your peer**

4b. Will clone your code, review it, and ask questions about it.

5. Be ready to present both your notes and reflections **<u>as pairs</u>**