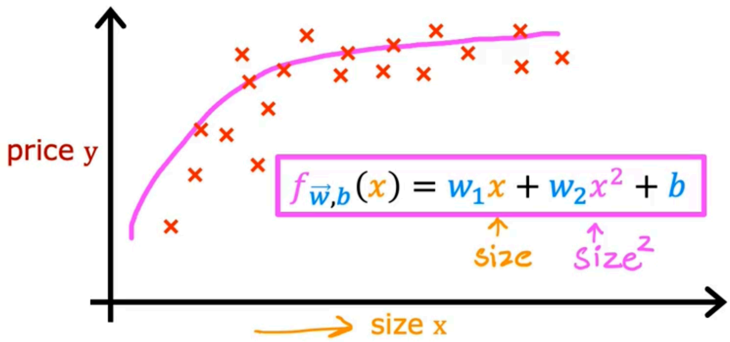
A. Polynomial Regression

Tuesday, 9 September 2025 10:47 PM

1. Introduction:

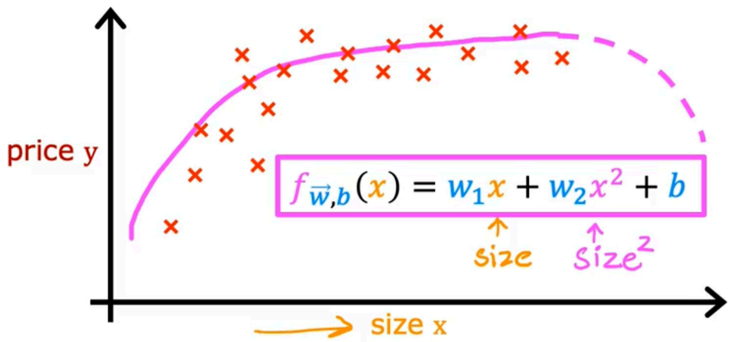
Polynomial Regression lets you fit curves, non-linear functions to your data.

For example, let's say we have a housing data as:

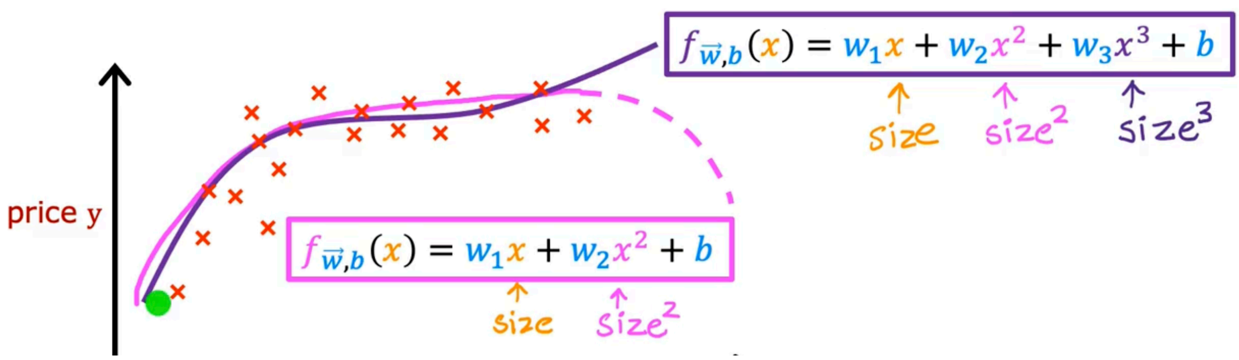


We couldn't have fit a straight line through the data points efficiently. But a quadratic curve can somewhat do the job.

In the above example, the coefficient of the second-degree term, w_2 must be -ve to fit the data points well. Now, since w_2 is -ve, the curve opens downwards (like a frowny face), and will eventually come down as the size of x increases.



Now we wouldn't really expect the housing prices to go down when the size increases, right? So, we can use a cubic feature (x^3). As we recall, introducing new features that improve our model is part of Feature Engineering.



Also, since we are introducing new features like x^2 or x^3, a small change in these features can sway the model results drastically and it's likely we wouldn't achieve a good result with our model.

So if we create features that are powers of the original feature, then Feature Scaling becomes increasingly important since we'll also use Gradient Descent.

This is Polynomial Regression. It is a form of linear regression where the relationship between the independent variable (x) and the dependent variable (y) is modelled as an n^{th} degree polynomial. It is useful when the data exhibits a non-linear relationship allowing the model to fit a curve to the data.

---

2. Need for Polynomial Regression:

- **Non-linear Relationships:** Polynomial regression is used when the relationship between the independent variable (input) and dependent variable (output) is non-linear. Unlike linear regression which fits a straight line, it fits a polynomial equation to capture the curve in the data.

- **Better Fit for Curved Data:** When a researcher hypothesizes a curvilinear relationship, polynomial terms are added to the model. A linear model often results in residuals with noticeable patterns which shows a poor fit. It can capture these non-linear patterns effectively.

- **Flexibility and Complexity:** It does not assume all independent variables are independent. By introducing higher-degree terms, it allows for more flexibility and can model more complex, curvilinear relationships between variables.

---

3. Working:

Polynomial Regression is an extension of Linear Regression where higher-degree terms are added to model non-linear relationships. The general form of the equation for a polynomial regression of degree *n* is:

$$\mathbf{y = w_1x^1 + w_2x^2 + w_3x^3 + \ldots\ w_nx^n + b + e}$$

where,
y is the dependent variable on x
x is the independent variable
$\mathbf{w_1, w_2, \ldots, w_n}$ are the coefficients of the polynomial terms
b represents the y-intercept
e represents the error rate

relationships. The general form of the equation for a polynomial regression of degree *n* is:

$$\mathbf{y = w_1x^1 + w_2x^2 + w_3x^3 + \dots w_nx^n + b + e}$$

where,
y is the dependent variable on x
x is the independent variable
$\mathbf{w_1,w_2,\dots,w_n}$ are the coefficients of the polynomial terms
b represents the y-intercept
e represents the error rate

While the regression function is linear in terms of the unknown coefficients, $\mathbf{w_1,w_2,\dots,w_n}$, the model itself captures non-linear patterns in the data. The coefficients are estimated using techniques like Least Square technique to minimize the error between predicted and actual values.

Choosing the right polynomial degree **n** is important: a higher degree may fit the data more closely but it can lead to overfitting. The degree should be selected based on the complexity of the data. Once the model is trained, it can be used to make predictions on new data, capturing non-linear relationships and providing a more accurate model for real-world applications.
