2.4 – Modeling with Polynomial Functions

1 Introduction to Polynomials

A **polynomial** function takes the form $y = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$ where n is a nonnegative integer and a_i is the i^{th} constant coefficient. The **degree** of the polynomial is n, the largest exponent of the independent variable [1]. For example, the function

$$F(x) = 2x^6 - x^4 + \frac{2}{5}x^3 + \sqrt{2}$$

is a polynomial of degree 6 [3]. A polynomial of degree 0 is a horizontal line $y = mx^0$, and a polynomial of degree 1 is of the form $y = mx + b$, so it is a linear function. Degree 2 polynomials of the form $y = ax^2 + bx + c$ are called **quadratic** and their graph is always a parabola. Polynomials of degree 3 are called **cubic**. The figures below show graphs of a quadratic, cubic, degree 4, and degree 5 polynomials.

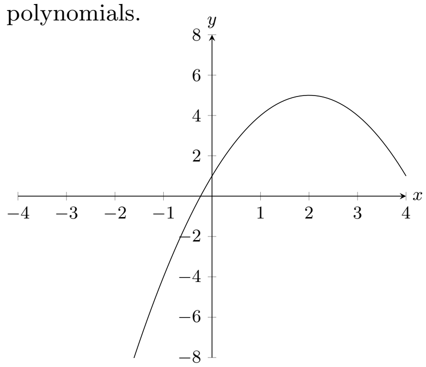


Figure 1: Graph of the quadratic function $y = -(x-2)^2 + 5$. This quadratic is in the form $y = a(x-b)^2 + d$. Notice that the sign of the leading coefficient a caused the graph to reflect over the line $y = d$, the constant b caused a translation in the positive x direction, and the constant d caused a translation in the positive y direction.

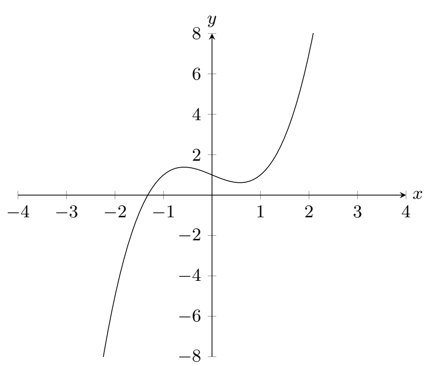


Figure 2: Graph of the cubic function $y = x^3 - x + 1$. Notice that as the degree of the polynomial increases, so does the number of possible *turning points*. As you progress in your mathematical career, you may recognize these as *local extrema*. A polynomial can have at most $n-1$ local extrema, where n is the degree of the polynomial [1]

.

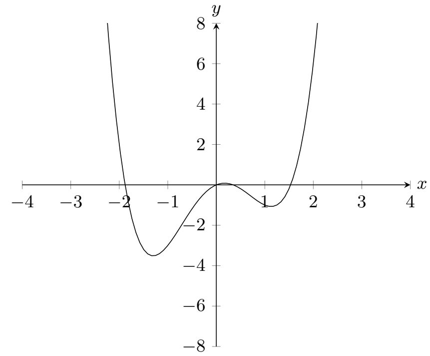


Figure 3: Graph of the degree 4 polynomial $y = x^4 - 3x^2 + x$. As we continue to increase the degree of our polynomials, notice that as $|x|$ increases, $|y|$ increases. In other words, as x increases, y either increases to infinity or decreases to negative infinity. Likewise, as x decreases, y either increases to infinity or decreases to negative infinity. This is the case for all polynomials, regardless of shape [1].

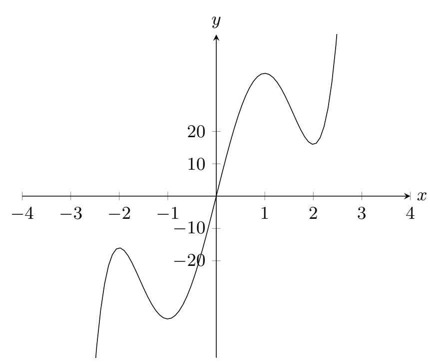


Figure 4: Graph of the degree 5 polynomial $y = 3x^5 - 25x^3 + 60x$. Notice that the general shape of the polynomial depends on if its degree, n, is even or odd. If n is even, then the polynomial is an even polynomial, and its graph is similar to a parabola in that y either increases or deceases as $|x|$ increases. If n is odd, then the polynomial is an odd polynomial, and its graph is similar to a cubic in that y both increases and decreases as $|x|$ increases.

1.1 Long Term Behavior

What we noted in Figure 3 is called the *long term behavior* of polynomial functions. We saw that as $|x|$ increases, $|y|$ increases. In other words, as x increases, y either increases to infinity or decreases to negative infinity. Likewise, as x decreases, y either increases to infinity or decreases to negative infinity. This is the case for all polynomials, regardless of shape [1].

1.2 Even and Odd Functions

What we noted in Figure 4 is the important distinction between *even* and *odd* polynomials. We saw that the general shape of the polynomial depends on if its degree, n, is even or odd. If n is even, as seen in Figure 5, then the polynomial is an even polynomial, and its graph is similar to a parabola in that y either increases or deceases as $|x|$ increases.

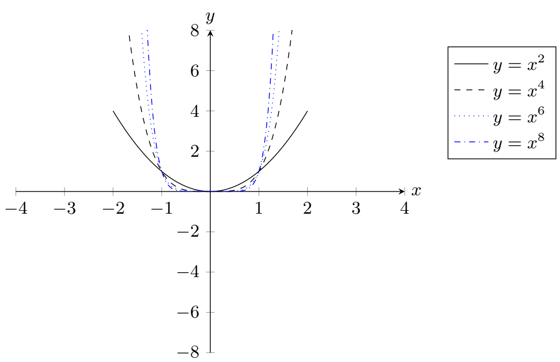


Figure 5: Even functions

If n is odd, as seen in Figure 6, then the polynomial is an odd polynomial, and its graph is similar to a cubic in that y both increases and decreases as $|x|$ increases. Notice from both Figure 5 and Figure 6, that as n increases, the graph of $y = x^n$ becomes flatter near 0 and steeper when $|x| \geq 1$ [3].

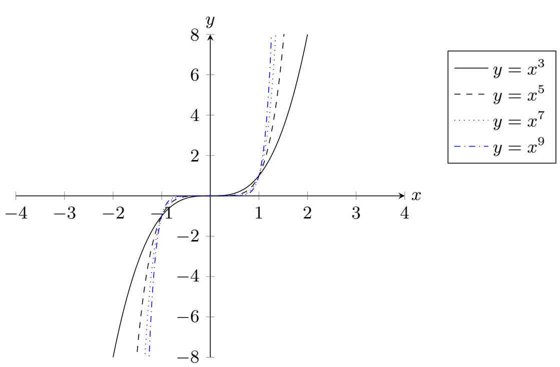


Figure 6: Odd functions

The following box summarizes behaviors of polynomials based on the degree of the function.

Polynomial Behavior Summary

- the graph of a polynomial function is an unbroken smooth curve
- the graph of a polynomial function of degree n has at most $n - 1$ local extrema
- for the graph of any polynomial function (other than a constant function), as $|x|$ gets very large, $|y|$ grows very large [1]

2 Worked Example

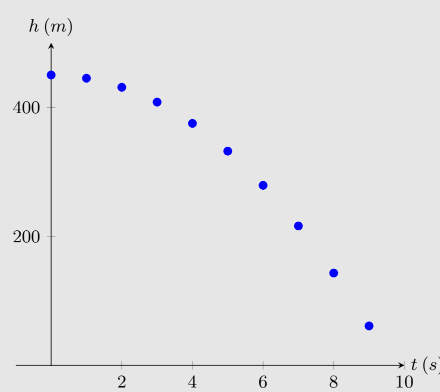
The following problem is an example where we will use an empirical approach, but can use our understanding of the context of the problem to do a first principles"sanity check" of our empirical model.

Problem 2.4.1: Modeling with Polynomials Problem

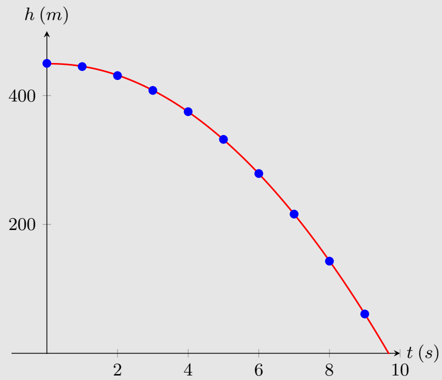
A ball is dropped from the upper observation deck of the CN Tower, 450 m above the ground, and its height h above the ground is recorded at 1-second intervals in the table below. Find a model to fit the data, predict the time at which the ball hits the ground.

Time (s)	Height (m)
0	450
1	445
2	431
3	408
4	375
5	332
6	279
7	216
8	143
9	61

Solution: We start by creating a scatter plot of the data and observe that a linear model is inappropriate, but it looks like the data points might lie on a parabola, so we try a quadratic model instead.



Using Excel, we find the quadratic model $h = 449.36 + 0.96t - 4.90t^2$ with $SSE = 1.5$ and $R^2 = 1$. Here we can do our "sanity check." We know from physics that projectile motion is parabolic and our data and model seem to follow this principle. This is an example where we can use our contextual knowledge to do a cursory check of our model, asking ourselves"does this make sense?"



The ball hits the ground when $h = 0$, so we solve the following equation using the quadratic formula.

$$-4.90t^2 + 0.96t + 449.36 = 0$$

The positive root is $t \approx 9.67$, so we predict that the ball will hit the ground after falling about 9.7 seconds [3].

3 Model Assessment

We will use SSE and R^2 to assess and compare the fit of our polynomial models. Remember that an R^2 value closer to one is an indicator of better fit, but is not a guarantee of a good model. This is because of the dangers of over and underfitting.

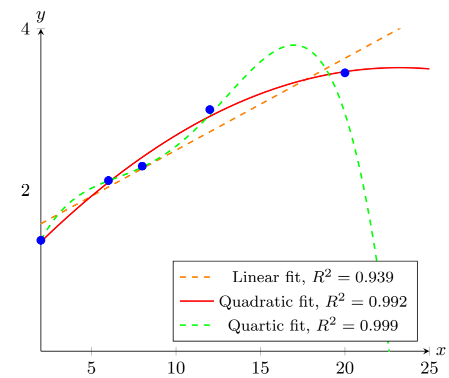
Overfitting makes a predictive model less accurate because the model aligns *too* closely to the given data instead of allowing for the inherent error or *noise* due to randomness [2]. An overfit models are typically more complex than necessary because they attempt to model noise instead of the overall trend of the data. Often a model that less precisely matches the data, but better captures the general structure of the relationship between the explanatory and response variables, will perform significantly better when making predictions [2].

Underfitting occurs when a model is too simple to capture the underlying relationship between the explanatory and response variables, leading to poor performance. This can occur when we have small datasets or variables that do not relate to each other.

The figure below shows us examples of both over and underfitting. We plotted the following data

Explanatory	Response
2	1.376
6	2.118
8	2.296
12	2.998
20	3.454

and added linear, quadratic, and fourth degree polynomial trendlines in Excel, noting the value of R^2 for each [2].



Notice that the fourth order trendline has the highest value of R^2 because the curve attempts to fit every single data point, some of which are likely due to random noise. This curve is unlikely to perform well when making predictions because it is **overfit**. Notice also that while the fourth order polynomial manages to fit every data point almost perfectly, the long term behavior of the function does not match the overall trend of the data. The data seems to increase toward an asymptote, but the quartic decreases dramatically after the final observation. The linear model is also likely to perform poorly because it is **underfit**, it is too simple and does not appear to follow the overall trend of the data. Note that, similary to the fourth order polynomial, the long term behavior of the linear model does not match the overall trend of the data. The quadratic model appears to follow the general shape of the data without over or underfit [2].

3.1 Data Partitioning

We can use data partitioning to help identify if our model is over or underfit. We train the model using data from the training set. Then we use data points in our test set to determine how well our model performs. This works because we know the results for the response variable in our test set and can compare to the model prediction [2]. If our model performs well on the training set, but very poorly on the test set, it is likely overfit. If the model performs poorly on the training set, it is likely underfit.

4 Ethical Checklist

As we discussed in previous readings, we must consider data and model validity, and clearly communicate our polynomial models. Below are some questions to think through as you consider whether your polynomial model is ethical.

- **Data validity**. Where did the data come from? How was it collected? How much did you need to clean the data? What are the possible implications about this data set?
- **Model validity**. Examine your modeling decisions.
  - Did you make modeling decisions while cleaning or exploring your data? How did you identify your explanatory and response variables, are they the best variables to choose? Does your model answer the question you are trying to find?
  - Is your predictive model accurate? How did you evaluate your model? How does your model perform on unseen data? Does your choice of polynomial model make sense in the context of the problem?
  - Did you identify possible cases of over or underfitting as you worked through the modeling process? How did you determine the model was overfit? Did you partition data into training and test sets? How did you partition the data, what assumptions did you make when partitioning?
- **Communication**. Is your visualization clear? Is it misleading, or does it convey an honest representation of the data? Did you thoroughly communicate each of your assumptions and modeling decisions? Did you communicate the limitations of your model?

References

[1] David Cohen, Theodore Lee, and David Sklar. *Precalculus*. Brooks/Cole, 2012.

[2] Frederick Hillier et al. *MA103 Mathematical Modeling: Introduction to Management, Science, & Business Analytics with Connect*. McGraw-Hill, 2024.

[3] James Stewart, Daniel Clegg, and Saleem Watson. *Calculus: Early Transcendentals*. Cengage, 2020.