Polynomial Regression Analysis: Exploring the Impact of Regularization

Raghda Al Taei

2024

1 Introduction

In the realm of data analysis and machine learning, understanding the relationships between variables is paramount. One effective way to model these relationships is through polynomial regression, which allows us to capture non-linear patterns in data. This article explores a dataset of 25 one-dimensional points, investigating how varying regularization parameters impact the fit of polynomial models of degrees 2, 3, and 7.

2 The Problem

The primary goal of this analysis is to predict the output variable y based on the input variable x. Given that the relationship between x and y may not be strictly linear, polynomial regression serves as a powerful tool. However, overfitting can occur, especially with higher-degree polynomials. To combat this, we introduce regularization through a parameter λ, which helps to control the complexity of the model and improve its generalization to unseen data.

3 Steps Taken

3.1 Loading the Data

We began by loading our dataset, which consists of two rows—one for x values and another for y values. The data was structured such that we could easily transpose it into column vectors for further analysis.

```
myVar = load('Data.mat');
data = myVar.data;
x = data(1,:)'; % Transpose to make it a column vector
y = data(2,:)'; % Transpose to make it a column vector
```

3.2 Defining Lambda Values

To analyze the effects of regularization, we defined three values for λ: 0, 10, and 100. These values help us observe how regularization influences the fit of polynomial models.

```
lambda = [0, 10, 100];
```

3.3 Preparing Design Matrices

We constructed design matrices for polynomial degrees 2, 3, and 7. Each design matrix included a column of ones to account for the bias term, followed by the powers of x corresponding to the degree of the polynomial.

```
X_poly2 = [ones(size(x)), x, x.^2];          % Degree 2
X_poly3 = [ones(size(x)), x, x.^2, x.^3];   % Degree 3
X_poly7 = [ones(size(x)), x, x.^2, x.^3, x.^4, x.^5, x.^6, x.^7]; % Degree 7
```

3.4 Calculating Theta

Using the normal equation, we calculated the coefficients θ for each degree of polynomial and each value of λ. The regularization term λ was incorporated into the normal equation to prevent overfitting.

```
theta_poly2{i} = (X_poly2' * X_poly2 + lambda(i) * eye(3)) \ (X_poly2' * y);
```

3.5 Generating Predictions

We created a set of predictions for a range of x values by evaluating the polynomial models fitted with different λ values.

```
y_pred_poly2 = cellfun(@(theta) [ones(size(x_pred)), x_pred, x_pred.^2] * theta, theta_poly2
```

3.6 Plotting the Results

Finally, we generated plots for each polynomial degree and regularization parameter combination. The original data points were shown alongside the fitted polynomial curves, providing a visual representation of the model fits.

```
plot(x, y, 'bo'); % Original data
plot(x_pred, y_pred_poly2{i}, 'r-');
```

4 Results

The resulting plots provide insight into how different polynomial degrees and regularization values affect the fit:

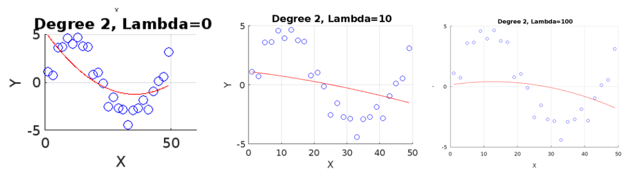

Figure 1: Degree 2 Results

- **Degree 2**: With $\lambda = 0$, the model captures the general trend without overfitting. As λ increases, the fit becomes more conservative, smoothing out the curve.
- **Degree 3**: Similar to degree 2, the inclusion of a cubic term allows for a better fit, but with higher λ values, the model begins to lose flexibility, which is evident in the plots.

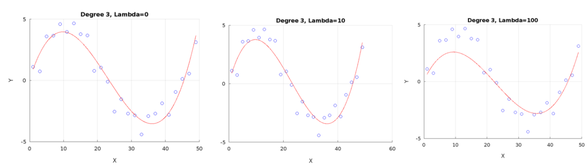

Figure 2: Degree 3 Results

- **Degree 7**: This model demonstrates the most complexity, capturing intricate patterns in the data. However, at $\lambda = 100$, we observe significant over-regularization, resulting in a much flatter curve that fails to capture the data's underlying trends.

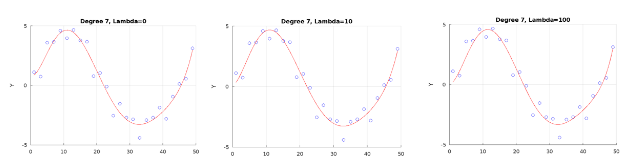

Figure 3: Degree 7 Results

5 Conclusion

Through this analysis, we illustrated the importance of regularization in polynomial regression, showcasing how varying λ values influence model complexity and fit. This exploration reinforces the notion that while higher-degree polynomials can provide better fits to the training data, regularization is crucial for ensuring that models generalize well to new data.

For further details, including the complete code and analysis, please visit my GitHub Repository.