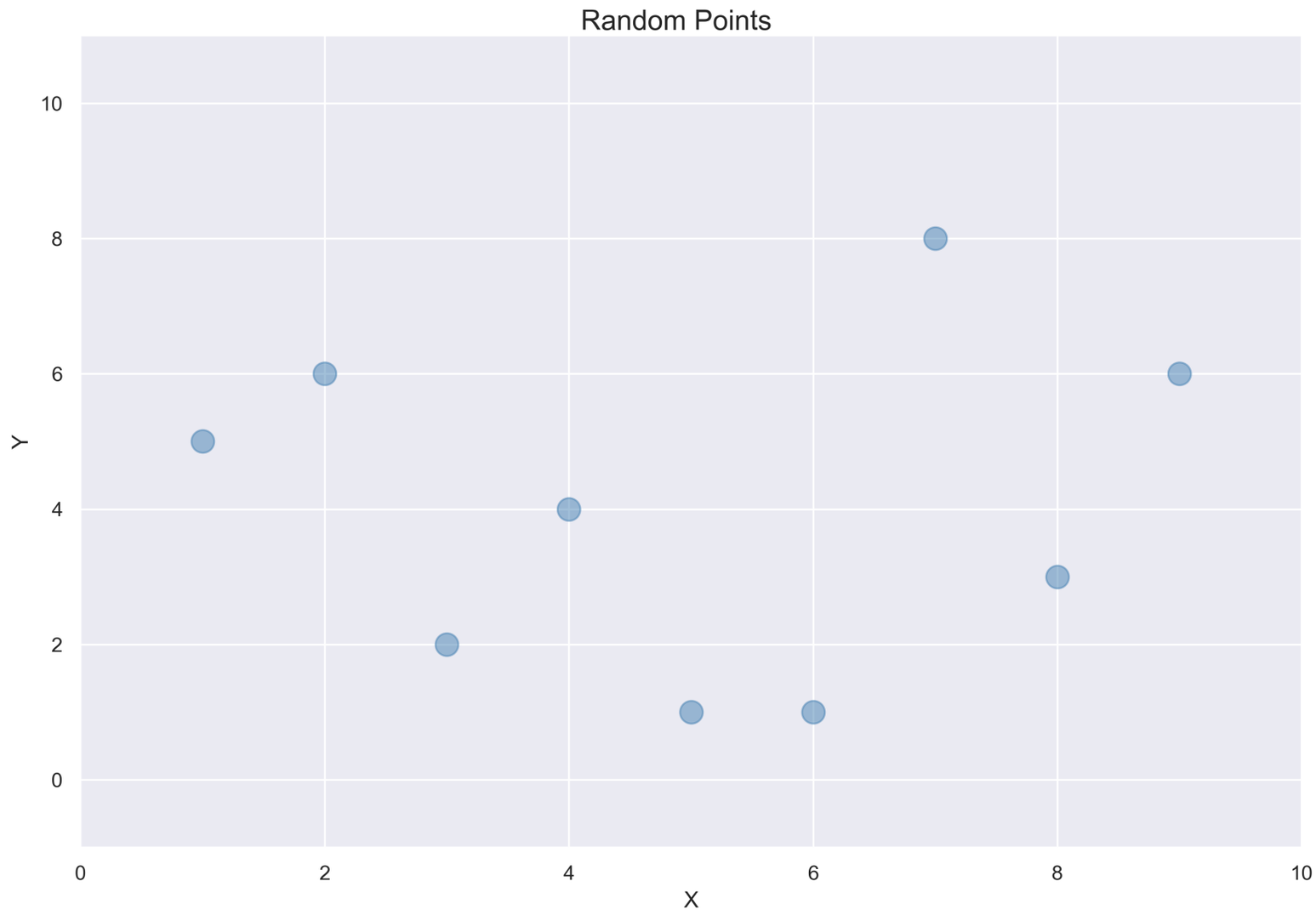
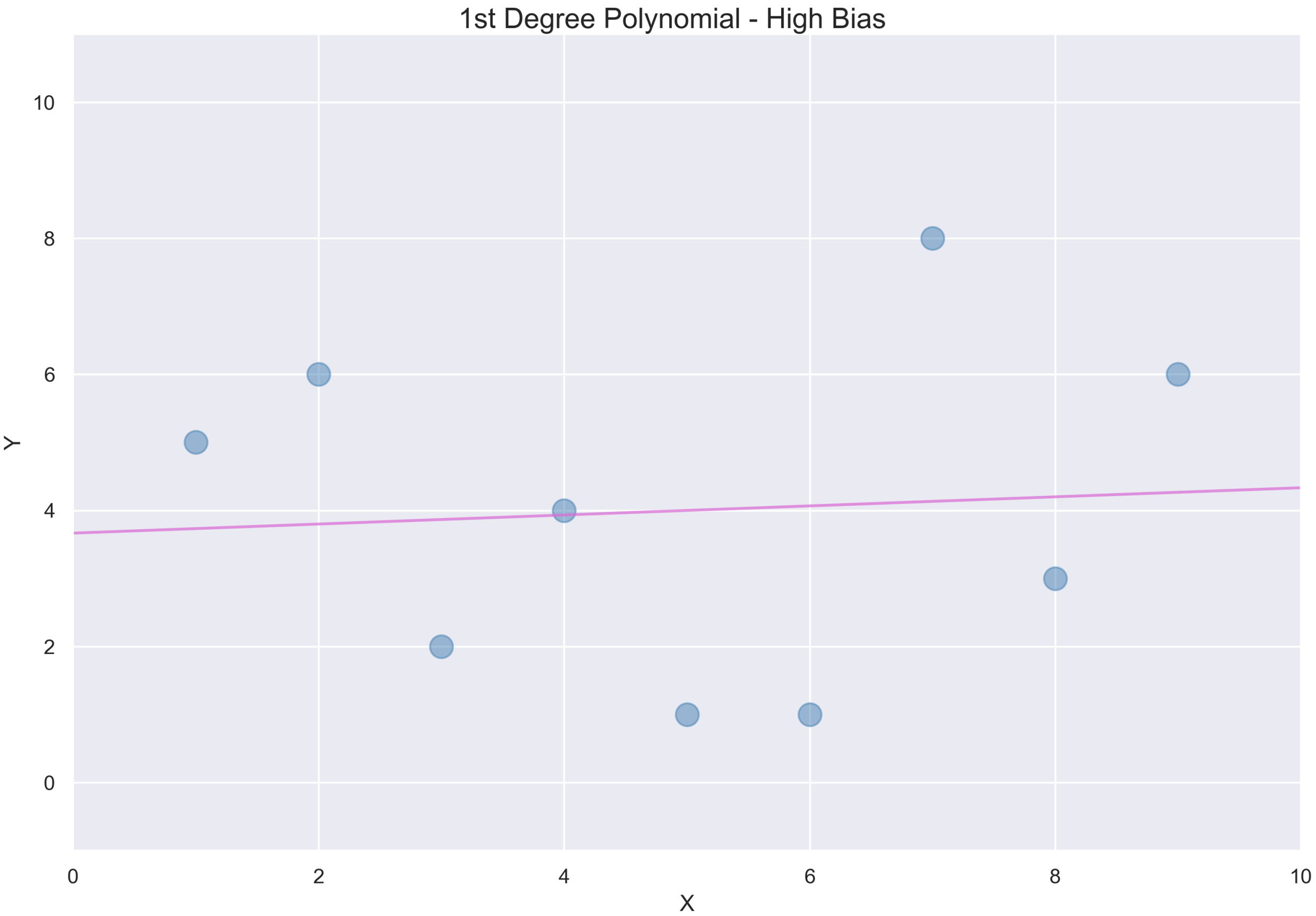
OVERFITTING THE DATA


Random Points
10
8
6
4
2
0
Y
0
2
4
6
8
10
X

POLYNOMIALS

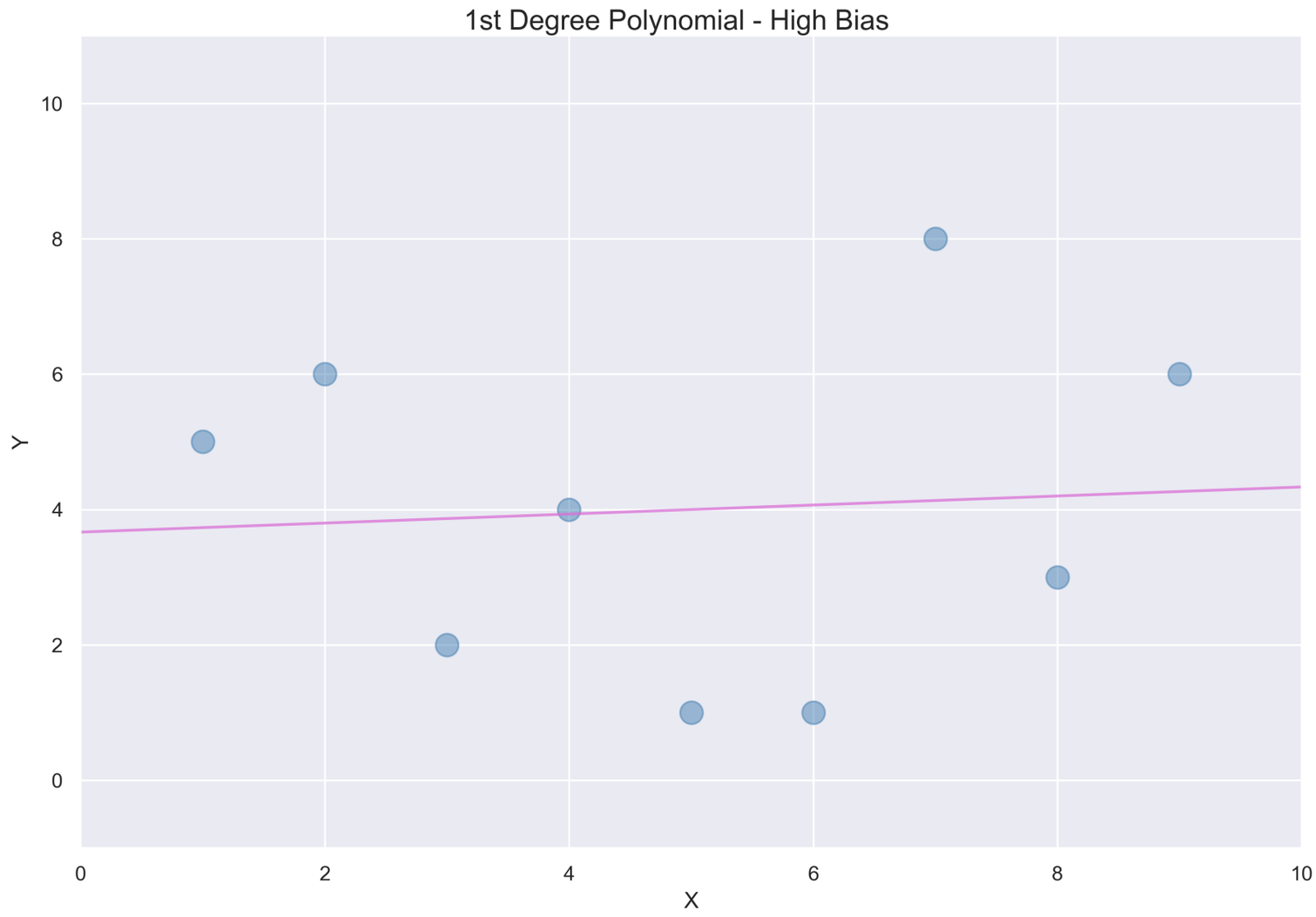
- What degree of polynomial will fit the data best?

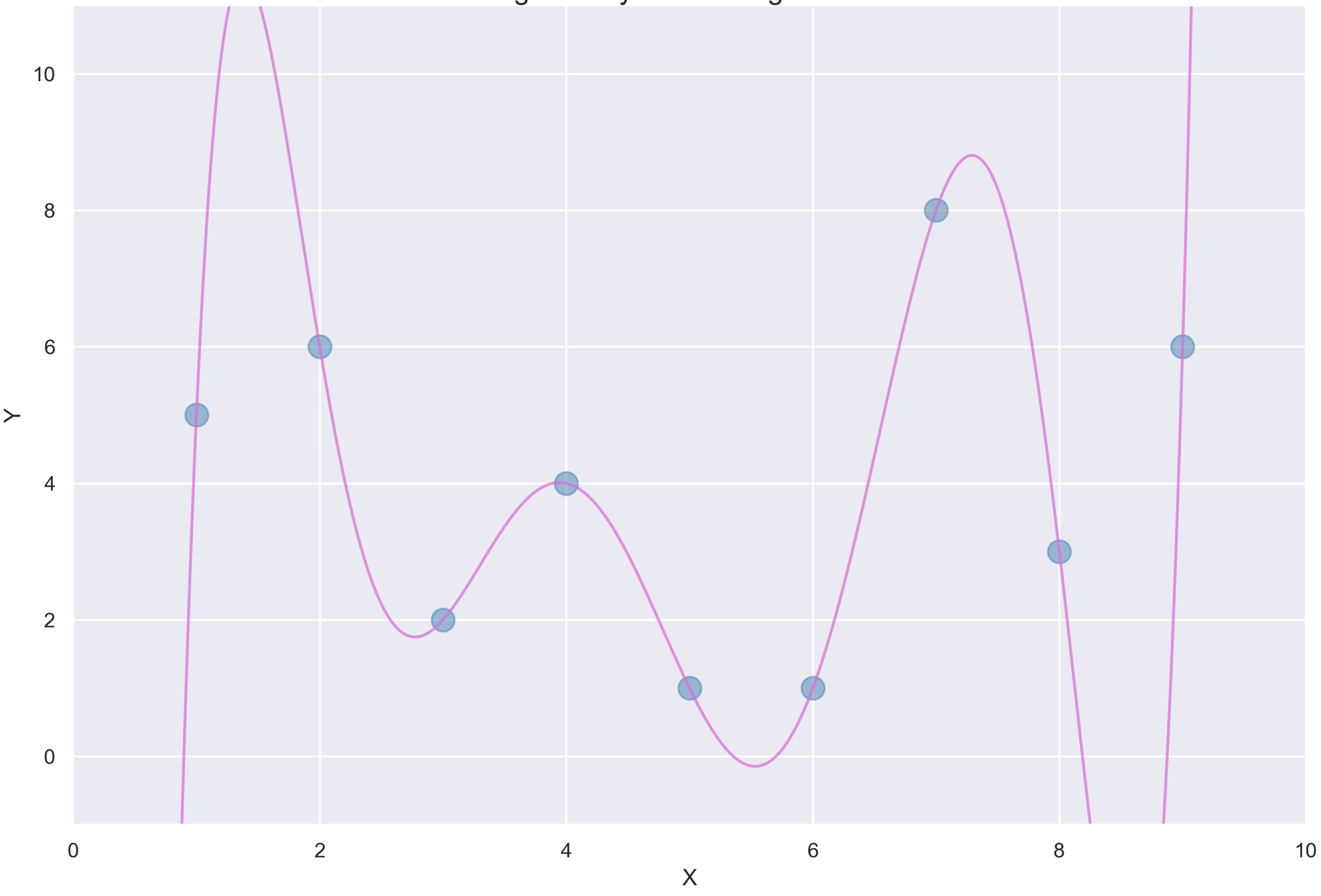



1st Degree Polynomial - High Bias
10
8
6
4
2
0
Y
0
2
4
6
8
10
X

UNDERFITTING

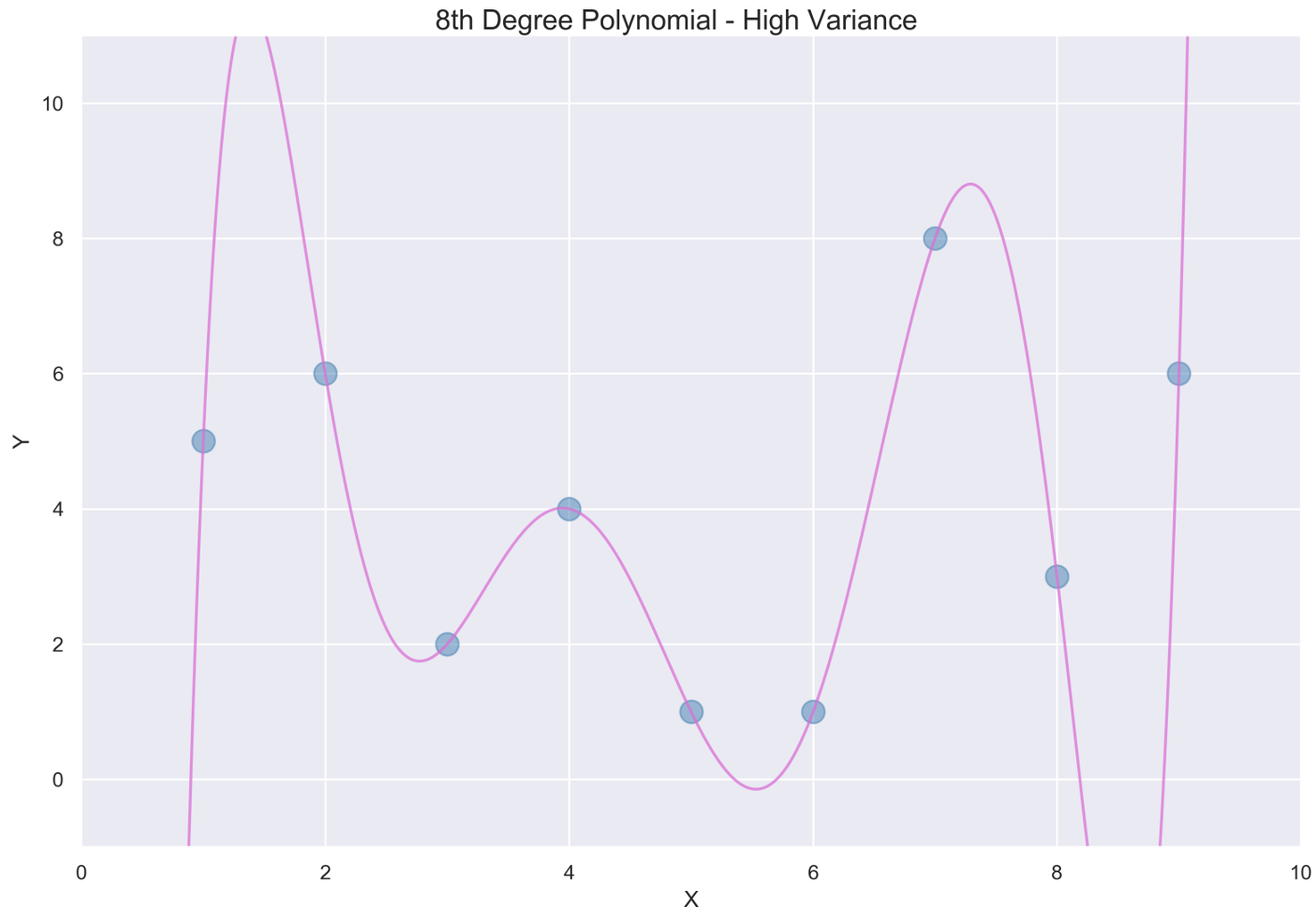
- Regression line fails to pick up on nuances of the data.

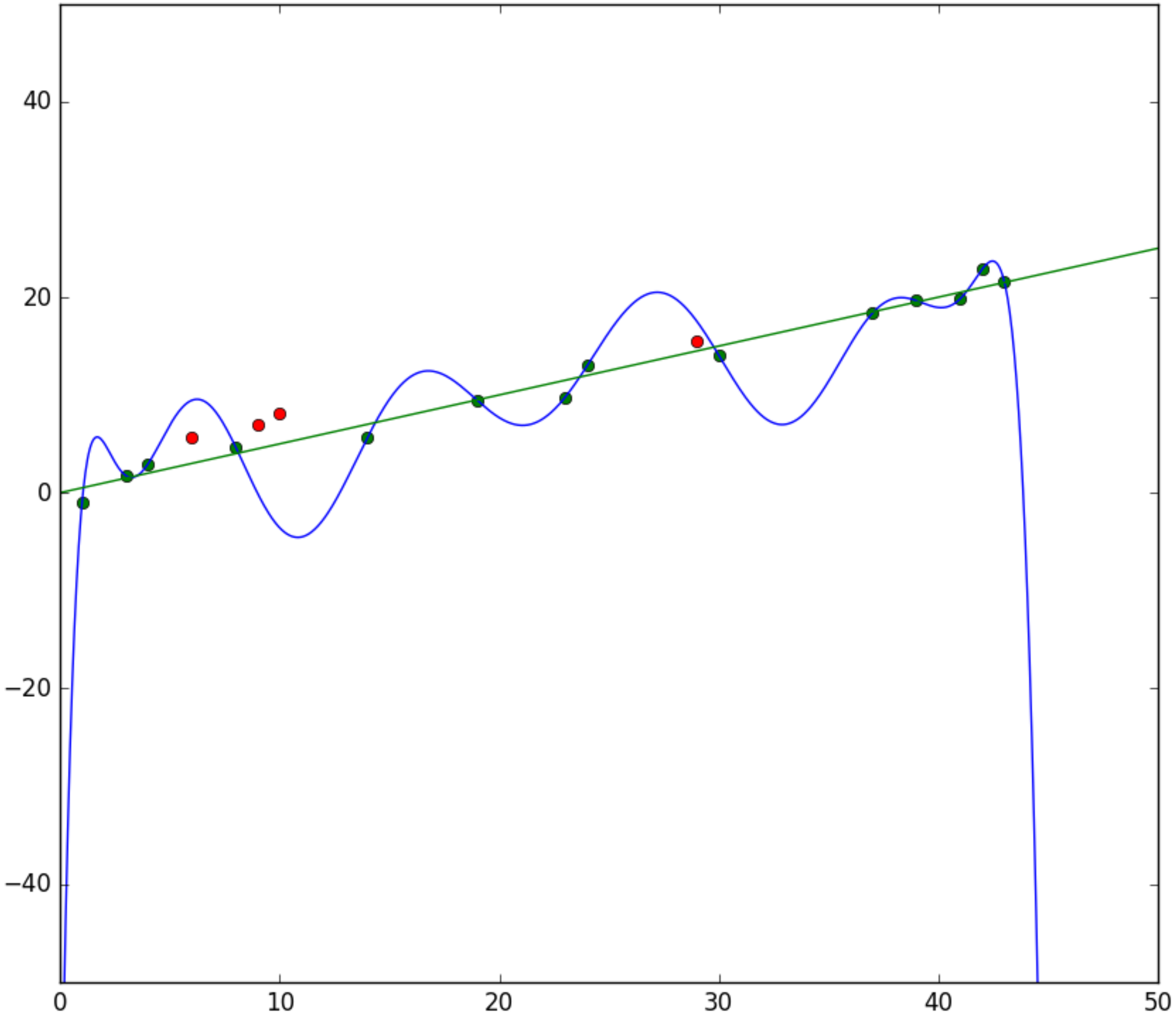


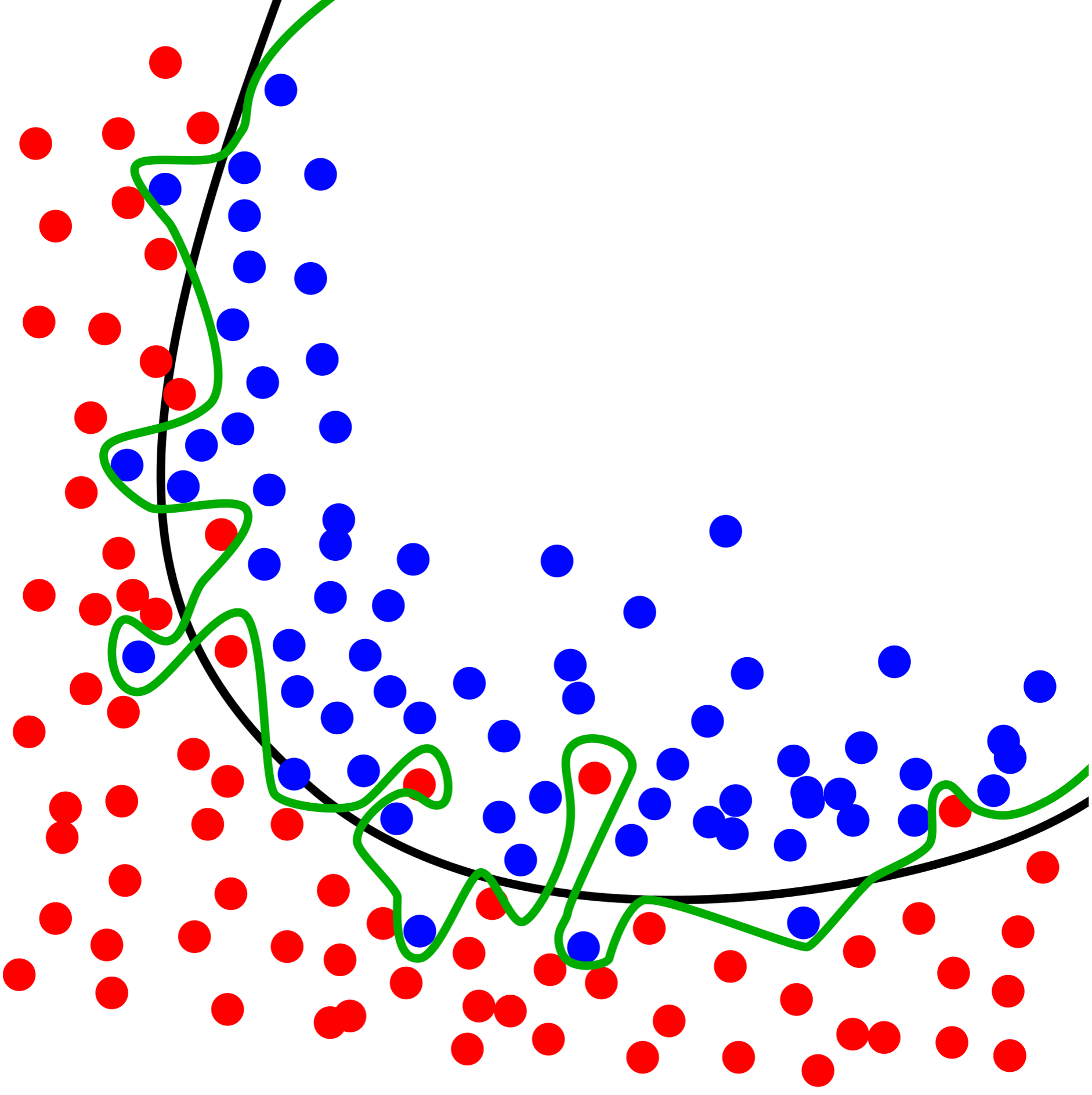

8th Degree Polynomial - High Variance
Y
10
8
6
4
2
0
0
2
4
6
8
10
X

OVERFITTING

- Polynomial generalizes poorly to new data.
- When x=1.5, y=11.
- When x=8.5, y=-3.
- When x=10, y=250.




40
20
0
−20
−40
0
10
20
30
40
50



OVERFITTING SOLUTIONS

- Use cross-validation to vary test sets.
- Use models with regularization built-in (XGBoost).
- Adjust hyperparameters to reduce overfitting.

HAPPY CODING!