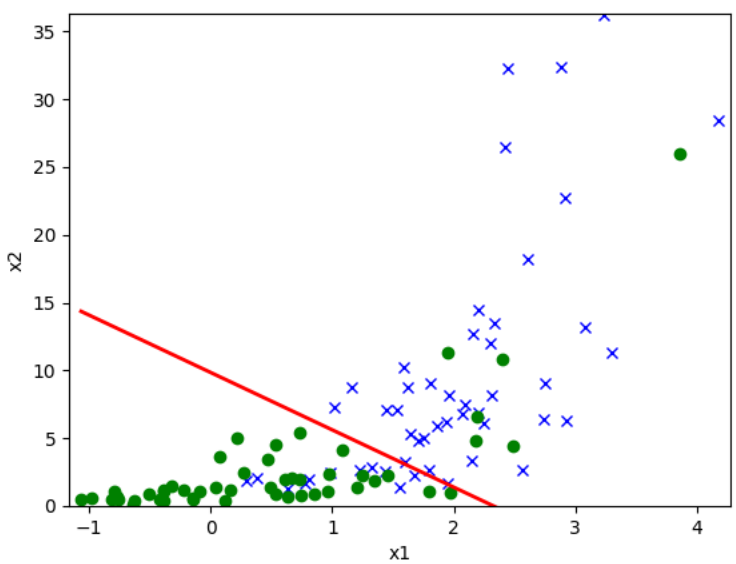
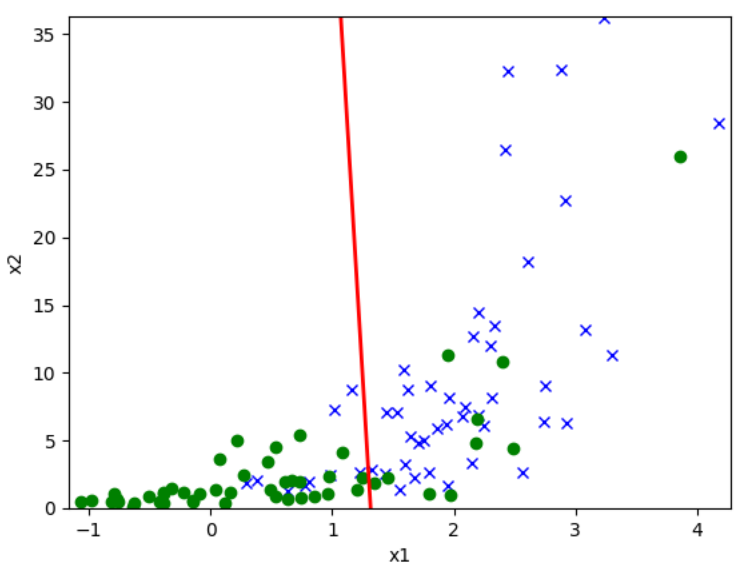
Stanford CS229 Machine Learning Problem Set #1 Coding Problems

1. Linear Classifiers (Logistic Regression and GDA)

1-b. Logistic Regression and separating hyperplane for Dataset 1

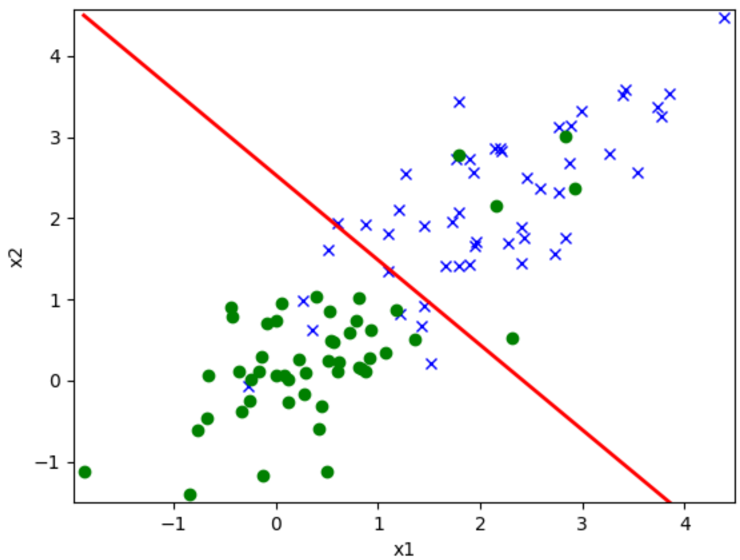


1-e. GDA model and separating hyperplane for Dataset 1

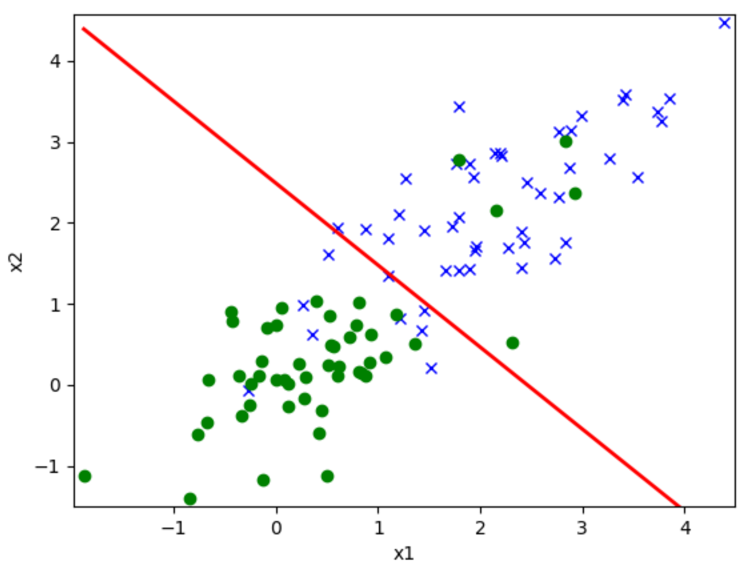


1-f. Compare results of Logistic Regression and GDA model on Dataset 1
The separating boundary for Logistic Regression seems more reasonable than for GDA.
For Dataset 1, Accuracy (Logistic Regression) = 0.83 and Accuracy (GDA) = 0.81, so the accuracy for Logistic Regression is also higher than GDA.

1-g. Logistic Regression and GDA on Dataset 2
Logistic Regression separating hyperplane on Dataset 2.



GDA model separating hyperplane on Dataset 2

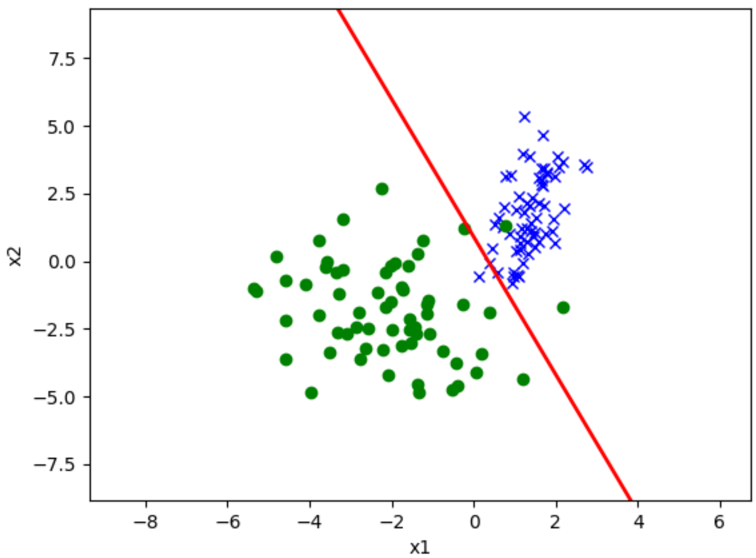


Results of Logistic Regression and GDA model on Dataset 2 comparison:
For Dataset 2, Accuracy (Logistic Regression) = 0.86 and Accuracy (GDA) = 0.86, so GDA does not perform worse than Logistic Regression.
For Dataset 1, X might not follow Gaussian Distribution, causing GDA to perform worse than Logistic Regression, which is a more robust model. For Dataset 2, X follows Gaussian Distribution, so it performs as good as Logistic Regression.

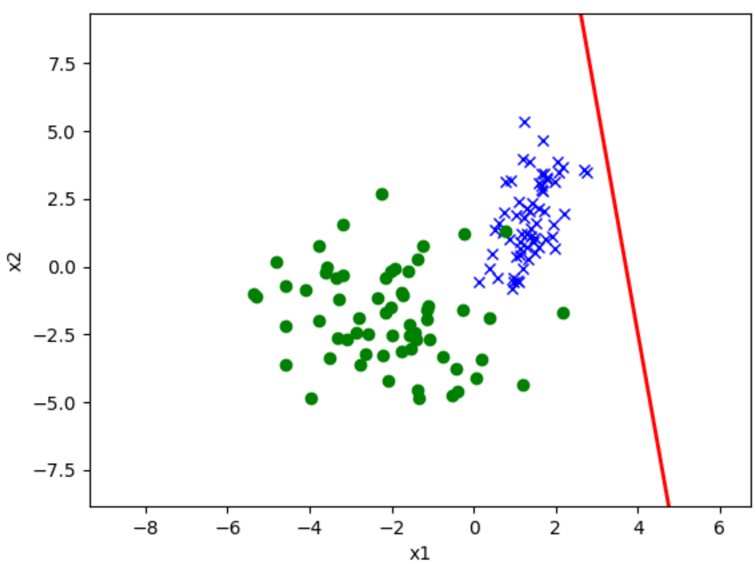
1-h. Possible transformation on x_2
A possible transformation is to set $x_2 \coloneqq log x_2$, which makes x_2 follow Gaussian Distribution.

2. Incomplete, Positive-Only Labels

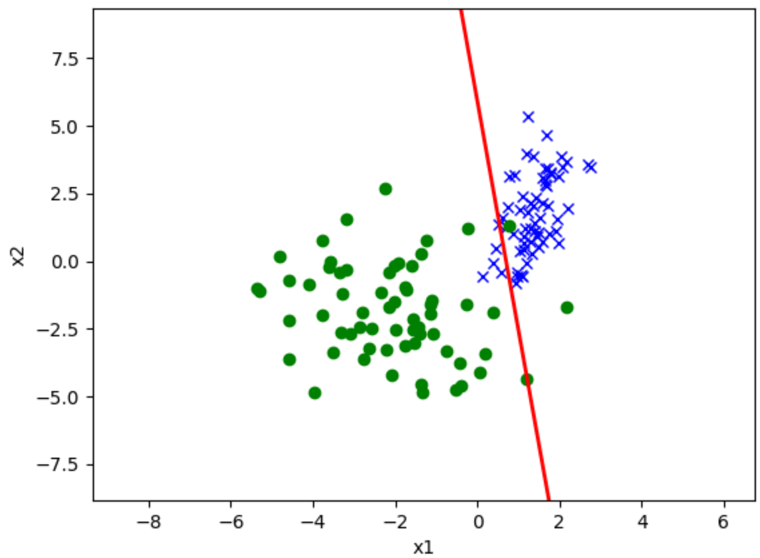
2-a. A plot to visualize the test set with x_1 on the horizontal axis and x_2 on the vertical axis. Use different symbols for examples $x^{(i)}$ with true labels $t^{(i)} = 1$ than those with $t^{(i)} = 0$. And plot the decision boundary obtained by Logistic Regression trained with t-labels.



2-b. A plot to visualize the test set with x_1 on the horizontal axis and x_2 on the vertical axis. Use different symbols for examples $x^{(i)}$ with true labels $t^{(i)} = 1$ than those with $t^{(i)} = 0$. And plot the decision boundary obtained Logistic Regression trained with $y^{(i)}$ and predicted naively with the parameter obtained.

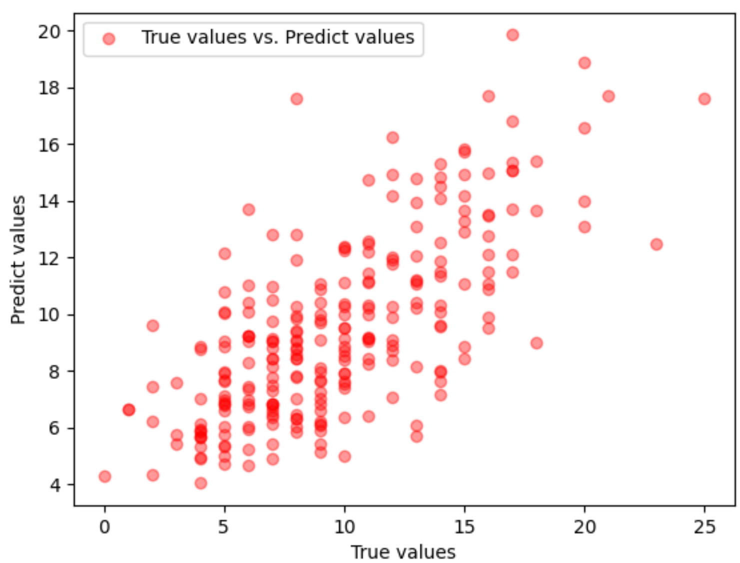


2-f. A plot to visualize the test set with x_1 on the horizontal axis and x_2 on the vertical axis. Use different symbols for examples $x^{(i)}$ with true labels $t^{(i)} = 1$ than those with $t^{(i)} = 0$. Estimate an $\alpha = \frac{1}{|V_+|}\Sigma_{x^{(i)}\epsilon V_+} h(x^{(i)})$. And predict the validation sample with Logistic Regression and scaled by this α. Plot the decision boundary obtained the adjusted Logistic Regression.



3. Poisson Regression

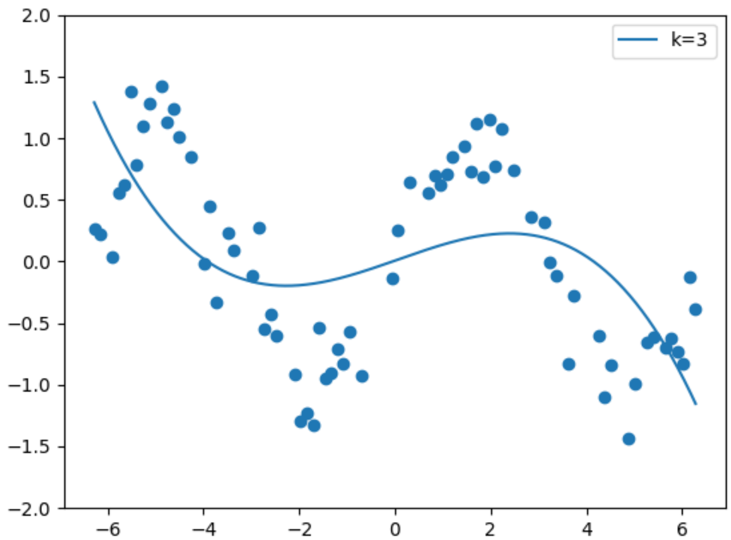
3-d. A scatter plot between the true counts vs predicted counts on the validation set.



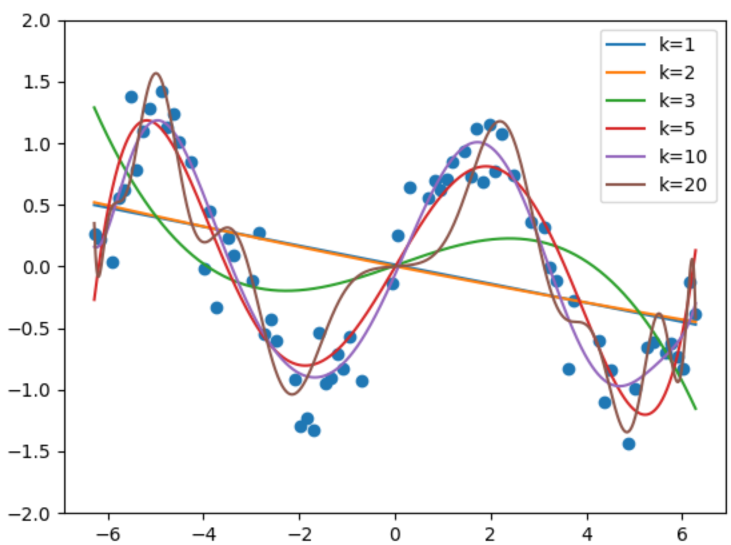
5. Linear regression: linear in what?

5-b. Degree-3 polynomial regression

A scatter plot of the training data, and the learnt hypothesis as a smooth curve over it

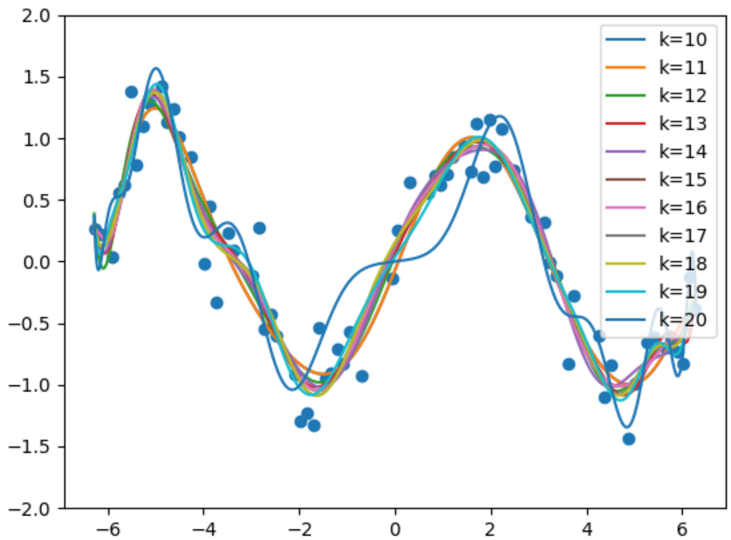


5-c. Degree-k polynomial regression
A scatter plot of the training data, and different learnt hypothesis curves for each value of k. k = 1, 2, 3, 5, 10, 20. In 5-c, $\phi(x) = [1, x, x^2, \ldots, x^k]^T$.



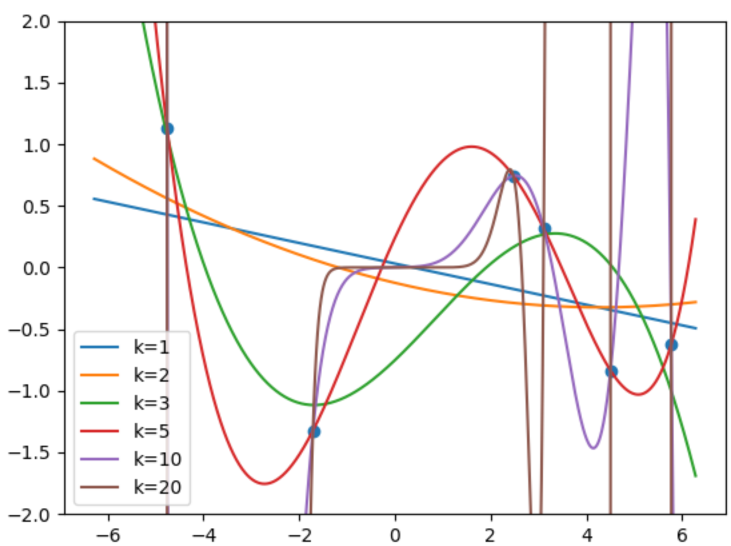
Higher degree of polynomials makes the model fit the data better, however it is also more numerically unstable.

5-d. Degree-k polynomial with sin(x) regression
A scatter plot of the training data, and different learnt hypothesis curves for each value of k. k = 1, 2, 3, 5, 10, 20. In 5-c, $\phi(x) = [1, x, x^2, \ldots, x^k, \sin x]^T$.

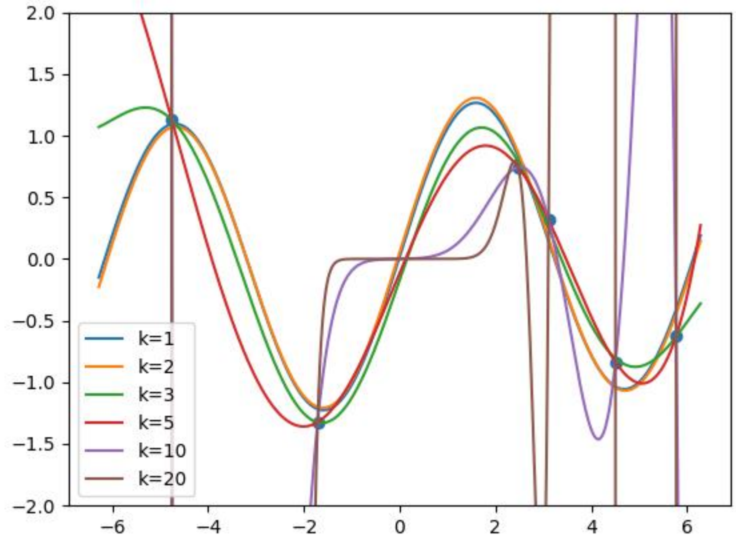


Higher degree of polynomials with an additional sin(x) makes the model fit the data even better, however it still exhibits the same numerically instability.

5-e. A smaller dataset.
Smaller dataset with just degree-k polynomial.



Smaller dataset with just degree-k and sin(x) polynomial.



With a much smaller dataset, although models with higher degree polynomials passes through all points, it still seems like a poor fit. Numerical instability remains a major problem regardless of whether sin(x) exists or not.