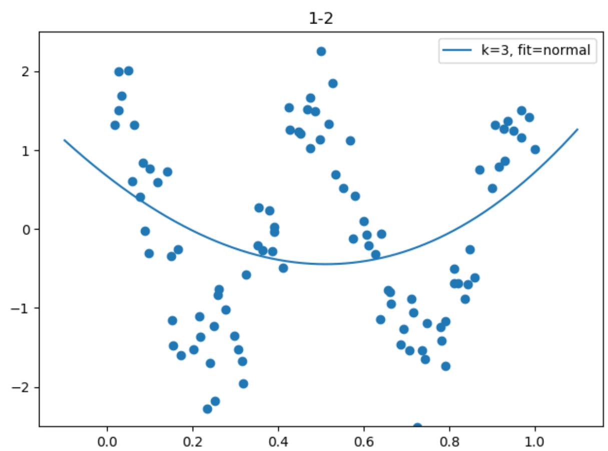
Jeremy Ng - 500882192
CPS 803 - Assignment 1

1. Linear Regression

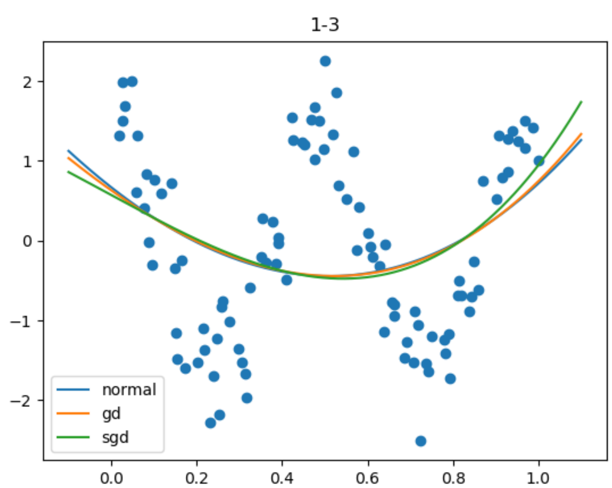
1.1. Objective Function: $J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(h_\theta(\widehat{x}^{(i)}) - y^{(i)})^2$

Update Rule of BGD: $\theta_j := \theta_j - \alpha(h_\theta(\widehat{x}^{(i)}) - y^{(i)})\widehat{x}_j^{(i)}$

1.2. Plotting learned hypothesis with polynomial degree 3.

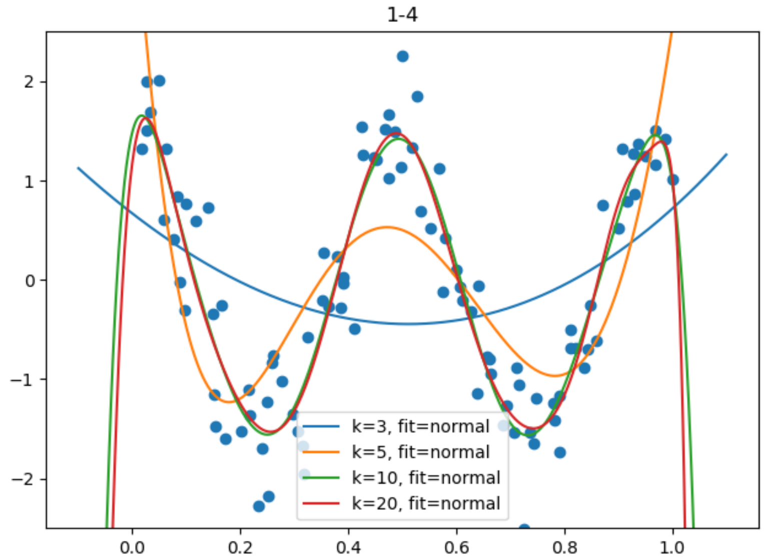


1.3. Using an alpha of 0.01 and degree 3 on various plots, across 10000 iterations; we can see that the GD and SGD fit get close to the normal.

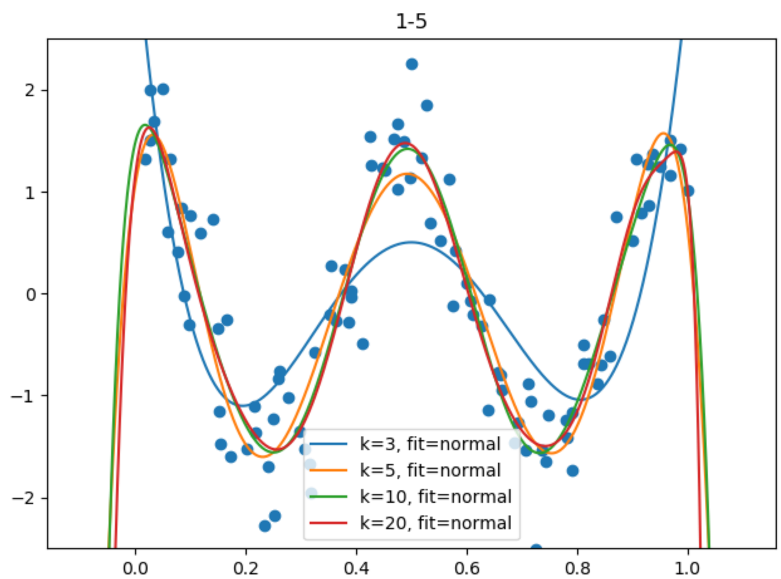


Jeremy Ng - 500882192
CPS 803 - Assignment 1

1.4. As the polynomial degree increases, the fit becomes more accurate.

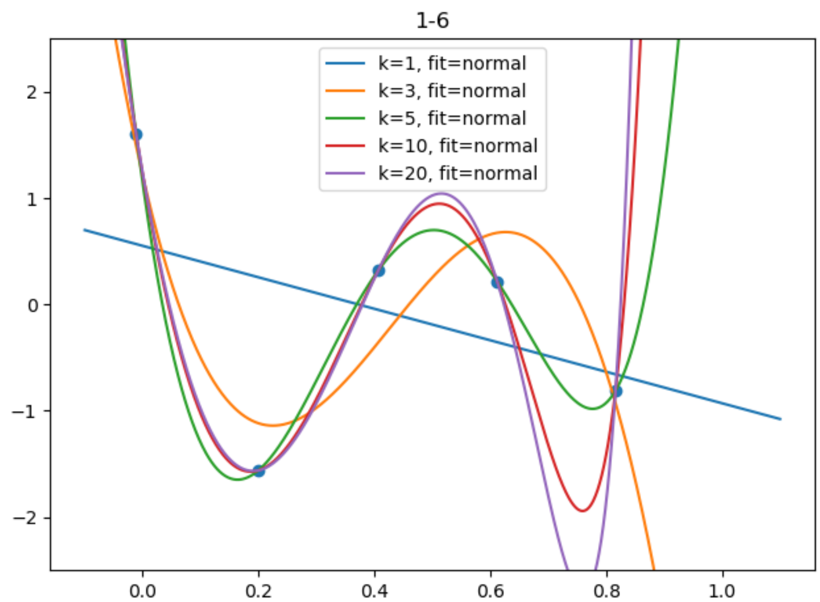


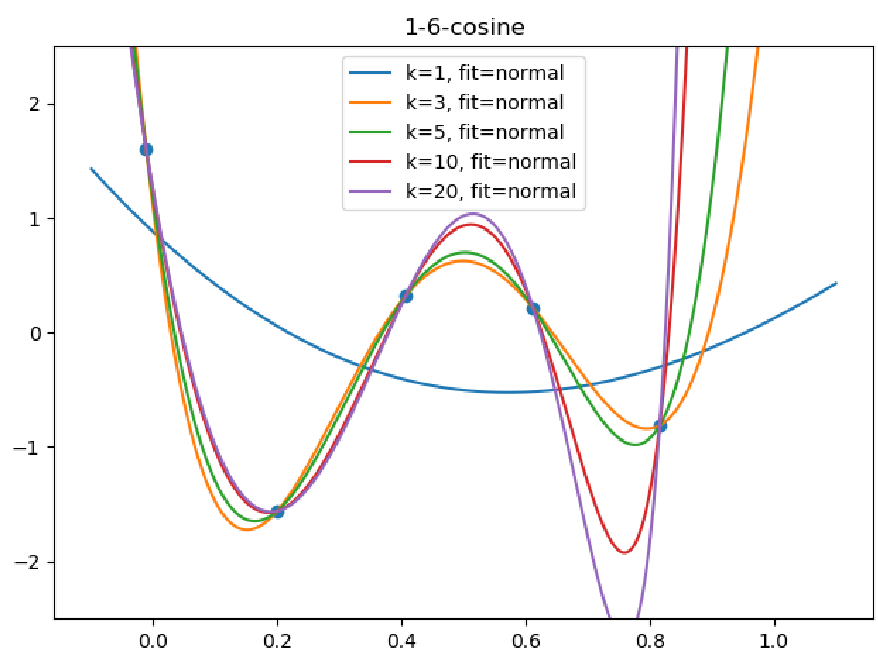
1.5. Compared to graphs 1-4, 1-5 has a more accurate starting point. If we look at k = 3, it is similar to 1-4 at k = 5. Overall I think this would save on computation time.



Jeremy Ng - 500882192
CPS 803 - Assignment 1

1.6. Below are the various hypothesis curves along with the cosine feature. We can see that a polynomial degree of 10 or 20 leads to a large curve in the range 0.6 <= y <= 0.8. This part of the curve represents data that is outside of our dataset and may cause problems for any new data that we try to add. This is overfitting.

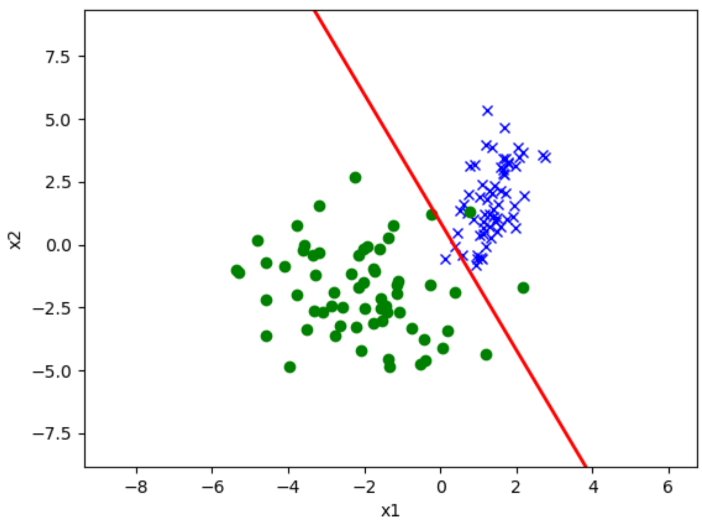




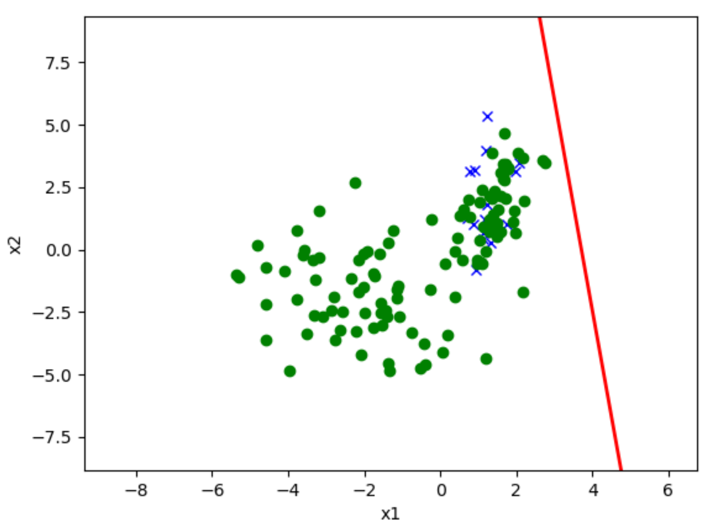
Jeremy Ng - 500882192
CPS 803 - Assignment 1

2. Incomplete, Positive-Only Labels

2.1. Using true labels, we can see that the decision boundary accurately separates the data.



2.2. Using partial labels, the model does not accurately separate the data by failing to predict the probability of interest.



Jeremy Ng - 500882192
CPS 803 - Assignment 1

2.3. $p(t^{(i)} = 1 \mid y^{(i)} = 1,\ x^{(i)}) = 1$

Let:

$T = (t^{(i)} = 1)$

$Y = (y^{(i)} = 1,\ x^{(i)} = 1)$

Identity:

$P(A) = P(A|B) * P(B) + P(A|\neg B) * P(B)$

Therefore:

$P(Y)$

$= P(y^{(i)} = 1,\ x^{(i)} = 1 \mid t^{(i)} = 1) * P(t^{(i)} = 1) + P(y^{(i)} = 1,\ x^{(i)} = 1 \mid t^{(i)} = 0) * P(\ t^{(i)} = 1)$

$$P(Y|\neg T) = P(y^{(i)} = 1,\ x^{(i)} = 1 \mid t^{(i)} = 0) = 0$$

Using Bayes' Theorem:

$P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$

We have:

$P(T|Y)$

$= \frac{P(y^{(i)}=1, x^{(i)}=1 \mid t^{(i)}=1)*P(t^{(i)}=1)}{(y^{(i)}=1, x^{(i)}=1)}$

$= \frac{P(y^{(i)}=1, x^{(i)}=1 \mid t^{(i)}=1)*P(t^{(i)}=1)}{P(y^{(i)}=1, x^{(i)}=1 \mid t^{(i)}=1)*P(t^{(i)}=1)+P(y^{(i)}=1, x^{(i)}=1 \mid t^{(i)}=0)*P(\ t^{(i)}=1)}$

$= \frac{P(y^{(i)}=1, x^{(i)}=1 \mid t^{(i)}=1)*P(t^{(i)}=1)}{P(y^{(i)}=1, x^{(i)}=1 \mid t^{(i)}=1)*P(t^{(i)}=1)+0}$

$= 1$

2.4. CP8318 Only Question

2.5. CP8318 Only Question

Jeremy Ng - 500882192
CPS 803 - Assignment 1

2.6. With correction and a validation test set, the model can accurately separate the data.