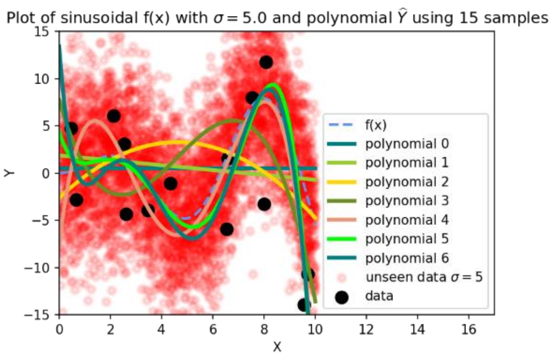
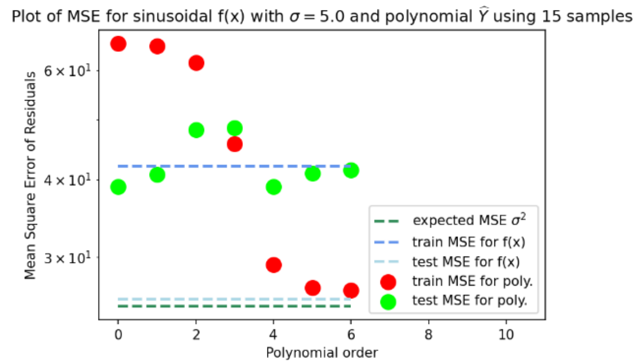
COMS 474/574 -- SP2022 -- HW 3 solutions

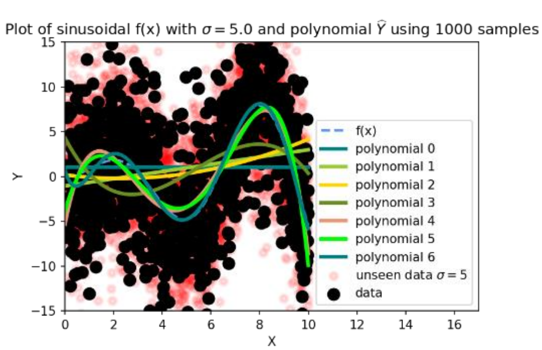
1. Sinusoidal f(X)

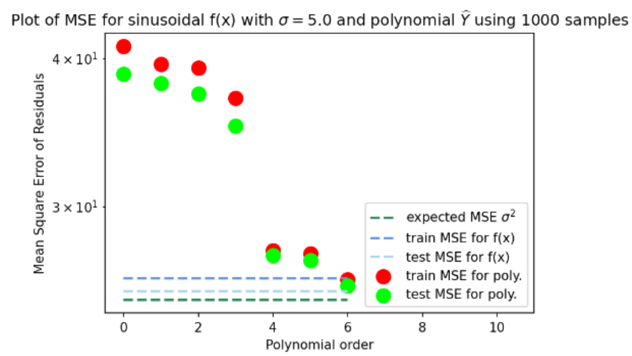
(a).i. n=15




(a).ii. N=1000




**N=15:** We begin by noting that in this example for n=15 samples, the training MSE is above the test MSE. While we would not expect this, we do see that even the training set MSE of f(X) is quite high, so the empirical variance of those 15 samples (about 40) is greater than the expected variance (25). Low order polynomials have high training and test MSE, and from the scatter plot clearly do not match the trend f(X), indicating under-fitting. We see the training error is driven low by increasing polynomial order. Even for n=15 it is above the expected noise variance (25), though that is likely due to the unusually large empirical variance of the noises in the n=15 data set (your results may have training MSE below 25 for n=15).' Nonetheless, the higher-order polynomials have training MSE well below that of f(X), while their corresponding test MSE is high, indicating over-fitting

In both cases (n=15 and n=1000), we see the training error is driven low by increasing polynomial order. Even for n=15 it is above the expected noise variance (25), though that is likely due to the unusually large empirical variance of the noises in the n=15 data set (your results may have training MSE below 25 for n=15).'

**N=1000**: For test MSE, with n=1000 there is clear improvement, and the test MSE remains very close to training MSE, so there is no evidence of over-fitting. Low order polynomials have high training and test

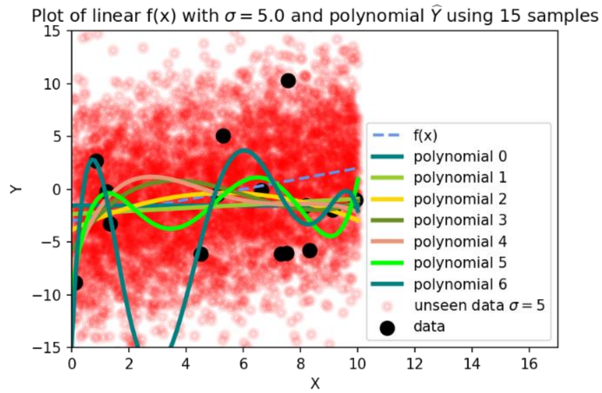
MSE compared to higher order polynomials for n=1000, indicating under-fitting. From the plots of the fitted polynomials, it is also clear that the lower-order polynomials do not come close to matching the curvature of f(X). Even for polynomial order 6, the training MSE and test MSE with n=1000 were close to each other and close to the expected noise variance of 25, indicating a good fit.

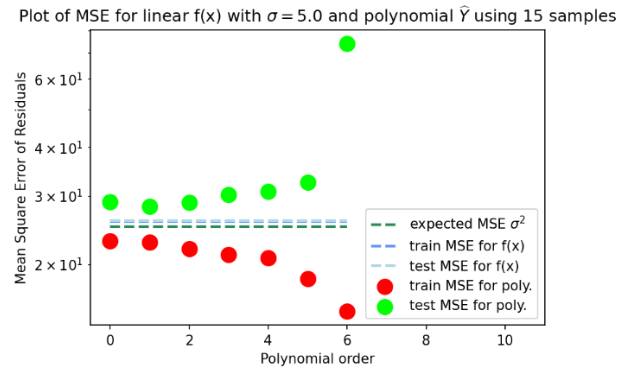
**Effects of the number of samples**: We saw that for both n=15 and n=1000, under-fitting was present. The test MSE of the constant and linear functions only slightly improved with more samples. Thus, having many more samples did not mitigate under-fitting.

For higher order polynomials, we saw that while they appeared to over-fit with n=15, there was no significant over-fitting for n=1000. Thus, having more samples helped mitigate over-fitting.

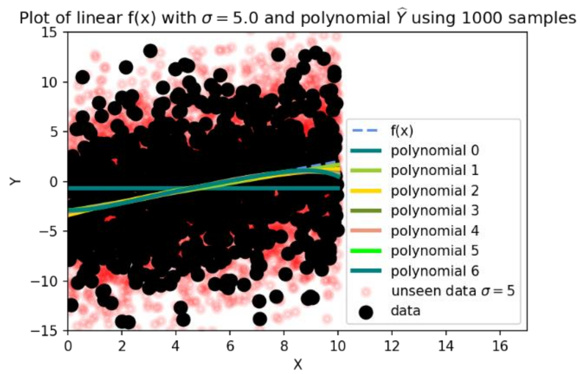
(b).

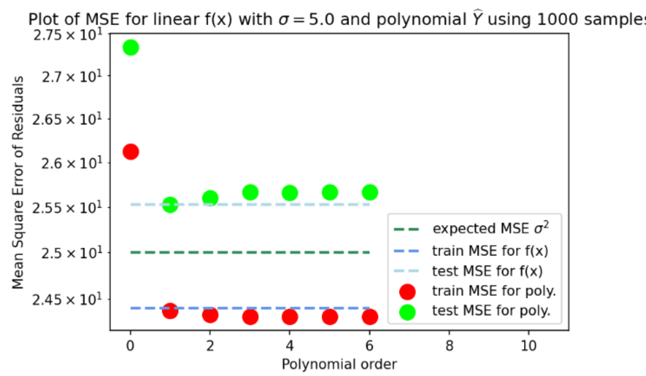
N=15




N=1000




**<u>N=15:</u>** The constant and linear models have similar training and test MSEs, with training MSE below the noise variance of 25, suggesting slight over-fitting. As the polynomial order increases, the gap between training and test MSE's increases, with training MSE on a downward trend, <u>so overfitting is clearly occurring</u> for high order polynomials and even to a minor extent for low order polynomials. We also know that for p=0 (constant model), it is a less flexible model class than f(X) and can see that visually from the scatter plot, so <u>under-fitting is occurring as well for p=0</u>.

**<u>N=1000</u>**: For test MSE, with n=1000 p=1 (linear model) has the best fit, with higher order models only slightly worse, and for p>0, all the fit models have test MSE almost identical to that of the trend f(X), suggesting that <u>there is no significant over-fitting</u>. The scatter plot also shows even the higher order polynomials are almost linear. The constant model (p=0) has higher test and training MSE than others, and scatter plot also shows it does not match f(X), <u>so underfitting is occurring</u>.

**Effects of the number of samples**: We saw that for both n=15 and n=1000, under-fitting was present for the constant model (p=0). The test MSE of the constant and function only slightly improved with more samples. Thus, having many more samples did not mitigate under-fitting.

For higher order polynomials, we saw that while they significantly over-fit with n=15 samples, there was no significant over-fitting for n=1000. Thus having more samples mitigated over-fitting.

(c). **Similarities and differences between (a) and (b):**

For these experiments, we saw over-fitting was more dramatic when f(X) was linear than when f(X) was sinusoidal with n=15 – specifically that even lower order models did better than the true linear f(X) while only the higher order models out-performed the true sinusoidal f(X). That may be due to the fluctuation in Y values for the training set being almost entirely due to noise when f(X) was linear compared to when f(X) was sinusoidal. Though it may not be a systematic result (i.e. it may depend on the specific X values and noise values in each training sample and the particular f(X)). For both types of f(X), having n=1000 mitigated over-fitting.

For both, under-fitting was present and increased data did not mitigate under-fitting. However, having more data removed the effects of over-fitting, so the poor performance of low order polynomials for n=1000 was largely due to just underfitting. Which polynomials were clearly underfitting varied between f(X) linear (for which just p=0 was underfitting) and f(X) sinusoidal ( esp. p<4 but even p=4 and p=5).

Grading

(a) 2 total for all plots, 8 for discussion

(b) “

(c) 6