Com S 474 Homework 3

1 . Sinusoidal

Chart, scatter chart

Description automatically generatedChart, scatter chart, bubble chart

Description automatically generatedi)Chart

Description automatically generated

ii)Chart, diagram

Description automatically generated

a) Over-fitting can be seen in occurring as most features are added the models tend to follow the trend of the past data and not the underlying trend. The extent of which over-fitting occurs is due to the number of samples. With more samples the models with more features do well

1 Linear

Chart, scatter chart, bubble chart

Description automatically generatedi)Chart, scatter chart

Description automatically generated

Chart

Description automatically generatedii)Chart

Description automatically generated

b) Under-fitting can be seen in the models with more polynomials, and this is present when the number of samples is small (e.g. 15). This can be seen since we know the underlying trend f(x) is linear but model predictions for what the trend can’t be captured. (e.g. too many peaks and troughs for a linear model)

c) With the sinusoidal models it can be seen that when the number of sample data is low over-fitting is possible as the degrees of polynomials increase whereas with the linear model it was showing under-fitting. There is a difference when the number of sample data was large. With the linear model as more polynomials are added the model still fit closely to the f(x) line, however with the sinusoidal model only the models containing 4 or more polynomials were able to capture the underlying trend f(x). Also, with linear models are good at predicting continuous data (e.g. GDP of the United States over some time) while the sinusoidal model is better at predicting a category (e.g. Will it rain on a given day) since the function oscillates over an interval.