**DATA 311 – Lecture 27 – Occam’s Razor, Overfitting, Data Splits, Cross-Validation**

1. Draw a function (not necessarily a straight line!) y = f(x) that “fits” the given data points for each of the plots below.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

1. Draw an function y = f(x) on the plot below (again, not necessarily a line) that **overfits** the training data depicted below.

Chart, scatter chart

Description automatically generated

1. Using a dashed line in the plot above, draw a function y = f(x) that **underfits** the training data.
2. Consider the “your model” (dotted) lines in the plots below. One of them is high-variance model, and the other is high-bias model. Which is which?

Chart, line chart

Description automatically generated

1. Suppose you split your data into training and validation sets, train a model, and evaluate its accuracy on both the training set and on the validation set. How would you describe each of the four possible scenarios? Focus on the left column first.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Validation Accuracy | |
|  |  | **Bad** | **Good** |
| Training Accuracy | **Bad** |  |  |
| **Good** |  |  |