Problem 1 (4,5)

The gradient algorithm was mentioned with no introduction of background nor pseudo code or equations to describe the concept. However, there was thorough discussion of the effect of step size and tolerance on gradient descent on both convex and non-convex function. The same is true for central difference section; no equation or introduction was given but good analysis and visualization of the effects of step size.

Problem 2 (3,5)

Equations for OLS and the prediction given, but no variables introduced. The basis function matrix shown is miscalculated resulting in discrepancies in results compared to that given in Bishop. For example data shown in Fig. 6 when M=9 for the OLS regression does not agree well at all with the expected values shown in Bishop. Also it is unclear why the author switches to considering M=8 for the second half of the question. The reported disadvantage of using sinusoidal basis functions is that it is “better suited for periodic functions” which does not describe what the disadvantage is for non-periodic functions and more or less just rephrases the question.

Problem 3(4, 4)

Again no description, introduction or equation describing ridge regression is included. Good discussion of the foreseen and calculated effect of the outlier. However there is a lack of discussion on purpose and observed effect of the regularization term. There was a clear connection to cost and a thorough search for to the optimal M and lambda for the real data set. The labels on the graphs in Fig 14 are somewhat confusing at first glance. Also legends on any figure with a mix of training, test and validation data would be helpful for readers.

Problem 4(4,5)

An equation or introduction of LAD would have been helpful. There is some confusion in the optimal parameters, as no error values were given and whether the value of $lambda$ is the same for both M=1 and M=2. Good discussion of the outlier is rejected in LAD.