Homework 1

1   Directions:

- **Due: Thursday February 3, 2022 at 9pm.** Late submissions will be accepted for 24 hours after that time, with a 15% penalty. (the enforcement is strict, beginning at 9:01pm, except for extreme situations; having a poor wifi connection or minor computer problems is not sufficient for the penalty to be waived.)

- Upload the homework to Canvas as a single pdf file.

- If the graders cannot easily read your submission (writing is illegible, image is too dark, or if the contrast is too low), then you might receive a zero or only partial credit.

- Any non-administrative questions must be asked in office hours or (if a brief response is sufficient) Piazza.

2   Problems

Problem 1.   [16 points; 4 each part] For each of parts (a) through (d), indicate whether we would generally expect the performance of fitted models on future data the be better or worse if the model class has many degrees of freedom (e.g. class of high degree polynomials) compared to a model class with few degrees of freedoms (e.g. constant or linear functions). Briefly explain why (1-3 sentences).

(a). The number of samples n is large, and the number of features p is small.

(b). The number of features p is large, and the number of samples n is small.

(c). The relationship between the features $\{X_1, \ldots, X_p\}$ and the response (Y) is highly non-linear.

(d). The variance of the noise terms, i.e. $\sigma^2 = Var(\epsilon)$, is extremely high.

Problem 2.   [18 points total; 6 each part] You will now think of some real-life applications for machine learning.

(a). Describe three real-life applications in which classification might be useful. Describe the response, as well as the predictors.

(b). Describe three real-life applications in which regression might be useful. Describe the response, as well as the predictors.

(c). Describe three real-life applications in which cluster analysis might be useful.

Problem 3. [10 points]

We want to learn a model to predict Y. Let n denote the number of samples of data. Using calculus, derive the optimal constant function $\widehat{Y}(X) = \beta_0^*$ under mean square error

$$\beta_0^* = \underset{\beta_0 \in \mathbb{R}}{\arg\min} \ \frac{1}{n} \sum_{i=1}^{n} (Y(i) - \beta_0)^2 \, .$$

Make sure to check that you found a minimizing β_0, not a maximizing β_0.