**Homework 1 (Due at 12 noon on Sep 27)**

**1**. For each of parts (a) through (d), indicate whether we would generally expect the performance of a ﬂexible statistical learning method to be better or worse than an inﬂexible method. Justify your answer.

1. The sample size *n* is extremely large, and the number of predictors *p* is small.
2. The number of predictors *p* is extremely large, and the number of observations *n* is small.
3. The relationship between the predictors and response is highly non-linear.
4. The variance of the error terms, i.e. σ2 = Var(), is extremely high.

**2**. We now revisit the bias-variance decomposition.

1. Provide a sketch of typical (squared) bias, variance, training error, test error, and Bayes (or irreducible) error curves, on a single plot, as we go from less ﬂexible statistical learning methods towards more ﬂexible approaches. The x-axis should represent the amount of ﬂexibility in the method, and the y-axis should represent the values for each curve. There should be ﬁve curves. Make sure to label each one.
2. Explain why each of the ﬁve curves has the shape displayed in part (a).

**3**. What are the advantages and disadvantages of a very ﬂexible (versus a less ﬂexible) approach for regression or classiﬁcation? Under what circumstances might a more ﬂexible approach be preferred to a less ﬂexible approach? When might a less ﬂexible approach be preferred?

**4**. Describe the diﬀerences between a parametric and a non-parametric statistical learning approach. What are the advantages of a parametric approach to regression or classiﬁcation (as opposed to a nonparametric approach)? What are its disadvantages?