CAINE, Wilbert (20584260)

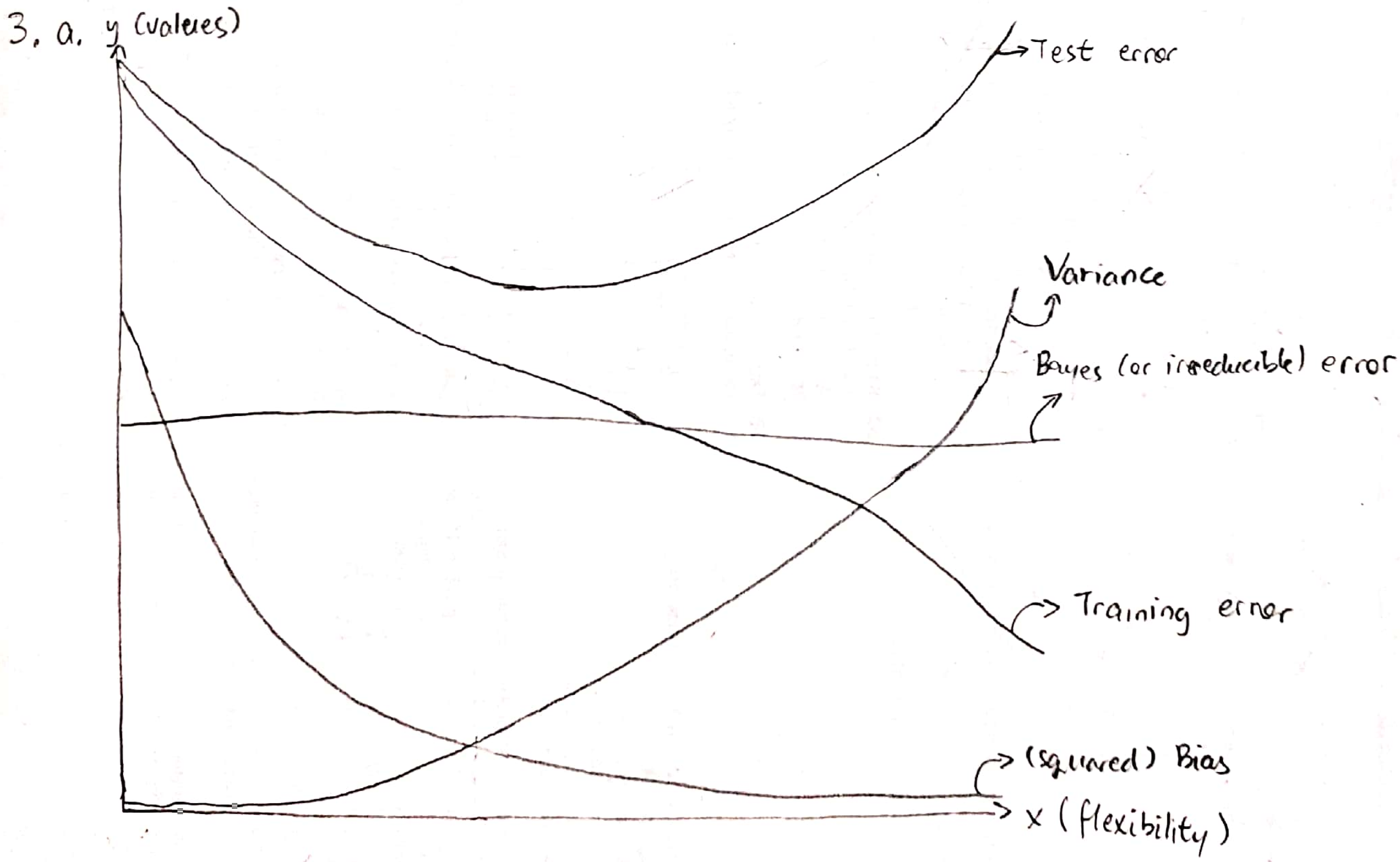
1\. a. Better. Large sample size and small number of predictor is likely to fit a model without overfitting. Flexible method allows to capture the relationship in the dataset better than inflexible method. Thus, flexible method can better fit the data while overfitting is unlikely to occur.

b\. Worse. In contrast to Q1a, overfitting to the noise will be observed with small number of observation and large number of predictors. Thus, inflexible method is better than the flexible ones.

c\. Better. Given the non-linear relationship, flexible method is required to capture the non-linearity between the predictors and response.

d\. Worse. Given the high variance of the error terms, a flexible method will overfit to the noise which is not desirable. When a flexible method is applied to a new dataset, it will generally perform worse than inflexible methods.

3\. a.



3.b. Bayes (or irreducible) error is a positive constant which describes the unknown noise that (almost) always exist in real-life problems

(Squared) Bias decreases because the relationship between predictors and outcome can be captured better with a more flexible model.
In real life problems, non-linear relationship often exists in the dataset.
Thus, flexible model can reduce the bias.

Variance is higher for a more flexible model because it captures the amount of change in the prediction given that the new data is changed. When flexibility increases, the methods overfits to the noise. Thus, the model fails to generalize the problem and produce high variance between new sets of data.

Test error is high when flexibility is too low, because of underfitting, or when flexibility is too high, because of overfitting. The test error changes with respect to the bias-variance tradeoff, and reaches its minimum when $E_D[(y-\hat{f}(x;D))^2] = (\text{Bias}_D[\hat{f}(x;D)])^2 + \text{Var}_D[\hat{f}(x;D)] + \sigma^2$ reaches its minimum.

Training error keeps on decreasing as the flexibility increases. Initially, the training error is high because of underfitting. By gradually increasing the flexibility, the training error will always decrease since the model will try to fit the training dataset and more flexible models are able to fit the training set better because they captures the nonlinear relationship. Eventually, the model overfits to the noise in the Training data which causes the training error to be low.

3.1. By constructing null hypothesis and the alternative hyhothesis accordingly:

* Intercept: $H_0: \beta_0 = 0$
  $H_1: \beta_0 \neq 0$

Intercept refers to the expected number of sales without any advertising budget to spend.
Given the p-value < 0.0001, we conclude that there is significant evidence to reject Ho.
This means that sales would not be zero given that there is no advertising budget

* TV: $H_0: \beta_1 = 0$
  $H_1: \beta_1 \neq 0$
* Radio: $H_0: \beta_2 = 0$
  $H_1: \beta_2 \neq 0$
* Newspaper: $H_0: \beta_3 = 0$
  $H_1: \beta_3 \neq 0$

The null hypothesis for TV means that the variable has no effect on sales, given that radio and newspaper is fixed

—— " —— Radio —— " —— " —— TV and newspaper —"

—— " —— Newspaper —— " —— " —— TV and radio —"

The p-value is significant for TV and Radio, which is <0.0001, while the p-value is NOT significant for Newspaper, which is 0.8599.
Thus, we reject Ho for TV and radio and we do NOT reject Ho for newspaper
We conclude that the budget for TV and Radio affect sales but the budget for newspaper does NOT affect sales

3.5. $$\hat{y}_i = x_i \hat{\beta} = x_i \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{k=1}^{n} x_k^2}$$

$$= \frac{\sum_{j=1}^{n} x_i \cdot x_j y_j}{\sum_{k=1}^{n} x_k^2}$$

$$= \sum_{j=1}^{n} \frac{x_i \cdot x_j y_j}{\sum_{k=1}^{n} x_k^2}$$

$$= \sum_{j=1}^{n} \frac{x_i \cdot x_j}{\sum_{k=1}^{n} x_k^2} y_j = \sum_{j=1}^{n} a_j y_j \quad \text{where } a_j = \frac{x_i \cdot x_j}{\sum_{k=1}^{n} x_k^2}$$