

# Stat 340 - Quiz 1

**What's Your Name?** \_\_\_\_\_

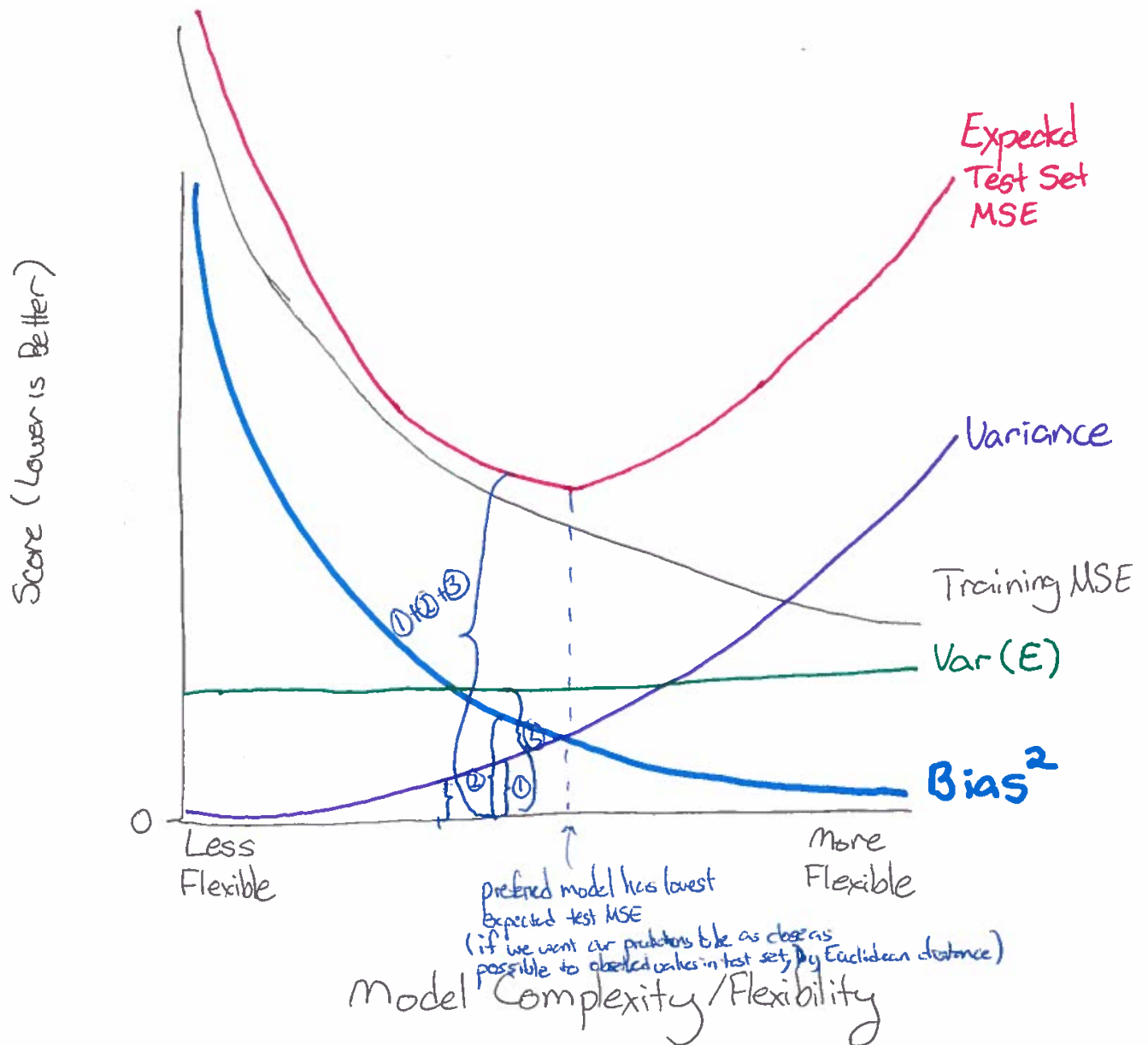
## **Problem 1.**

On the axes below, draw a sketch where the horizontal axis represents model flexibility (going from less flexible on the left to more flexible on the right) and the vertical axis represents model score, with four curves on it (these curves should add up appropriately):

1. Squared bias of predicted values
2. Variance of predicted values
3. Error in predictions due to noise in the data generating process (i.e.,  $\text{Var}(\varepsilon)$ )
4. Expected test set MSE

Label each of your curves.

**Solution:**



Since expected test set MSE =  $\text{Bias}^2 + \text{Variance} + \text{Variance}(E)$   
 the curves must sum up at each point on the x axis.

**Rubric: 50 points total**

- 10 points: correct shape for  $\text{bias}^2$
- 10 points: correct shape for variance
- 10 points: correct shape for  $\text{Var}(\varepsilon)$
- 10 points: correct shape for Expected test set MSE
- 5 points: correct relative location for expected test set MSE curve
  - 3 points if above all other curves
  - 5 points if they come close to adding up
- 5 points: correct location for preferred model, minimizing expected test set MSE

## Problem 2.

In problem 1, what was meant by “Bias of predicted values”? Your answer should include the following terms or phrases:

- “ $\hat{f}(x_0)$ ”
- “ $y_0$ ”
- “training sets”

## Solution:

The bias of the predicted values is the average difference between the predicted values  $\hat{f}(x_0)$  and observed values  $y_0$  in a test set, where the average is taken across all model fits  $\hat{f}$  from all possible training sets of a given size (and across all observed values  $y_0$  we might get in a test set).

(More explanation that you didn’t need to write: each training set gives us a different model fit, and so a different prediction  $\hat{f}(x_0)$ . On average across all training sets, what is the difference between these predictions and the observed values we might see in a separate test set?)

## Rubric:

- 15 points: average difference between observed and predicted values
  - 5 points for average
  - 10 points for talking about difference between observed and predicted
    - \* 5/10 on this part if you talk about “how close” they are
- 10 points: ... across all model fits that would be obtained from all possible training sets of a given size.

## Problem 3.

In problem 1, what was meant by “Variance of predicted values”? Your answer should involve the following terms or phrases:

- “ $\hat{f}(x_0)$ ”
- “training sets”

## Solution:

The variance of the predicted values describes the variability in predictions  $\hat{f}(x_0)$  obtained from all model fits  $\hat{f}$  from all possible training sets of a given size.

(More explanation that you didn't need to write: each training set gives us a different model fit, and so a different prediction  $\hat{f}(x_0)$ . How much variability is there in these predictions across all training sets?)

**Rubric:**

- 15 points: amount of variability of predicted values
- 10 points: ... across all model fits that would be obtained from all possible training sets of a given size.