

Names:

DATA 311 – Lecture 21 – Bias, Variance, Irreducible Risk, Overfitting

Suppose you aim to predict a student's final exam score given their scores on all the assignments in the first half of the quarter. For each of the following, say whether it is most likely to **decrease** (a) bias, (b) variance, or (c) irreducible risk.

1. Add additional information, such as amount of time spent studying, to your feature set.
2. Use a more complicated/flexible model.
3. Use a less complicated/flexible model.

Suppose you split your data into training and validation sets, train a model, and evaluate its accuracy on both the training set and on the validation set.

4. How would you describe each of the four possible scenarios? Focus on the left column first.

		Validation Accuracy	
		Bad	Good
Training Accuracy	Bad		
	Good		

