

# WS #14 - Variance + Bias

Monday, November 4, 2024

Your Name: \_\_\_\_\_

Names of people you worked with: \_\_\_\_\_

What is your favorite place to hang out on the 5Cs?

## Task:

*Bias* in a model happens when the model is too simple and doesn't quite capture the true (more complicated) model.

*Variance* in a model happens when the model is too complicated because it over fits the training data.

The two questions below are about the **model** (not about the data, test, training, etc.).

1. For  $k$ -NN, what would lead to a more biased model? What would lead to a higher variance model? (Hint: think about what you can control.)
2. For CART, what would lead to a more biased model? What would lead to a higher variance model? (Hint: think about what you can control.)

**Solution:**

1.  $k = n$  would be more biased,  $k = 1$  would be higher variance.
2. One split would be more biased ( $\alpha$  huge), as many splits as possible (each observation in its own terminal node) ( $\alpha$  tiny) would be higher variance.