

## Class 4 Pre-work: Bias-Variance Tradeoff

Read this excellent article, [Understanding the Bias-Variance Tradeoff](#), and be prepared to **discuss it in class**.

**Note:** You can ignore sections 4.2 and 4.3.

Here are some questions to think about while you read:

- \* In the Party Registration example, what are the features? What is the response? Is this a regression or classification problem?
- \* Conceptually, how is KNN being applied to this problem to make a prediction?
- \* How do the four visualizations in section 3 relate to one another? Change the value of K using the slider, and make sure you understand what changed in the visualizations (and why it changed).
- \* In figures 4 and 5, what do the lighter colors versus the darker colors mean? How is the darkness calculated?
- \* What does the black line in figure 5 represent? What predictions would the best possible machine learning model make, with respect to this line?
- \* Choose a very small value of K, and click the button "Generate New Training Data" a number of times. Do you "see" low variance or high variance, and low bias or high bias?
- \* Repeat this with a very large value of K. Do you "see" low variance or high variance, and low bias or high bias?
- \* Try using other values of K. What value of K do you think is "best"? How do you define "best"?
- \* Does a small value for K cause "overfitting" or "underfitting"?
- \* Why should we care about variance at all? Shouldn't we just minimize bias and ignore variance?