

Hw1

how bias and variance change when model complexity change

question: mathematica proof for the change ?

用什么来measure? cross-validation

Bonus Question

How bias and variance would change when model complexity changes?

1. Analyze from definition

- Bias: how much the prediction value obtained by the model differs from the real value.

sampling and *estimation* could both introduce bias. As the discussion is mainly about the impact of model complexity, more focus will be on *estimation error* (for sampling error, resampling and repeat the model building process then derive the average of prediction values could help, however, no change in model complexity).

So the causes for a larger bias could be a simpler model to do a simpler approximation, as a more complicated model usually could do a better fit. The risk of underfitting exists. Because with simpler model, less assumptions are made. For example, with the simplest model I could think (simple linear regression)

- Variance: with the same model, if a different training set is used, how much the estimate of the function will be changed. Intuitively, when a more complicated model is selected, more parameters need to be calculated. It would be much easier to obtain a result with a large variance compared to the former results. High-variance might fit the training data well (as with different training data, large change in estimate), but the risk of overfitting exist (fit too good to predict the future value). Large noise might be fitted by the complicated model.

So, there exists bias-variance tradeoff. From the lecture slides,

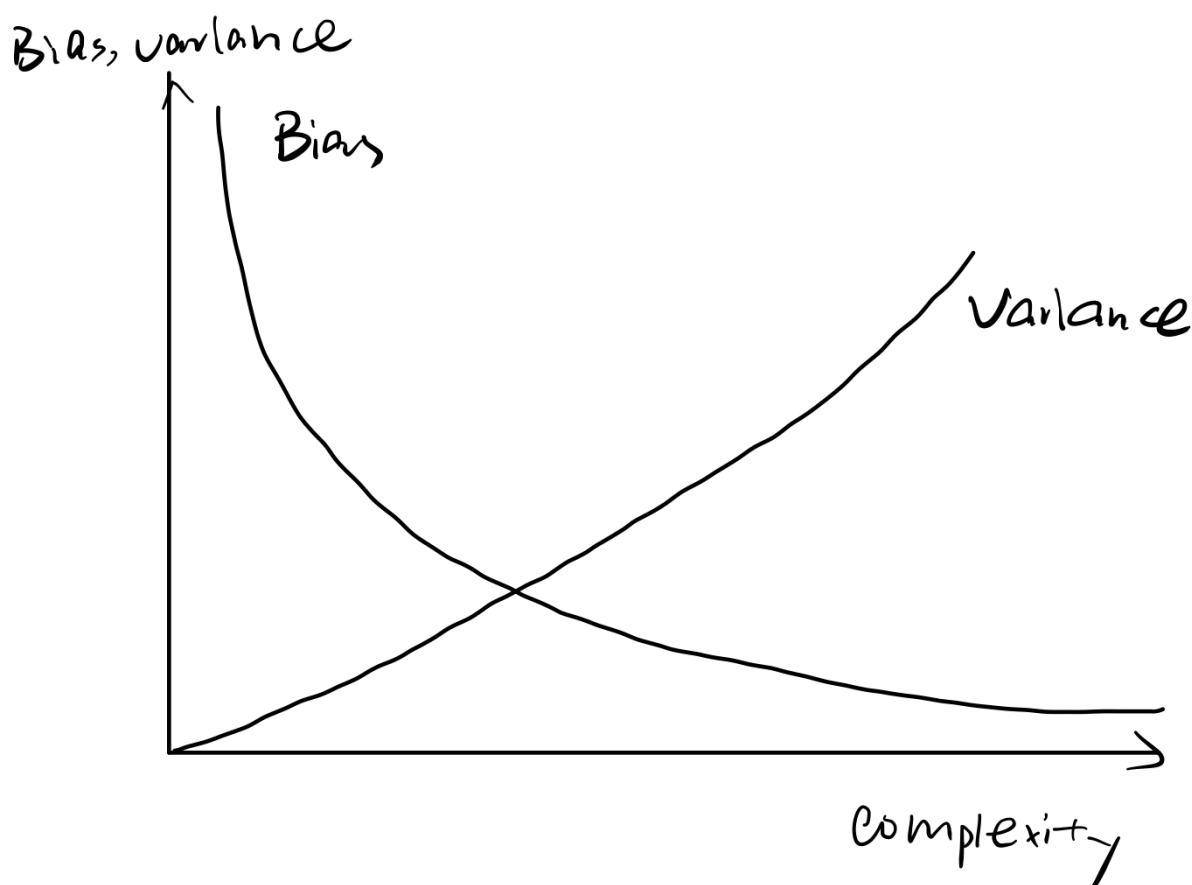
$$\text{MSE}(\mathbf{m}) = \mathbb{E}[(Y - \mathbf{m})^2] = \text{Var}[Y] + (\mathbb{E}[Y - \mathbf{m}])^2,$$

Model with small variance and high bias *underfit* the truth target, while with high variance and small bias *overfit* the truth target.

2. Answer for the trend

This answer is mainly based on the intuition in 1. and the fitted results shown in 3.

	Bias	Variance
monotone	yes	yes
change in rate	first change fast, then slower	roughly the same rate, do not differ much
Increase / decrease	Decrease	Increase



3. All the data will be

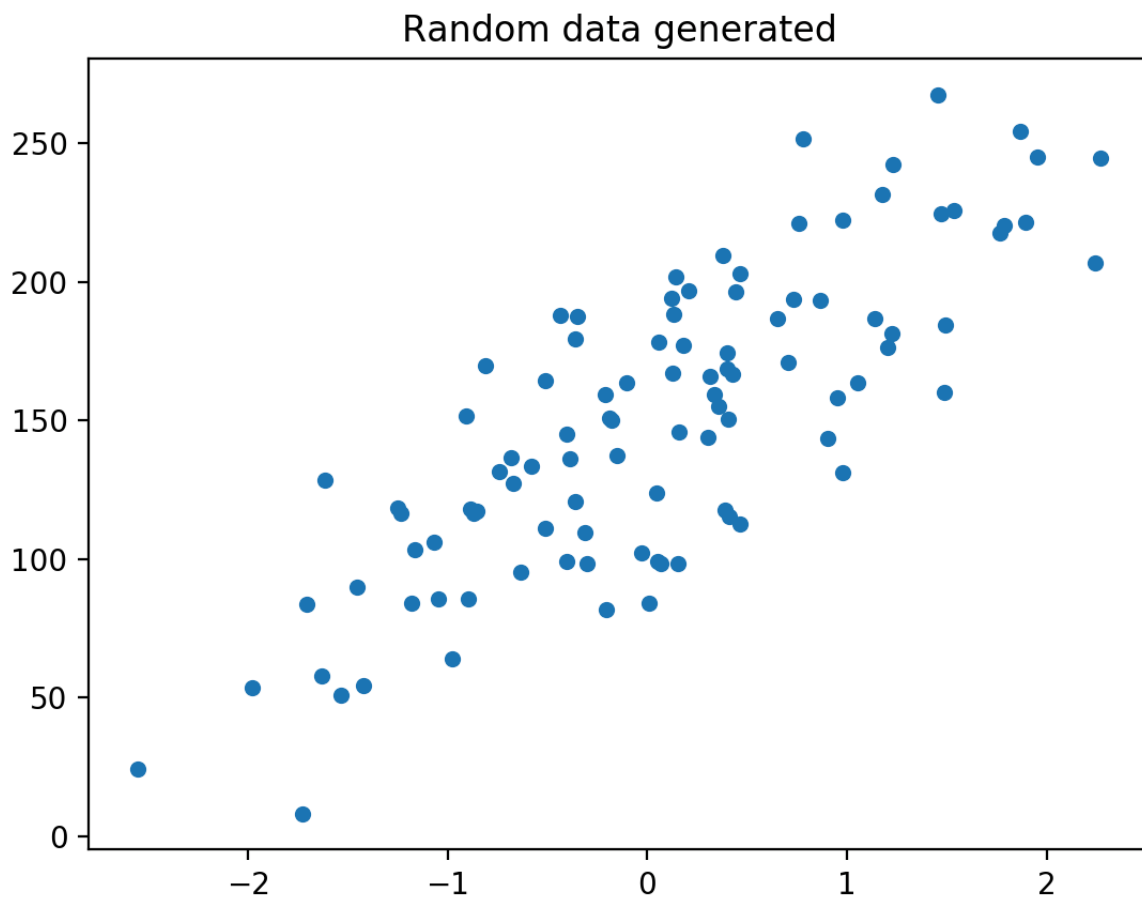
???真的这样吗

The data set is split into *training set* and *test set* and *K-fold Cross Validation* is used to train the data set. So for each model, we could get k different estimates, based on the prediction values (for each target value, k different prediction values) compared to the real value in the test set, the bias and the variance is calculated.

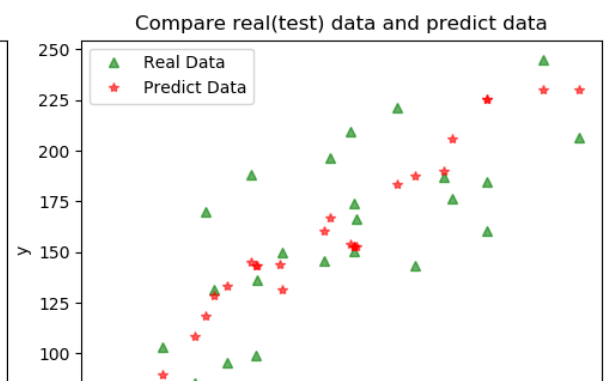
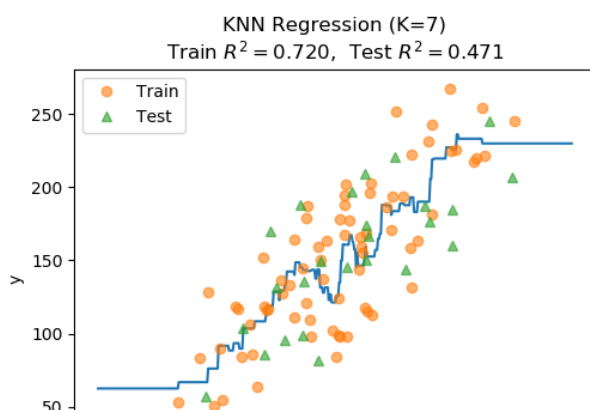
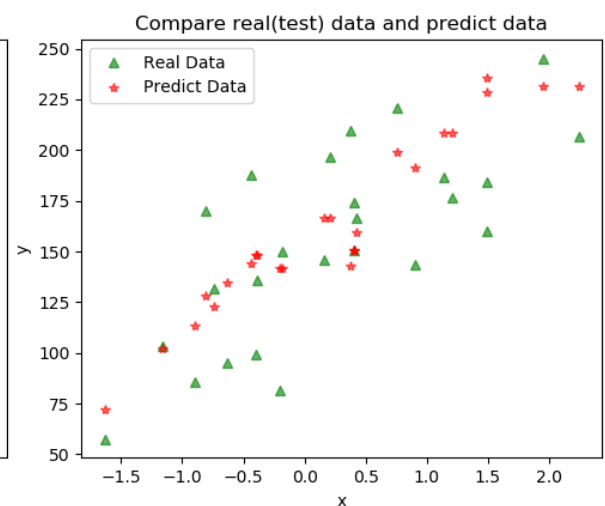
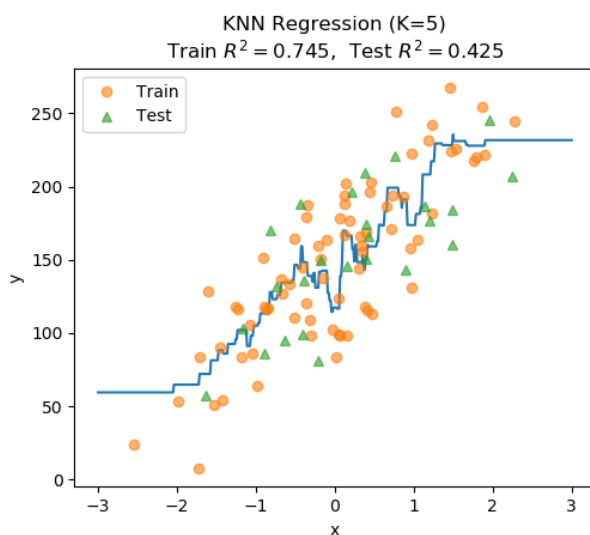
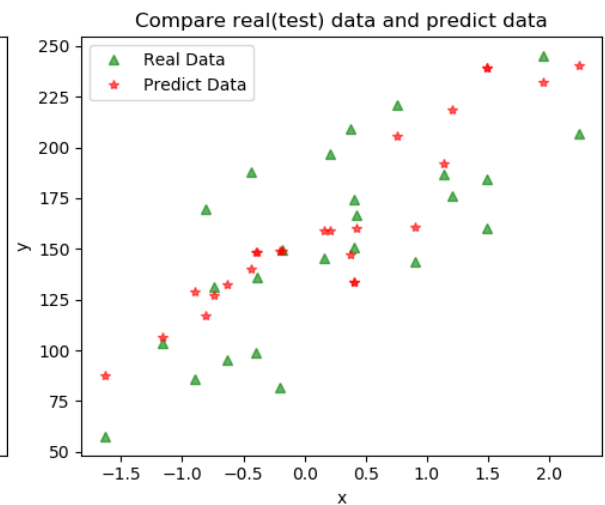
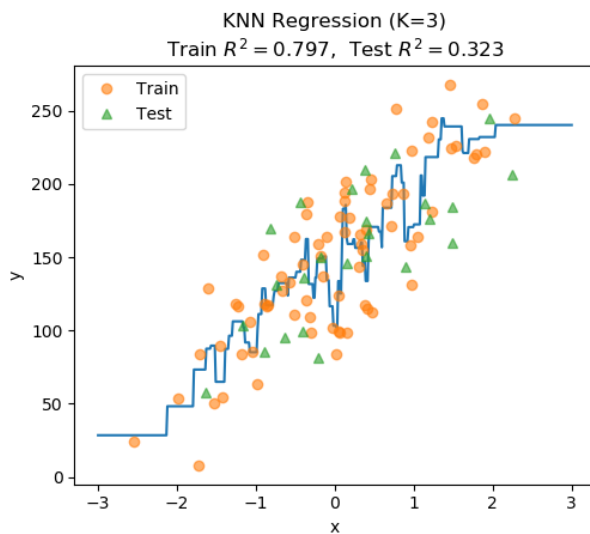
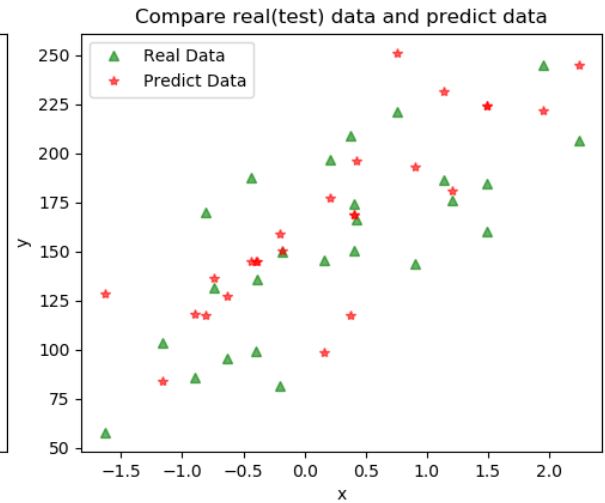
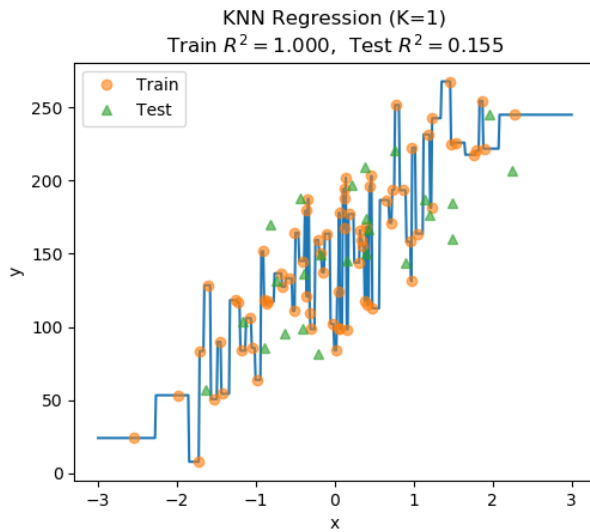
K-Nearest Neighbors method will be used. Because only change the number of K , different complexity model will be trained based on the equation.

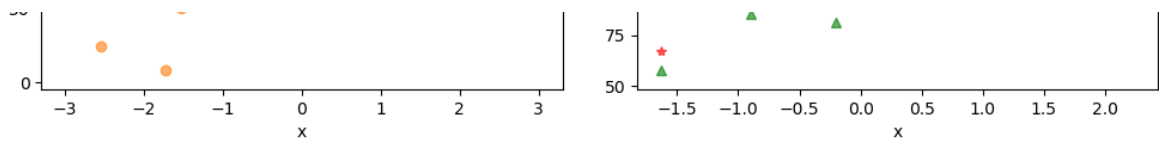
$$\hat{f}(x) = \text{Average}[y_i | x_i \text{ is in } N_k(x_i)]$$

data used is generated randomly, with 100 points (x_i, y_i) .



- The score calculated as $1 - u/v$, where u is the residual sum of squares $(y_{true} - y_{predict})^2$ and v is the sum of $(y_{true} - y_{test.mean})^2$. So a **better training score** means that the model explains the training set well, same as loss the flexibility, it would be much more likely to lose some ability to predict the future value, with a **lower testing score**. So *higher training score ~ lower variance ~ higher bias, higher testing score ~ higher variance ~ lower bias*.





可以直接用training test score 来画柱状图吗

Bias: test score? High test score, low bias?

是否需要formal 证明?

Denote the real value as y_1, y_2, \dots, y_n and the prediction value as y'_1, y'_2, \dots, y'_n , then $Bias = \sum_{i=1}^n (y'_i - y_i)^2$.