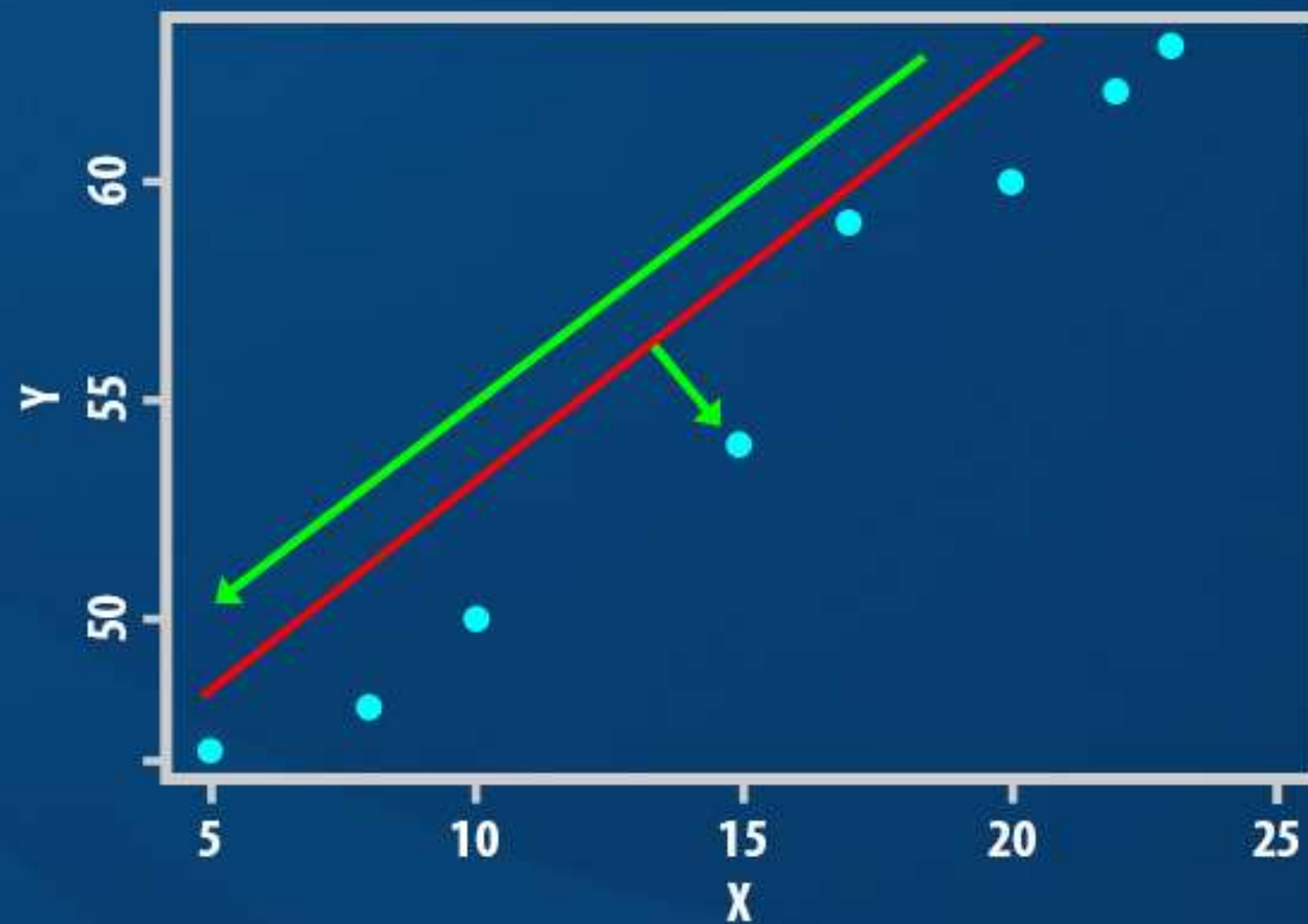


# UNDERFITTING AND OVERFITTING I

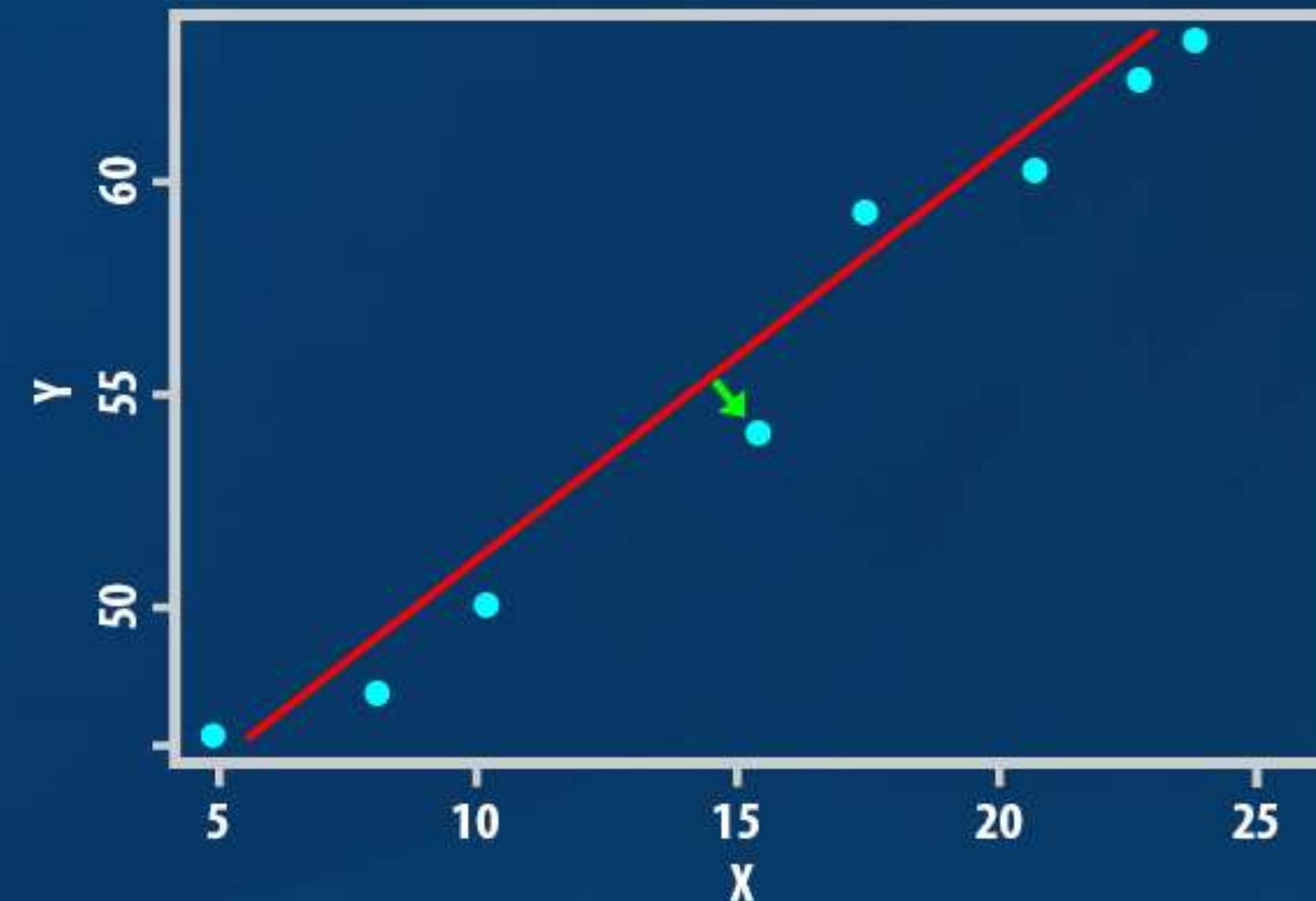
The concept of bias and variance in model selection raises issues of **underfitting and overfitting** respectively.

## Underfitting

- Underfitting is a situation where model is unable to capture the real trend of the data and, in result we receive poor performance.
- The supervised parametric models are more prone to underfitting problem. It is cause of high bias.



a) Example of "underfit".



(b) Example of "fit"

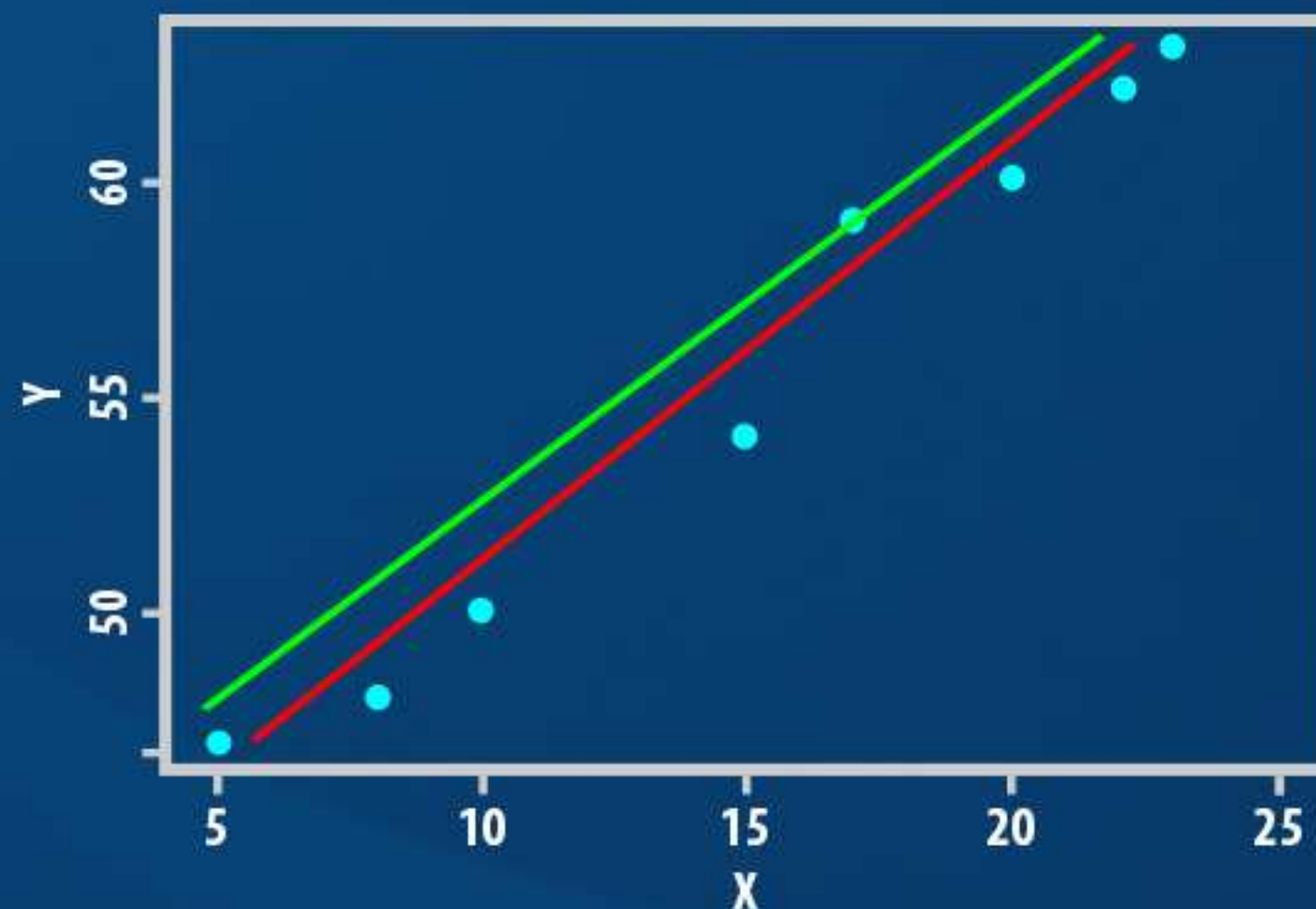
**Figure 38:** Example illustrations of "underfit", "fit" models. Red line on each plot represents the prediction line by the supervised model. Data point near or on the line is considered to be good prediction by the model



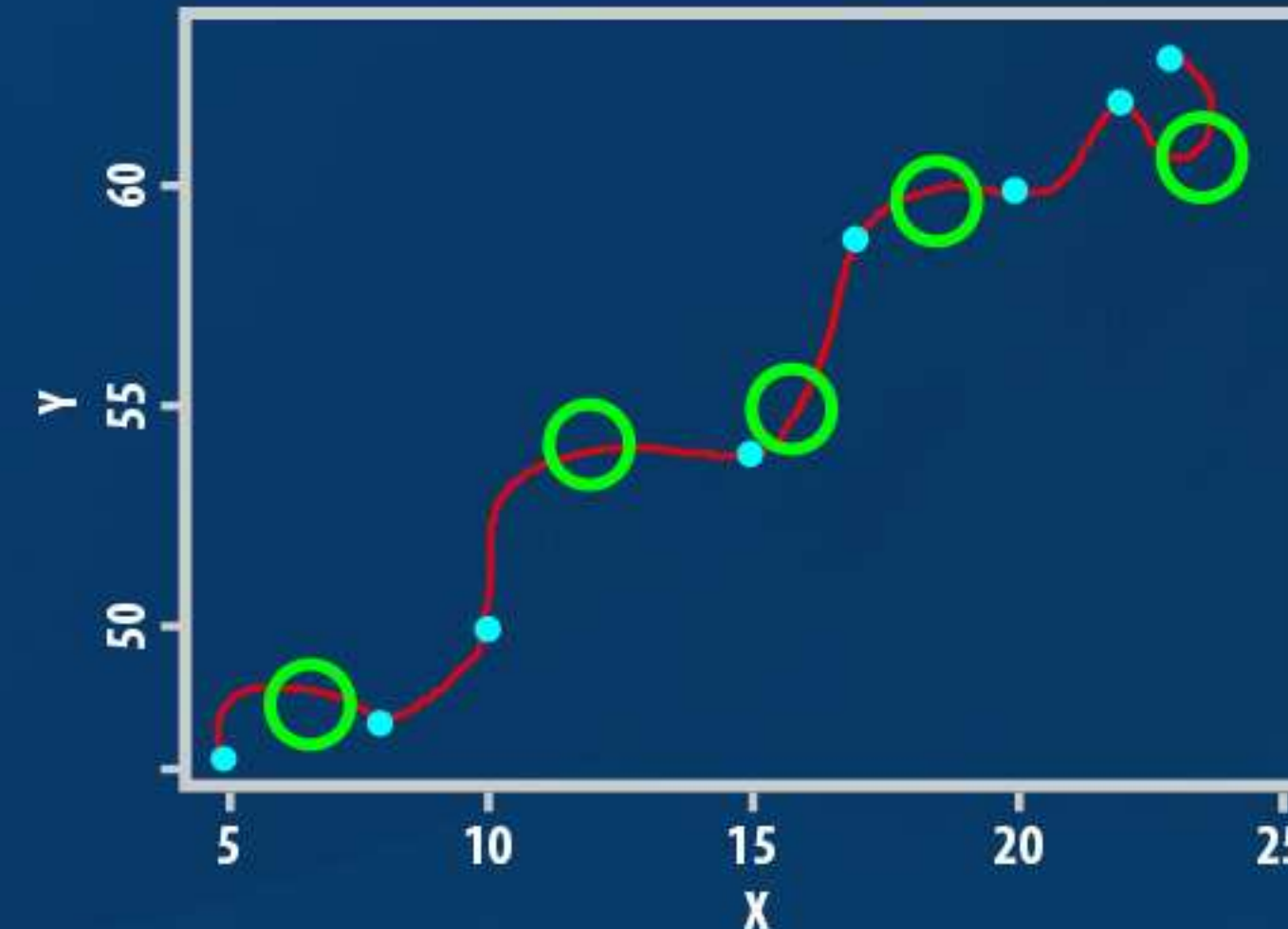
# UNDERFITTING AND OVERFITTING II

## Overfitting

- Overfitting arises when model behaves exceptionally good on training but fails during validation. The supervised non-parametric models falls into this category. It for the reason that non-parametric models have freedom to learn everything from the training data they are given. This may include noise and other irrelevant data not necessary for building predictive models.
- High variance is a cause of “overfitting”.
- Figure 39b illustrates example of overfitting where the model is too good to fit all the data points. This even captures randomness of the data which may not be there in the test data.



(a) Example of “fit” model



b) Example of “overfit” model

**Figure 39:** Example illustrations of “fit”, “overfit” models. Red line on each plot represents the prediction line by the supervised model. Data point near or on the line is considered to be good prediction by the model



# UNDERFITTING AND OVERFITTING III

## How to overcome “underfitting” and “overfitting”?

1. Adding more data and more features generally help reducing the model underfitting issue. More data results in better opportunity for model to learn.
2. Underfitting is usually caused by parametric supervised models where they are unable to capture the true trend of data resulting in poor performance. To overcome underfitting best practice is to use non-parametric models and reevaluate classifier.
3. Overfitting can be reduced by using Cross-validation evaluation method.
4. Bootstrap and ensembles can also be used to avoid overfitting.