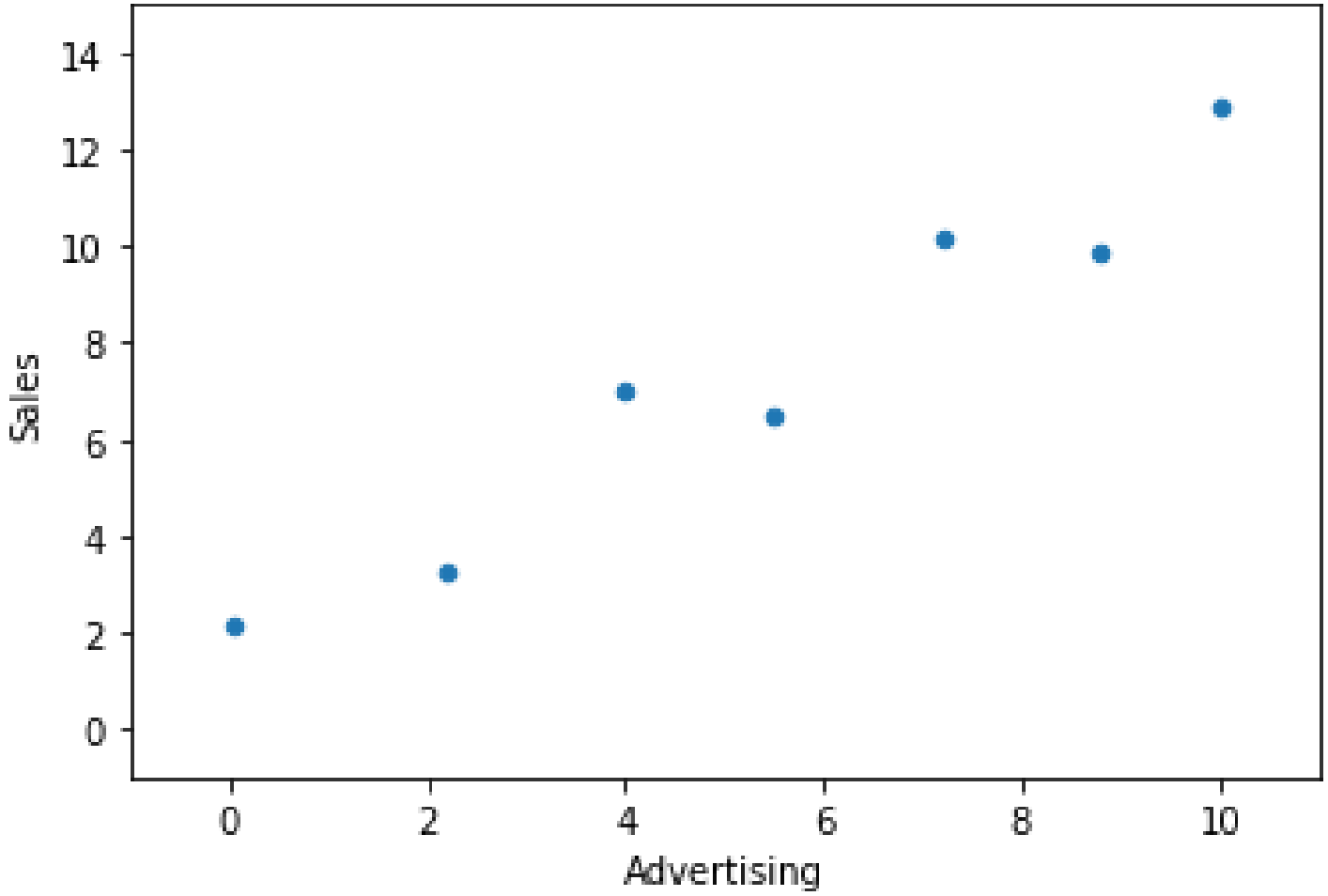
Overfitting and underfitting

Toni Ernvall

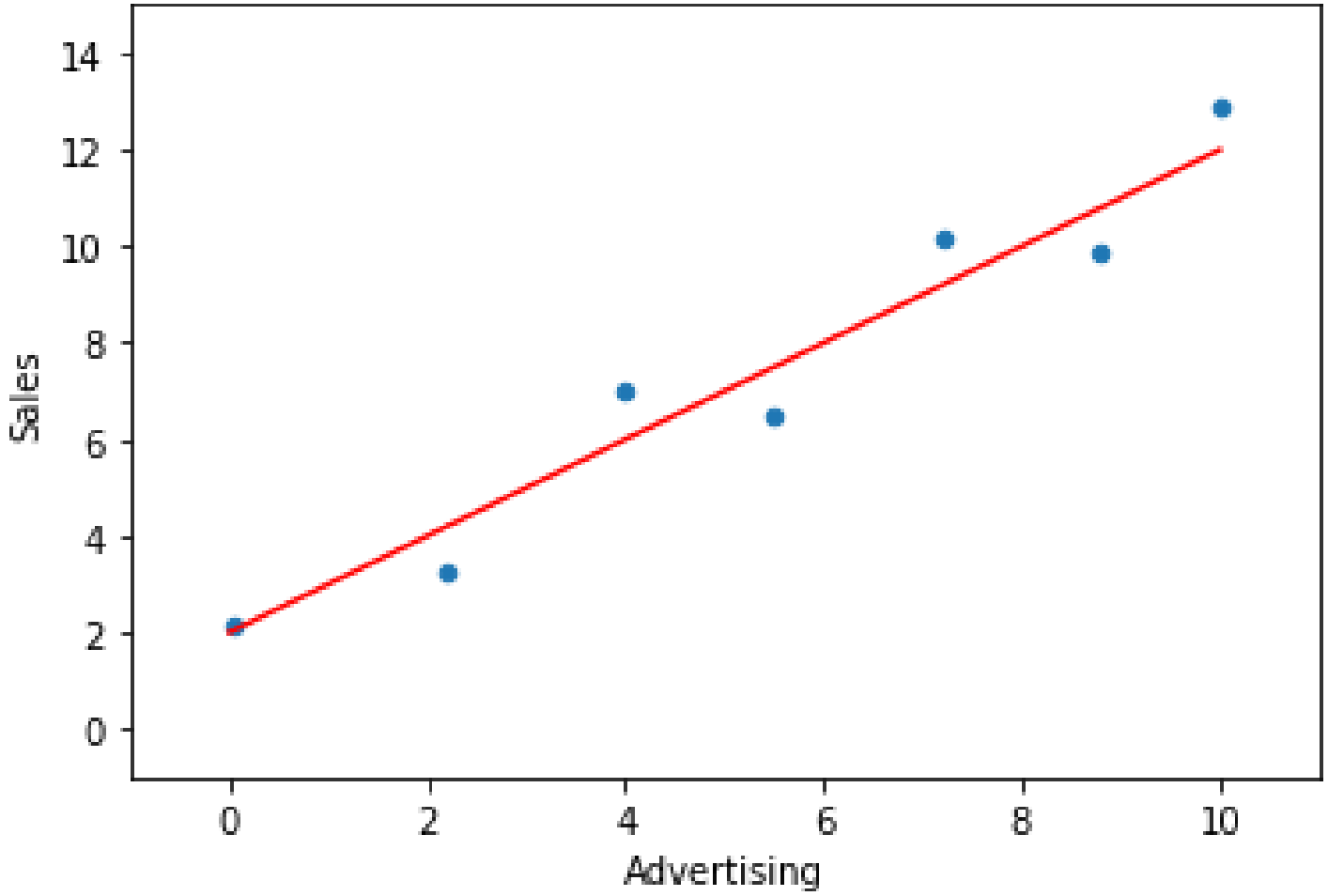
Overfitting

- Consider the scatter plot. It describes the connection between advertising and sales observed in the teaching data.
- We are trying to find a model that could describe this connection.



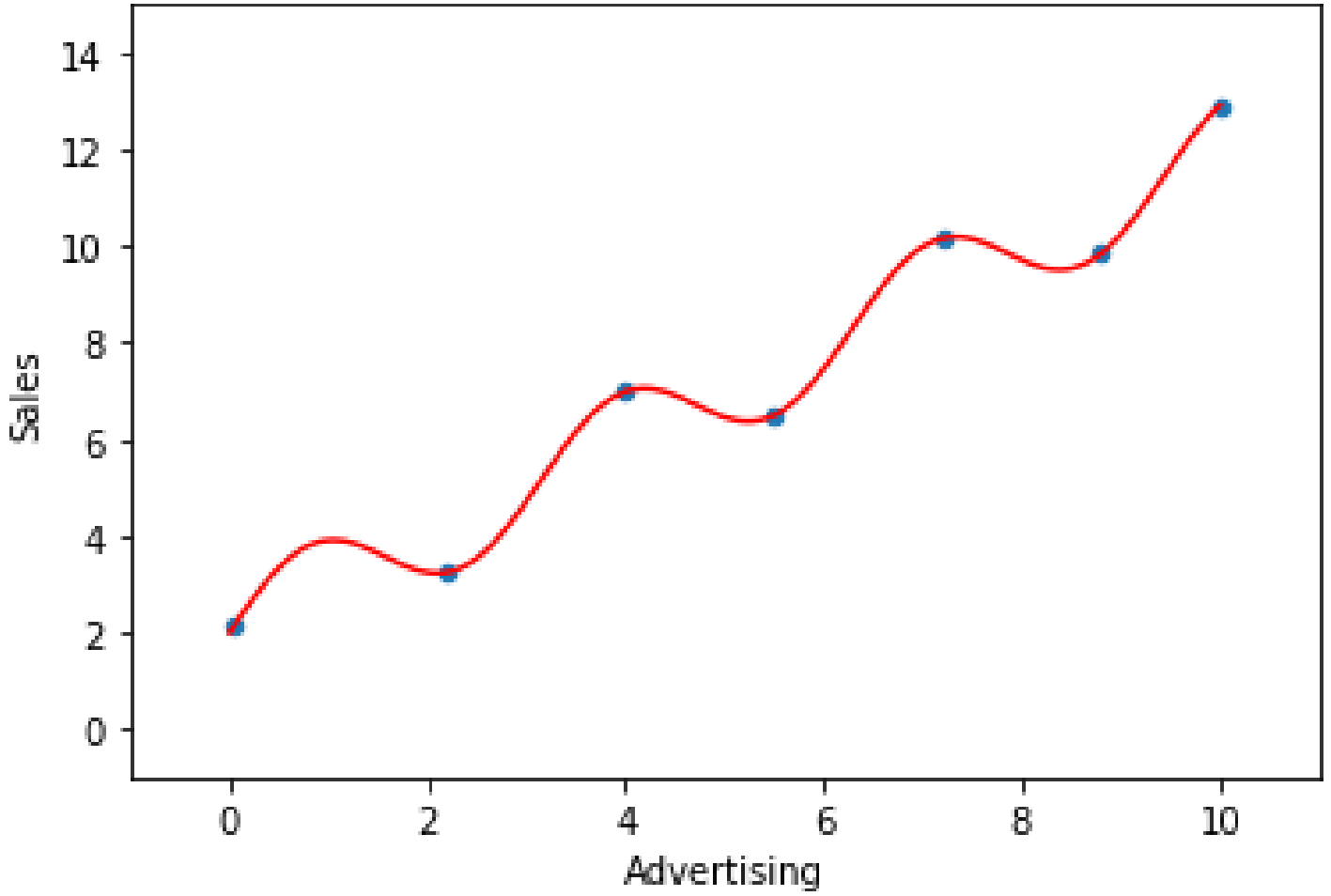
Overfitting

We could model the connection with a simple linear regression model.



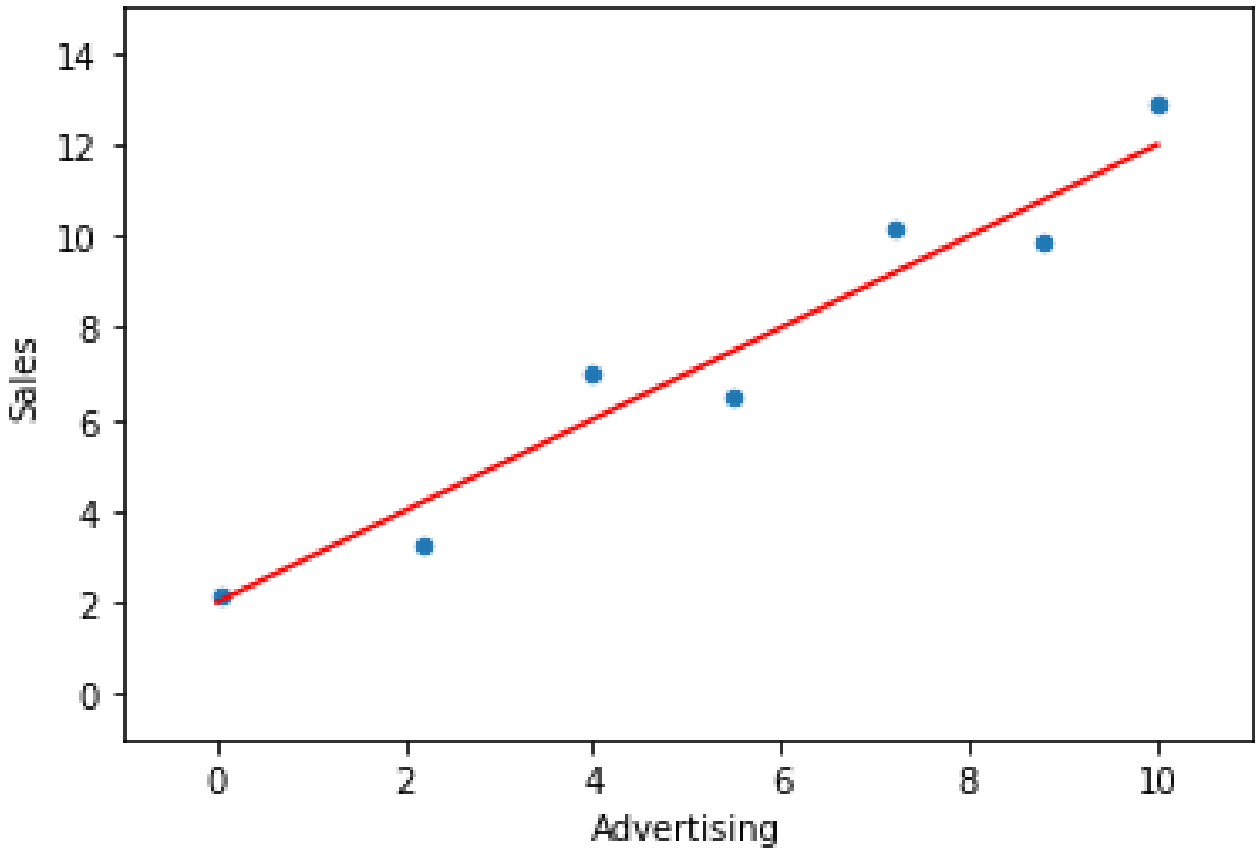
Overfitting

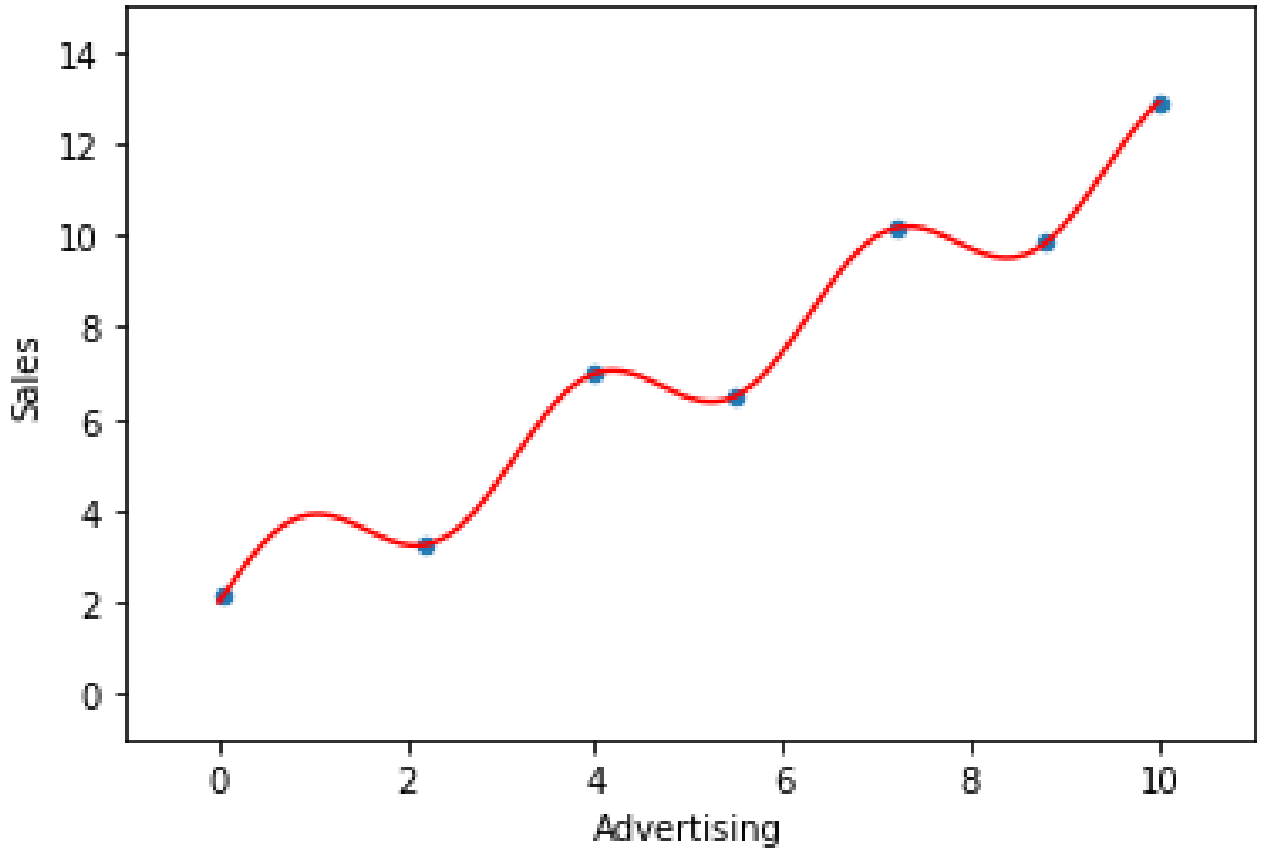
On the other hand, the connection could be modeled with a model that takes into account more precisely the variability observed in the teaching data. This has been done here.



Overfitting

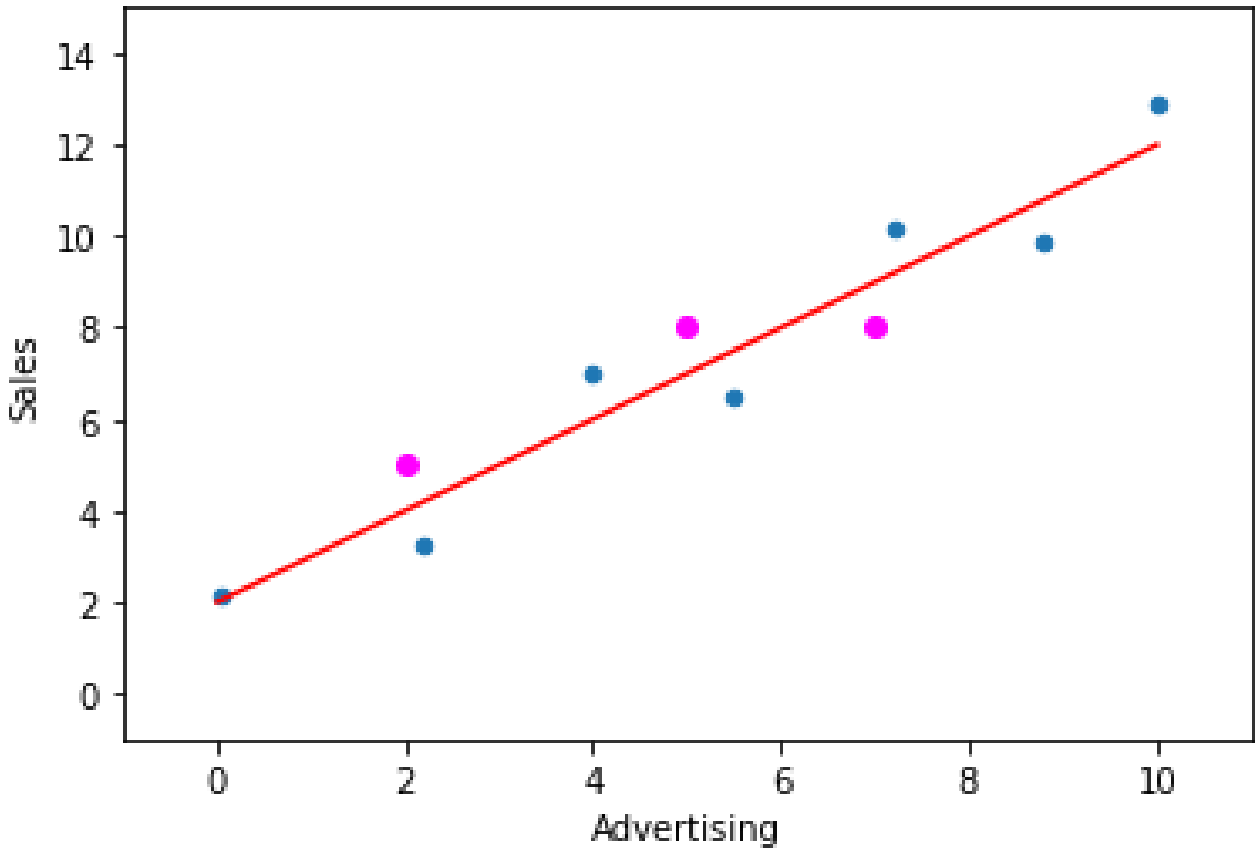
Now let's compare how these two models perform with the test data.

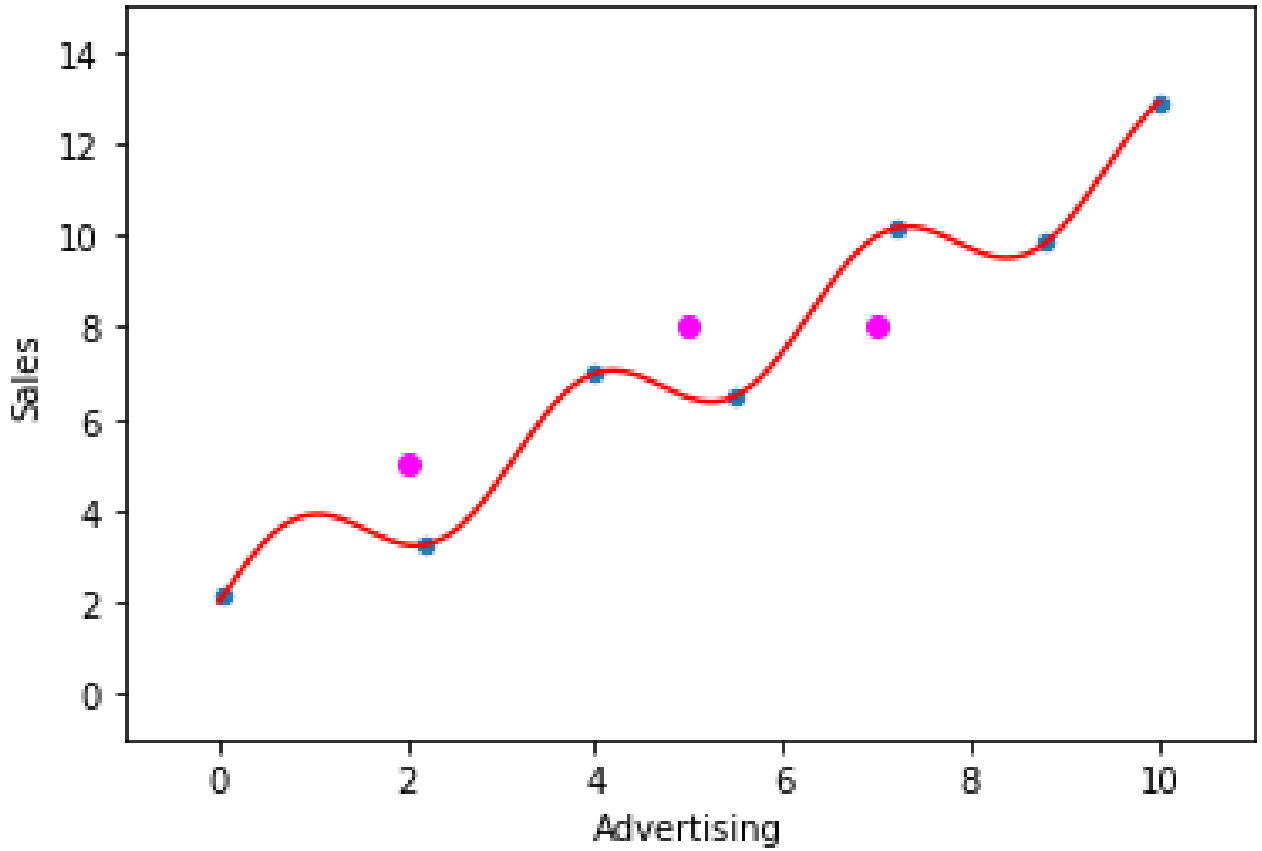




Overfitting

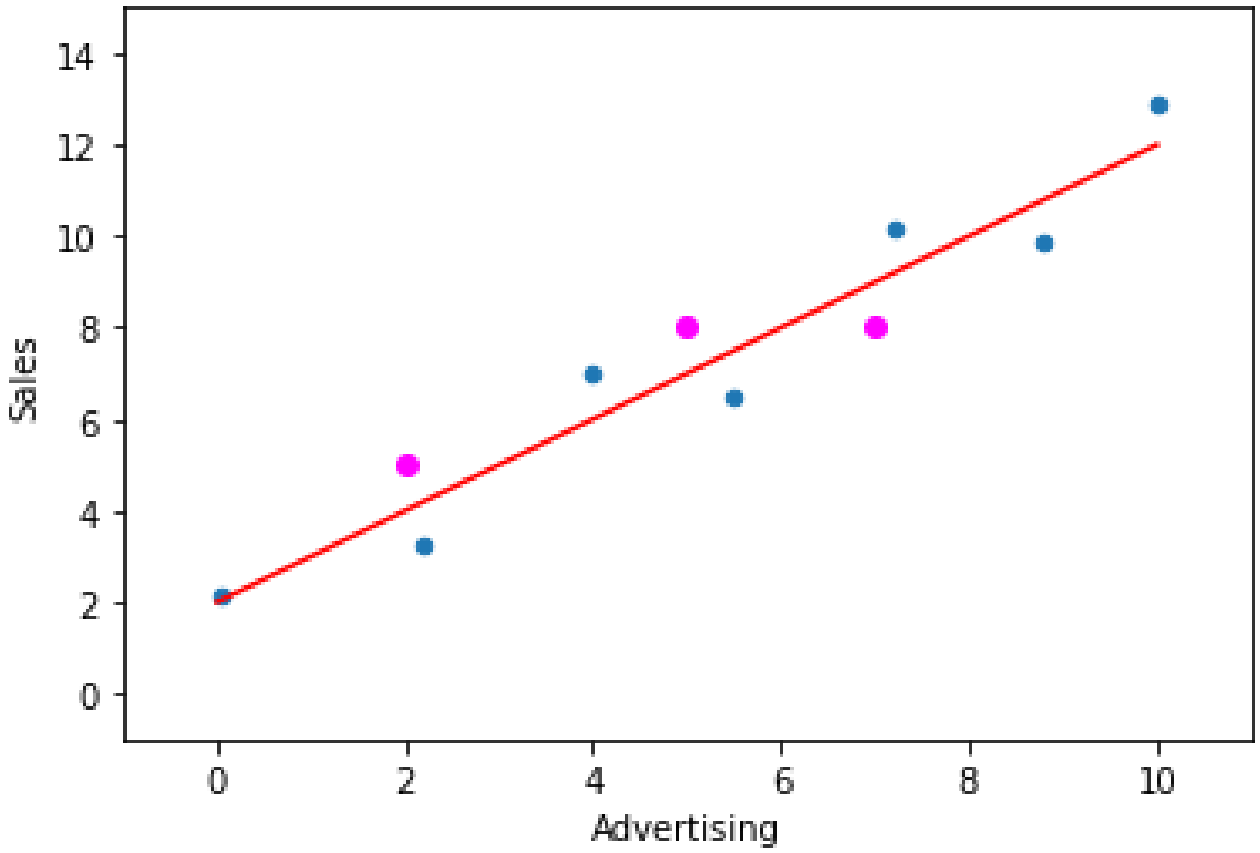
Now let's compare how these two models perform with the test data.

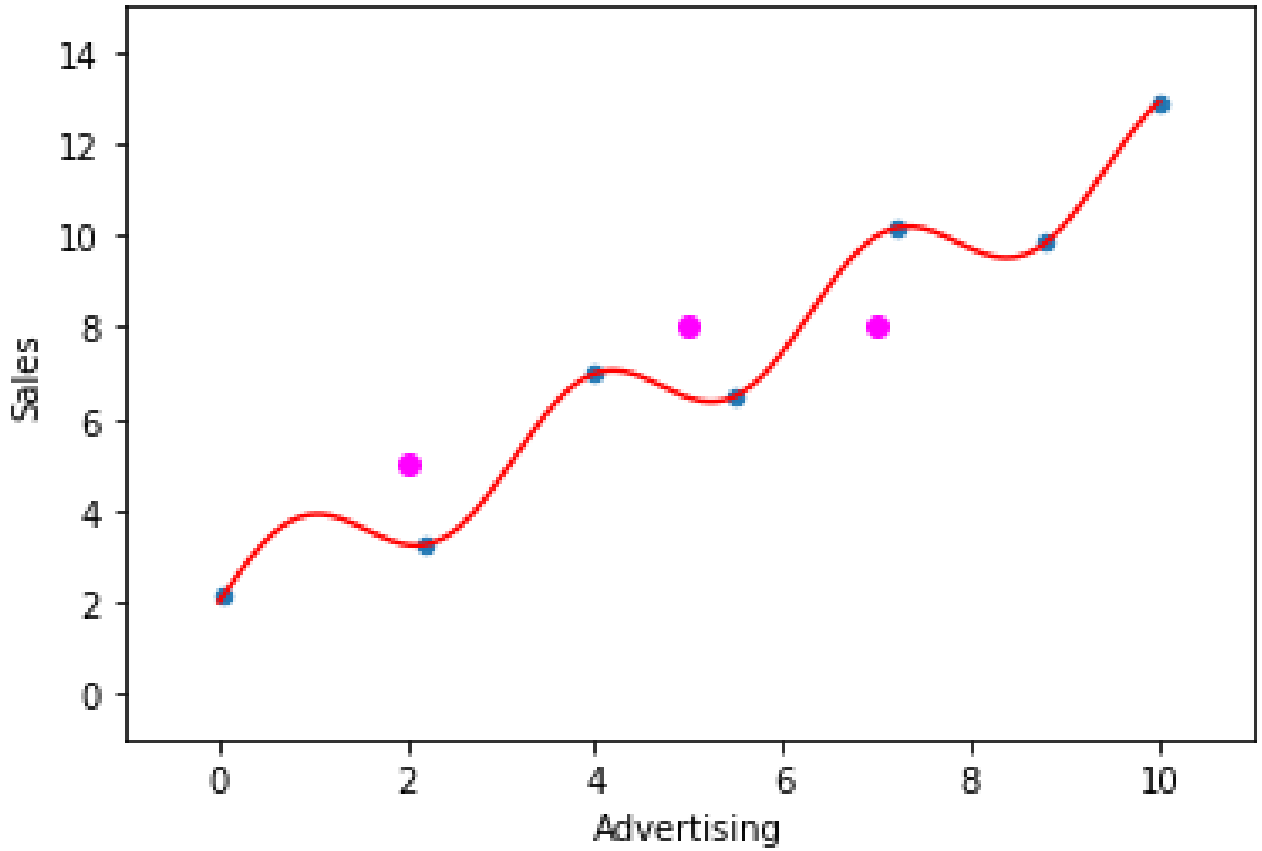




Overfitting

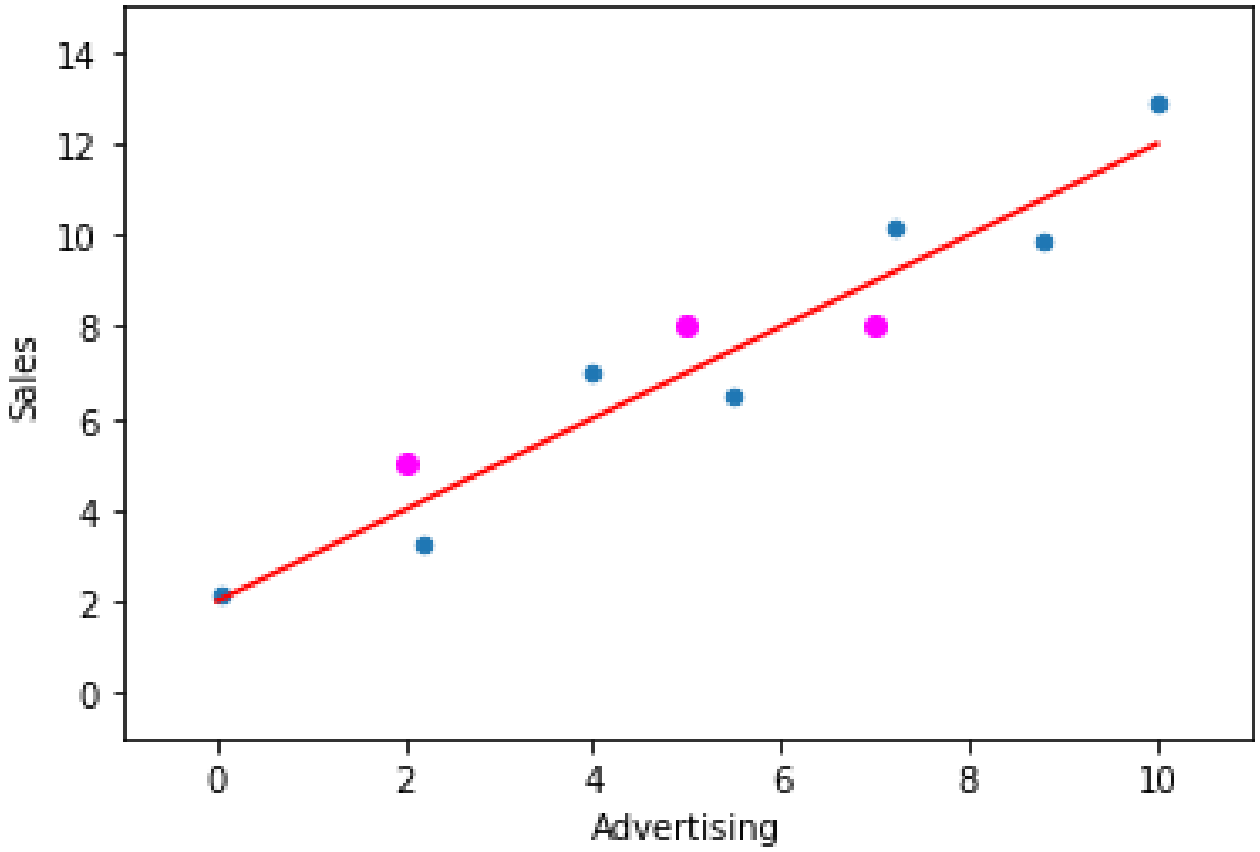
We notice that the simpler model predicts the test data better.

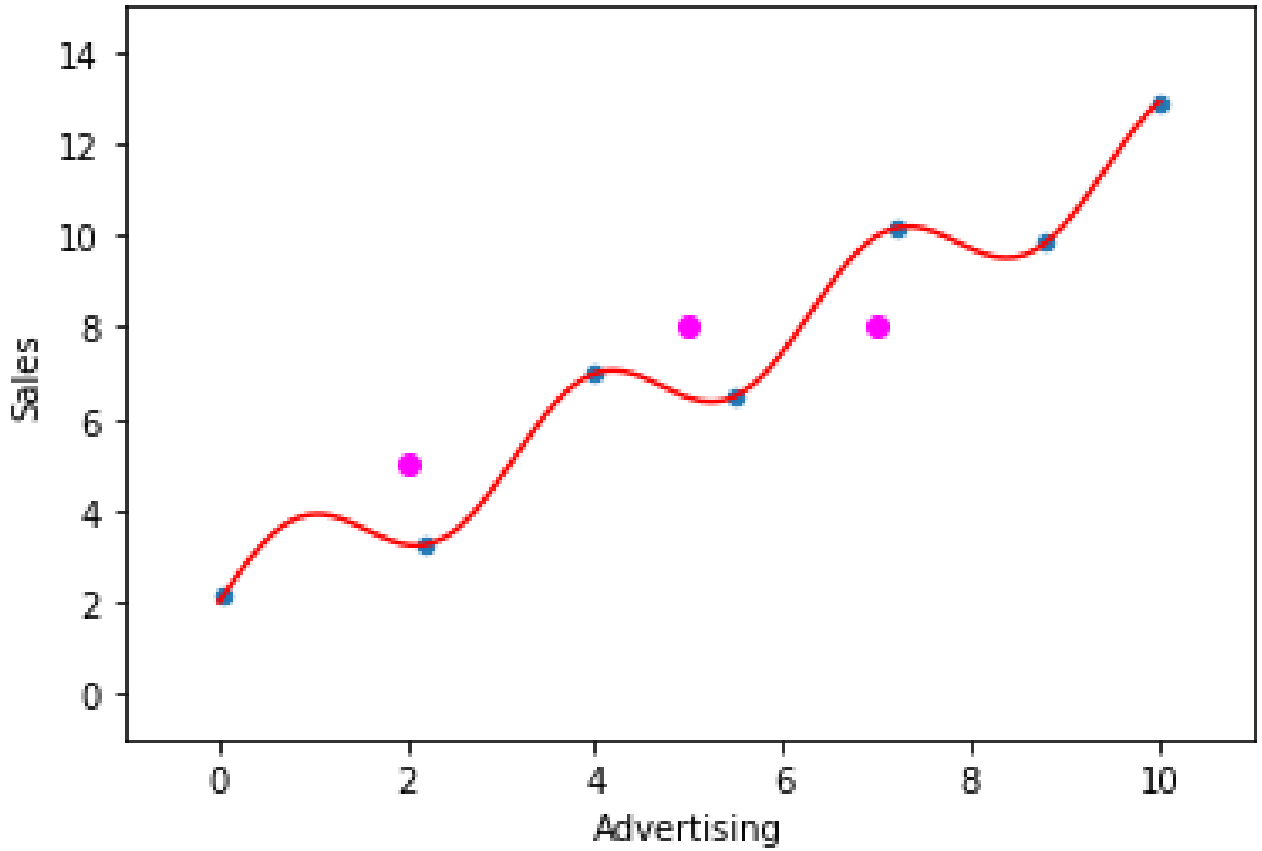




Overfitting

Overfitting refers to building a model so that it follows the features of the training data too closely. In this case, it may also take into account the random variation in the training data, which is not expected to be included in the model.



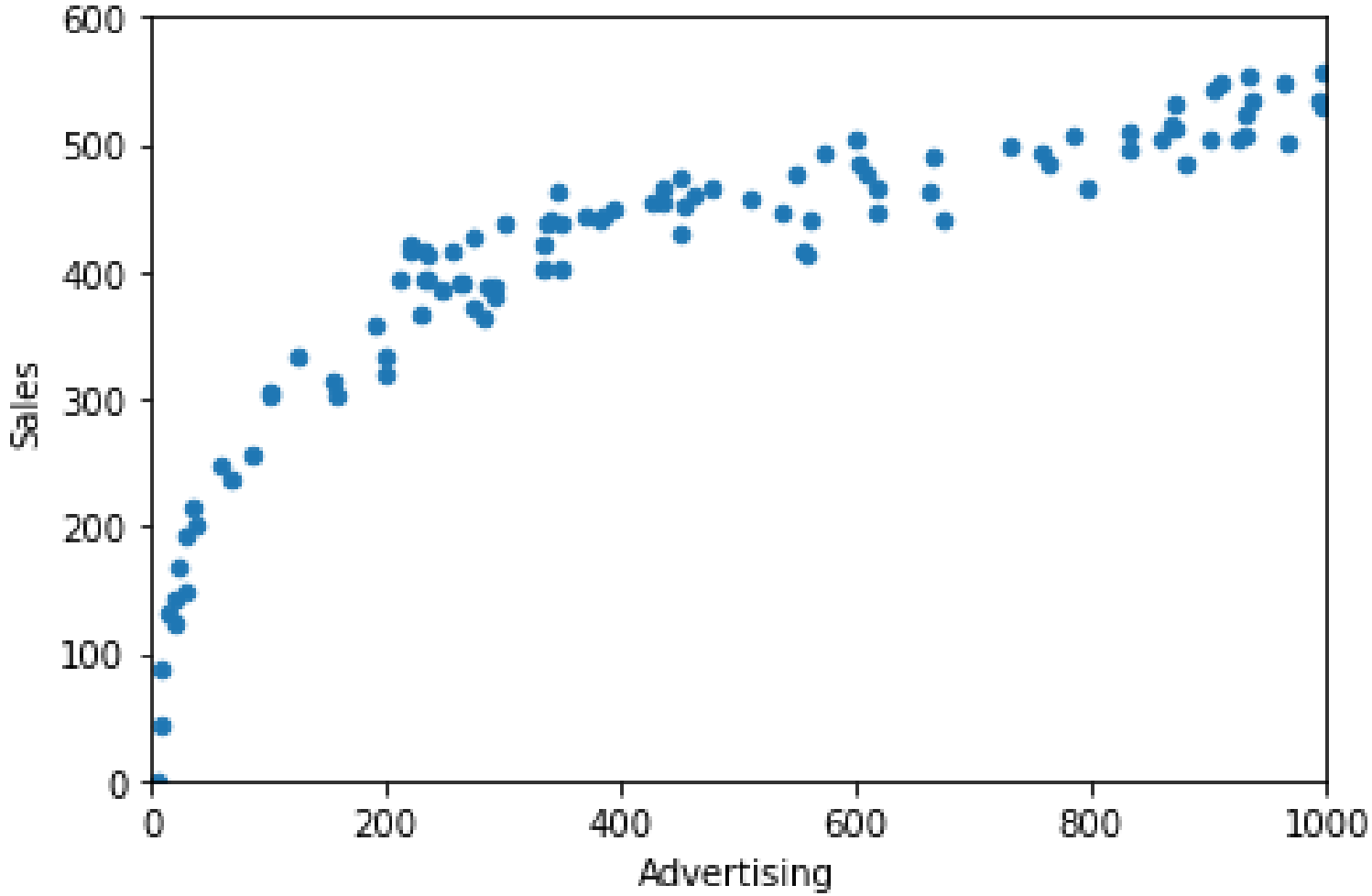


Underfitting

- It was noticed above that an overfitting model is not good.
- On the other hand, underfitting is not a good thing either.

Underfitting

- For example, the connection in the scatter plot cannot be modeled very well with a linear model (i.e. a straight line in a scatter plot).
- A more complex model would be needed here.



Underfitting

- For example, the connection in the scatter plot cannot be modeled very well with a linear model (i.e. a straight line in a scatter plot).
- A more complex model would be needed here.
- For example, this model could be good.