Occam's Razor, a problem solving principle states that
"Among competing hypotheses, the one with the fewest assumptions should be selected. Other, more complicated solutions may ultimately prove correct, but—in the absence of certainty—the fewer assumptions that are made, the better."

One of the most common problem data science professionals face is to avoid overfitting. Have you come across a situation where your model performed exceptionally well on train data, but was not able to predict test data. Or you were on the top of a competition in public leaderboard, only to fall hundreds of places in the final rankings? Well...

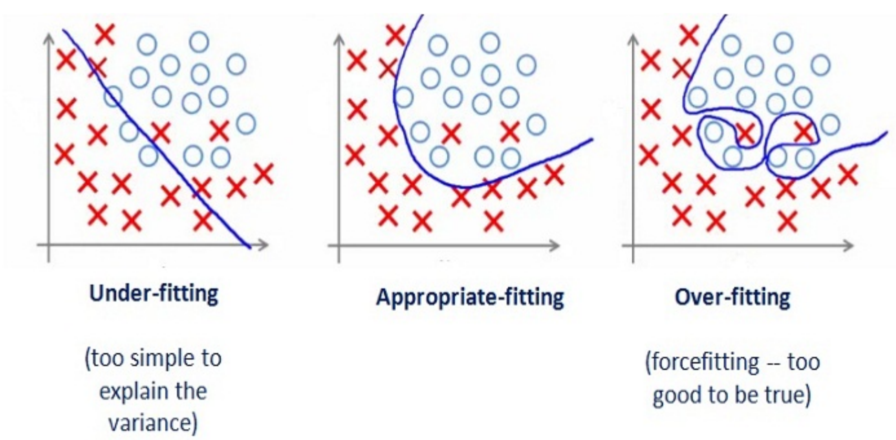
Let's consider an example here, we have 10 students in a classroom. We intend to train a model based on their past score to predict their future score. There are 5 females and 5 males in the class. The average score of females is 60 whereas that of males is 80. The overall average of the class is 70.

Now, there are several ways to make the prediction:

- Predict the score as 70 for the entire class.
- Predict score of males = 80 and females = 60. This a simplistic model which might give a better estimate than the first one.
- Now let's try to overkill the problem. We can use the roll number of students to make a prediction and say that every student will exactly score same marks as last time. Now, this is unlikely to be true and we have reached such granular level that we can go seriously wrong.

The first case here is called under fit, the second being an optimum fit and last being an over-fit.

The above data looks like a quadratic trend over independent variable X. A higher degree polynomial might have a very high accuracy on the train population but is expected to fail badly on test dataset. We will briefly touch up on various techniques we use to avoid over-fitting. And then focus on a special technique called Regularization.



1 Regularization

This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.

2 What does Regularization achive ?

A standard least squares model tends to have some variance in it, i.e. this model won't generalize well for a data set different than its training data. Regularization, significantly reduces the variance of the model, without substantial increase in its bias. So the tuning parameter λ, used in the regularization techniques described above, controls the impact on bias and variance. As the value of λ rises, it reduces the value of coefficients and thus reducing the variance. Till a point, this increase in λ is beneficial as it is only reducing the variance(hence avoiding overfitting), without loosing any important properties in the data. But after certain value, the model starts loosing important properties, giving rise to bias in the model and thus underfitting. Therefore, the value of λ should be carefully selected.

References:

For more detailed explaination can refer below:
`https://en.wikipedia.org/wiki/Regularization`
`https://www.analyticsvidhya.com/blog/2015/02/avoid-over-fitting-regularization/`