Hyper-parameters

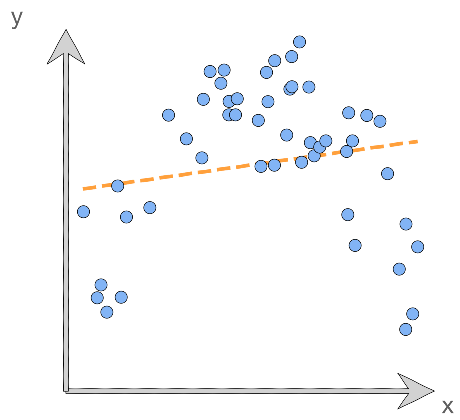
Bookmark this page

Fine tuning hyper-parameters

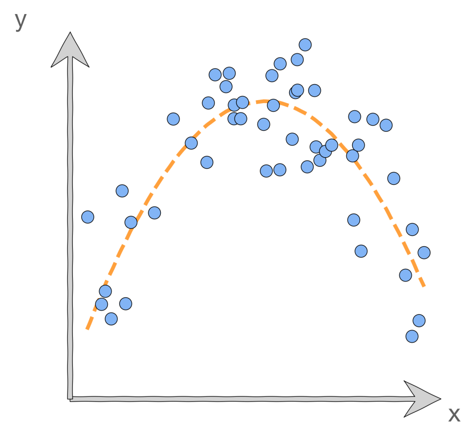
We can recast the problem of model selection as a problem of choosing hyper-parameters.

> **Definition: Hyper-parameter**
>
> A **hyper-parameter** is a parameter of the model that is not itself learned from the data.
>
> For example, polynomial regression requires choosing a degree – this can be thought of as model selection – and so we select a model by choosing a value for the hyper-parameter. This process of choosing a hyper-parameter value is called **tuning**.

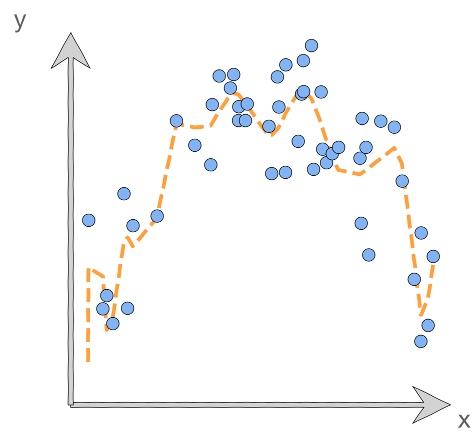
There are some hyper-parameters that we've seen before! For example, here are three models resulting from different choices of the polynomial degree fit on the same data. The degree of the polynomial is the hyper-parameter.



Underfitting: When the degree is too low, the model cannot fit the trend. Training and validation error is high here.

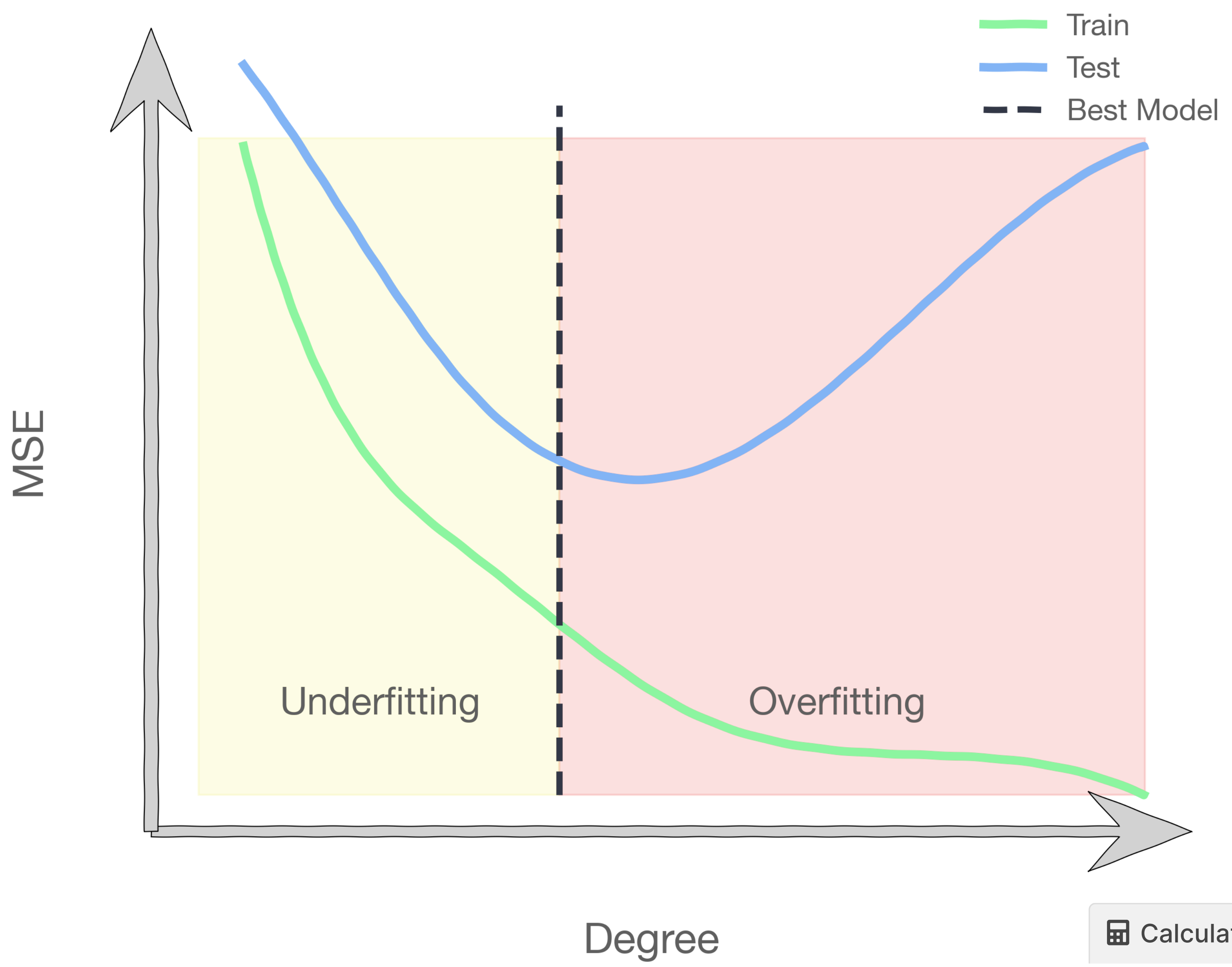


Best Model: We want a model that fits the trend and ignores the noise. Here the validation error is minimized.



Overfitting: when the degree is too high, the model fits all the noisy data points. Here the training error is low, but validation error is high.

We observe that when the degree is too low the model will have high train and validation error. As the degree increases up to a point we see both errors decrease. But eventually the model will start to do worse on the validation data even as the training error continues to drop.

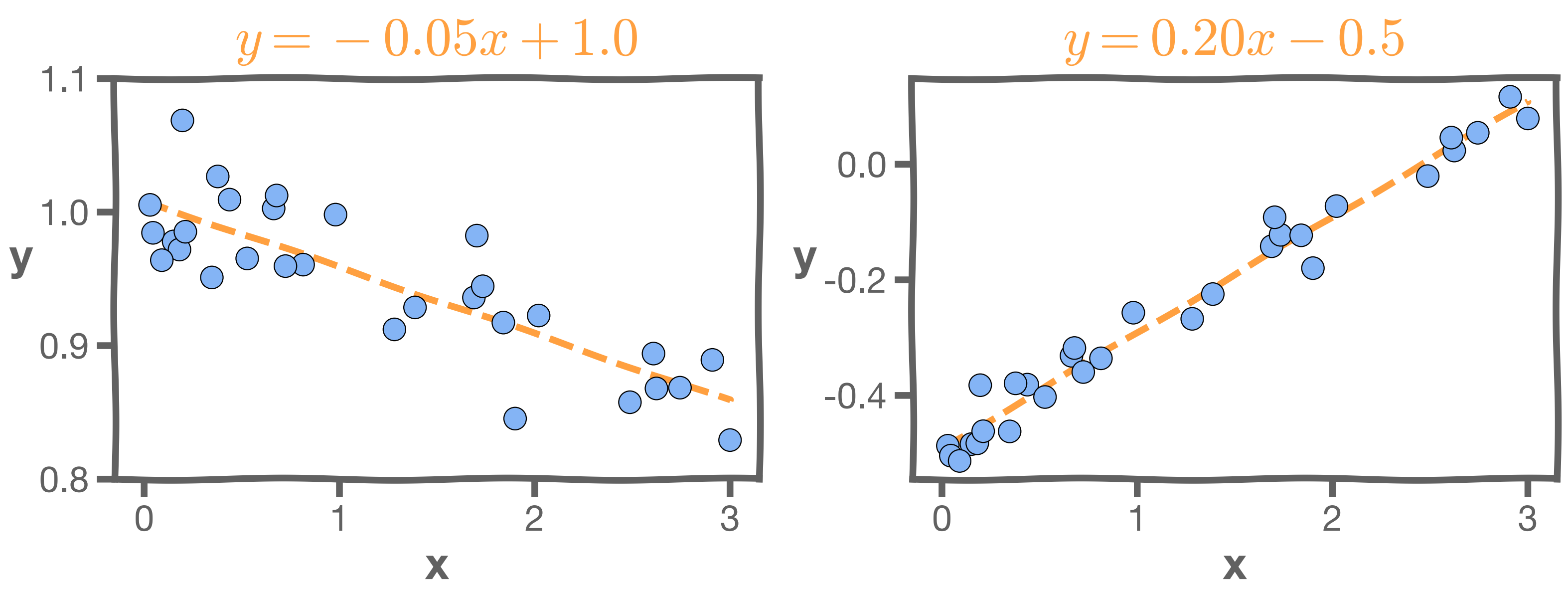


Evaluation: Model Interpretation

In addition to evaluating a model using a metric like MSE, we can also try to interpret the model and make to sure that what the model tells us makes sense.

For linear models we can interpret the parameters which are the β's.

Consider these two models, which you've seen before:



For the model on the left, the MSE of this model is very small. But the slope is -0.05. That means the larger the budget the less the sales. This isn't very likely. For the model on the right, the MSE is very small, but the intercept is -0.5 which means that for very small budget we will have negative sales which is not possible.

When we encounter nonsensical interpretations like these we should reinspect our modeling methodology and the data being used.

Discussion Board (External resource)

Haga clic en Aceptar para que su nombre de usuario y dirección de correo electrónico se envíen a una aplicación de terceros.

Aceptar

< Previous Next >

Connect

Calculator