# Overfitting

If we find a model which is very very well tuned with our training data (quite complex model) but doesn’t generalize well to other observations that is called overfit model.

If we have only “training” and “test” dataset, then we might train our model very specific to the training data, thus make overfit.

In order to prevent such situation, we need to have another set of data for cross-validation.

**Ridge regression:** Systematically finding the reasonable fit model (not overfit) with reasonable coefficient values.

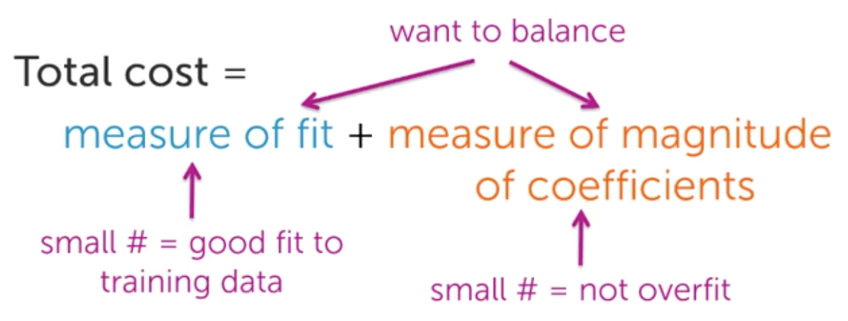
Coefficient values close to zero are more resonable. Huge coefficient values produces wigly and crazy fit which means overfit model. **Ridge regression quantify overfitting through the magnitude of the coefficients.**

**[Quality Matric]**

Previously our quality matric was the difference between our predicted house price vs actual price. Now quality matric also take into account the complexity of our model, in order to bias us towards simpler model.

Desired total cost format:

1. How well our function fits the data (**RSS)**.
2. Magnatude of coefficient.



**How to choose the lambda tuning parameter?**

If we have some tuning parameter (lambda) that controls model complexity, then we can think for every value of that tuning parameter, we can fit our model on our training data. Then we can assess the performance of that fitted model on a validation set, and we can tabulate this for all values of lambda that we might consider. And choose the specific model complexity according to the error on this validation set, and then assess the performance of the selected model on our test set.

# Feature selection

**Lasso regression (L1 regularized regression):** Similar to “ridge regression” solution is governed by a continious parameter (lambda).

Here with this tuning parameter we measure/tune performance of the model for certain set of features.

