5.3 Training Error

May 9, 2019

1 5.3 Training Error

In the previous sections, we learned to build regression models. In this section, we will learn one way to evaluate the quality of a regression model: the training error. We will also discuss the shortcomings of using training error to measure the quality of a regression model.

In [1]:
```
%matplotlib inline
import numpy as np
import pandas as pd
pd.options.display.max_rows = 5

housing = pd.read_csv("https://raw.githubusercontent.com/dlsun/data-science-book/master
                      sep="\t")
housing
```

Out[1]:

	Order	PID	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	...
0	1	526301100	20	RL	141.0	31770	Pave	
1	2	526350040	20	RH	80.0	11622	Pave	
...	...	...	...	...	...	...	...	
2928	2929	924100070	20	RL	77.0	10010	Pave	
2929	2930	924151050	60	RL	74.0	9627	Pave	

	Alley	Lot Shape	Land Contour	...	Pool Area	Pool QC	Fence	...
0	NaN	IR1	Lvl	...	0	NaN	NaN	
1	NaN	Reg	Lvl	...	0	NaN	MnPrv	
...	...	...	...	...	...	...	...	
2928	NaN	Reg	Lvl	...	0	NaN	NaN	
2929	NaN	Reg	Lvl	...	0	NaN	NaN	

	Misc Feature	Misc Val	Mo Sold	Yr Sold	Sale Type	Sale Condition	...
0	NaN	0	5	2010	WD	Normal	
1	NaN	0	6	2010	WD	Normal	
...	...	...	...	...	...	...	
2928	NaN	0	4	2006	WD	Normal	
2929	NaN	0	11	2006	WD	Normal	

	SalePrice