HW3

N14855366 Tianyuunyang Wang

1

- **(a)**The model class is linear and no under-modeling.
  $\beta =${1,2,0}
- **(b)**The model class is not linear and no under-modeling
  $\beta =${3,3,2,3}
- **(c)**The model classis nonlinear and under-modeling. Because there is no xixa parameter in the model class.

2.

(a). $\hat{y} = f(x, \beta) = \beta_0 + \beta_1 x$

$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$ $\quad\quad$ $\hat{\beta}_1 = S_{xy} / S_x^2$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. $\quad\quad$ $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $\quad\quad$ $S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

(b). $y_1 = f_0(x) = \beta_{00} + \beta_{01} x + \beta_{02} x^2$.

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\quad\quad$ $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \beta_{00} + \beta_{01} x_i + \beta_{02} x_i^2$.

$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $\quad\quad$ $S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$.

$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$ $\quad\quad$ $\hat{\beta}_1 = \frac{S_{xy}}{S_x^2}$

2.(c)(d)

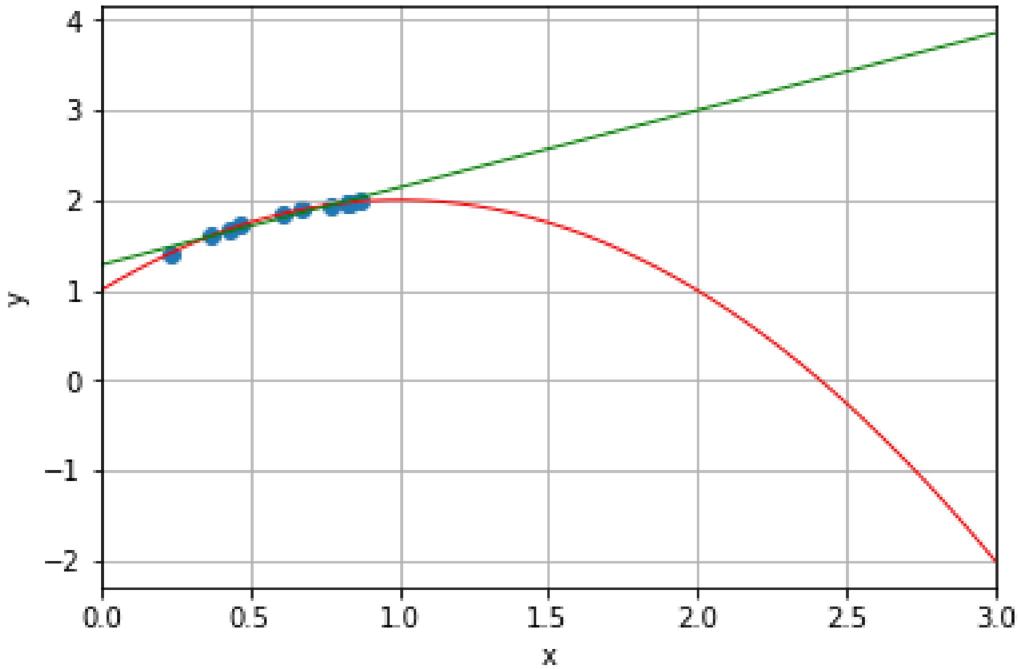
February 14, 2018

In [7]:
```
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
import numpy.polynomial.polynomial as poly
```

In [8]:
```
beta=np.array([1,2,-1])
dtrue=len(beta)-1
nsamp=10
xdat=np.random.uniform(0,1,nsamp)
```

In [11]:
```
xp=np.linspace(0,3,300)
yp=poly.polyval(xp,beta)
ydat=poly.polyval(xdat,beta,dtrue)
plt.plot(xp,yp,'r-',linewidth=1)
beta_hat=poly.polyfit(xdat,ydat,1)
yp_hat=poly.polyval(xp,beta_hat)
plt.plot(xp,yp_hat,'g-',linewidth=1)

plt.xlim(0,3)
plt.scatter(xdat,ydat)
plt.xlabel('x')
plt.ylabel('y')
plt.grid()
plt.show()
```




```
In [13]: #(d)
         #The largest bias is at x=3.
         #Because when x=(0,1) the linear is estimate equal to square.
         #When x get larger the bias get larger.
```

3. (a) x_1 : cancer volume. x_2 : patients age x_3 : cancer type.

$$x_3 = \begin{cases} 1 & \text{type I} \\ 0 & \text{type II} \end{cases}$$

Model 1: $\hat{y} = \beta_0 + \beta_1 x_1$

Model 2: $\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

Model 3: $\hat{y} = \beta_0 + \beta_1 x_1 x_3 + \beta_2 x_1 (1 - x_3) + \beta_3 x_2$

one-hot code.

	cancer_volume1 : x_{31}	cancer_volume2 : x_{30}	patient_age : x_2
Type 1	1	0	1
Type 2	0	1	1.

(b). Parameters

	Parameters
Model 1	2
Model 2	3
Model 3.	4

Model 3 is the most complex

(c) Model 1:

$$A = \begin{bmatrix} 1 & 0.7 \\ 1 & 1.3 \\ 1 & 1.6 \\ \vdots & \end{bmatrix}$$

Model 2:

$$A = \begin{bmatrix} 1 & 0.7 & 55 \\ 1 & 1.3 & 65 \\ 1 & 1.6 & 70 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

Model 3:

$$A = \begin{bmatrix} 1 & 0.7 & 0 & 55 \\ 1 & 0 & 1.3 & 65 \\ 1 & 0 & 1.6 & 70 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

(d) Least ~~test~~ error RSS = 0.70 (Model 3)
std deviation = 0.05.

$SE = \cancel{RSE}\ std/\sqrt{k-1} = 0.05/\sqrt{9} \approx 0.0167.$

$MSE\_[G] = mse\_mean + MSE\_std = 0.7167 < 0.72.$

∴ We select model 3 based on one standard error rule.