Homework 3

Jingyi Lou

1.

1. (a) the model is not linear and there is no under-modeling. $\beta = [1, 2, 0]$

(b) the model is not linear and there is no under-modeling

$a_0 = 3,\ a_1 = 3,\ b_0 = 2,\ b_1 = 3$

(c) the model is not linear and it is under-modeling.

2. (a) $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$

$S_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N}\sum x_i y_i - \bar{x}\bar{y}$

$S_x^2 = \frac{1}{N}\sum (x_i - \bar{x})^2 = \frac{1}{N}\sum x_i^2 - \bar{x}^2$

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

(b) $\bar{x} = \frac{1}{N}\sum x_i \qquad \bar{y} = \frac{1}{N}\sum y_i = \frac{1}{N}\sum f_0(x_i)$

$S_{xy} = \frac{1}{N}\sum (x_i - \bar{x})\left(f_0(x_i) - \frac{1}{N}\sum f_0(x_i)\right)$

$S_{x^2} = \frac{1}{N}\sum x_i^2 - \bar{x}^2$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{x^2}}, \quad \hat{\beta}_0 = \bar{y} - \beta_1 \bar{x} = \frac{1}{N}\sum f_0(x_i) - \beta_1 \bar{x}$

2 (c)

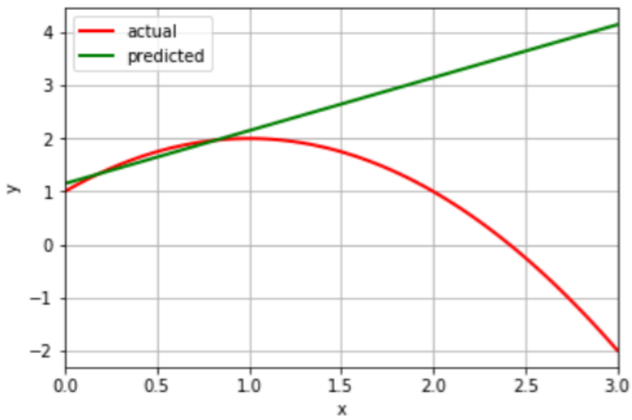
In [8]:
```
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
from sklearn import datasets, linear_model, preprocessing
import numpy.polynomial.polynomial as poly
%matplotlib inline


beta = np.array([1, 2, -1])
nsamp = 10
xdat = np.linspace(0, 1, nsamp)
ydat = poly.polyval(xdat, beta)
d = 1
beta_hat = poly.polyfit(xdat, ydat, d)
```

```
[ 1.14814815  1.        ]
```

In [14]:
```
xp = np.linspace(0, 3, 100)
yp = poly.polyval(xp, beta)
yp_hat = poly.polyval(xp, beta_hat)
```

In [21]:
```
plt.xlim(0, 3)
plt.plot(xp, yp, 'r-', linewidth=2)
plt.plot(xp, yp_hat, 'g-', linewidth=2)
plt.legend(['actual', 'predicted'], loc = 'upper left')
plt.grid()
plt.xlabel('x')
plt.ylabel('y')
```

Out[21]: `<matplotlib.text.Text at 0x1151fc400>`



2(d)

In [22]:
```
bias = np.zeros(100)
for i in range(100):
    bias[i] = (yp_hat[i] - yp[i])**2
bias_order = np.argsort(-bias)
print(xp[bias_order[0]])
```

```
3.0
```

3.

3. (a)

Model 1 : cancer volume : x $\hat{y} = \beta_0 + \beta_1 x$

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

Model 3 : cancer I volume : x_1 , cancer II volumn : x_2 , age : x_3

	Cancer I volumn (x_1)	Cancer II volumn (x_2)	Age (x_3)
Type I	1	0	1
Type II	0	1	1

y_hat = $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

(b)

Model 1 : 2 parameters

Model 2 : 3 parameters

Model 3 : 4 parameters

Model 3 is the most ~~p~~ complex one.

(c) Model 1 : $A = \begin{bmatrix} 1 & 0.7 \\ 1 & 1.3 \\ 1 & 1.6 \end{bmatrix}$

Model 2 : $A = \begin{bmatrix} 1 & 0.7 & 55 \\ 1 & 1.3 & 65 \\ 1 & 1.6 & 70 \end{bmatrix}$

Model 3 : $A = \begin{bmatrix} 1 & 0.7 & 0 & 55 \\ 1 & 0 & 1.3 & 65 \\ 1 & 0 & 1.6 & 70 \end{bmatrix}$

d) Model 2