Answer -1

1(a) $f_0(x) = 1+2x$, $f(x,\beta) = \beta_0 + \beta_1 x + \beta_2 x^2$

(i) The model is linear

(ii) However, it is under modeling.

If it was $f(x,\beta) = \beta_0 + \beta_1 x$ then it ~~was~~ would have been ideal and the true parameters would have been $\beta = (1, 2)$

1(b) $f_0(x) = 1 + \frac{1}{(2+3x)}$, $f(x, a_0, a_1, b_0, b_1) = \frac{(a_0 + a_1 x)}{(b_0 + b_1 x)}$

(i) The model is not linear since it has [$1/x$] component

(ii) The model is ~~not~~ under modeling.

1(c) $f_0(x) = (x_1 - x_2)^2$ and $f(x, a, b_1, b_2, c_1, c_2)$

$= a + b_1 x_1 + b_2 x_2 + c_1 x_1^2 + c_2 x_2^2$

(i) The model is not linear

(ii) However it is not under modeling. ~~True par~~

True parameters $= (a, b_1, b_2, c_1, c_2)$

Answer -2

```
# 2(a) For splitting training and testing
from sklearn.model_selection import train_test_split

xtr,xts,ytr,yts=train_test_split(X,y,test_size=0.5)



#2(b)

dtest=np.array(range(1,10))
RSStest=[]
RSStr=[]
for d in test:
  #Fitting the data
  beta_hat=poly.polyfit(xtr,ytr,d)
  #Measure the RSS on the training data
  yhat=poly.polyval(xtr,beta_hat)
  RSSd=np.mean((yhat-ytr)**2)
  RSStr.append(RSSd)
  #Measure RSS on test data
  yhat=poly.polyval(xts,beta_hat)
  RSSd=np.mean((yhat-yts)**2)
  RSStest.append(RSSd)
plt.plot(dtest,RSStr,'bo-')
plt.plot(dtest,RSStest,'go-')
plt.xlabel('Model Order')
plt.ylabel('RSS')
plt.grid()
plt.legend(['Training','Test'],loc='upper left')





#2(c) Order with lowest mean squared error
imin=np.arg.min(RSStest)
print("Estimated model order={0:d}".format(dtest[imin]))
```

Answer-3

3(a) Actual value $y = f_0(x_{test}) + \epsilon$

where ϵ is the noise on the training & test data

Output mean squared error:

$$MSE_y(x_{test}) = \underbrace{E}_{\text{Expectation}}(y - \hat{y})^2$$

$$MSE_y(x_{test}) = MSE_f(x_{test}) + \sigma_\epsilon^2$$

Since no noise $\sigma_\epsilon^2 = 0$

$$MSE_y(x_{test}) = E(f_0(x_{test}) - f(x_{test}, \hat{\beta})]^2$$

$$= \beta x^2 - \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i^2}$$

(b) ~~The training data.~~ Given $Bias = E(f(x,\hat{\beta}) - f(x,\beta_0))$

$$Bias(x) = E(f(x,\hat{\beta}) - \beta_0 x^2]$$

$$= E(\hat{\beta} x^2) - \beta_0 x^2$$

$$Bias(x) = \hat{\beta} E(x^2) - \beta_0 x^2$$

The training data is $y_i = f(x_i, \beta_0) + \epsilon_i$

3(b) where noise is $\epsilon_i \sim N(0, \sigma^2)$

The noise is now external.

$$MSE_y(x_{test}) = E\left[f_0(x_{test}) \cancel{f_0(x_{test})} + \epsilon - f(x_{test}, \hat{\beta})\right]^2$$

$$Bias(x) = E(f(x,\hat{\beta})) - f(x, \beta_0)$$

$$= E(f(x,\hat{\beta}) + E(\epsilon) - \beta_0 x^2 \rightarrow \text{True parameter}$$

$$= \beta E(x^2) + \underbrace{E(\epsilon)}_{\text{it is the noise}} - \beta_0 x^2$$

Since $E(\epsilon) = 0$ → given in slide number 17.
Noise on test sample is independent of $\hat{\beta}$ & x_{test}.

$$\therefore \boxed{Bias(x) = [\hat{\beta} E(x^2)] - \beta_0 (x^2)}$$

~~3(c) The training data is $y_i = f(x_i, \beta_0)$ where $\epsilon_i \sim N(0, \sigma^2)$~~

3(c) The training data is $y_i = f(x_i + \epsilon_i, \beta_0)$ where the

$$Bias(x) = E(f(x+\epsilon), \hat{\beta}) - f(x, \beta_0)$$

$$= E(f(x+\epsilon), \hat{\beta}) - \beta_0 x^2$$

$$= E(f(x)\hat{\beta} + E(\epsilon, \hat{\beta}) - \beta_0 x^2$$

$$\downarrow$$

$$\boxed{E(\epsilon) = 0}$$

$$= E[f(x, \beta_0)] - \beta_0 x^2$$

$$\boxed{= \beta_0 E(x^2) - \beta_0 x^2}$$ Ans

Homework-4

Answer - 4

$\boxed{4(a)}$ $MSE_{(f)}(x_{test}) = E[f_0(x_{test}) - f(x_{test}, \hat{\beta})]^2$

$\downarrow$

Function MSE

Bias: $Bias(x_{test}) = E[f(x_{test}, \hat{\beta})] - f_0(x_{test})$

Variance: $Var(x_{test}) = E[f(x_{test}, \hat{\beta}) - E[f(x_{test}, \hat{\beta})]]^2$

$MSE_f(x_{test}) = Bias(x_{test})^2 + Var(x_{test})$

$MSE_f(x_{test}) = M_1 + M_2 - 2M_3$

$M_1 = E[f_0(x_{test}) - \bar{f}(x_{test})]^2 = [f_0(x_{test}) - \bar{f}(x_{test})]^2$

$M_2 = E[f(x_{test}, \hat{\beta}) - \bar{f}(x_{test})]^2$

$M_3 = E[(f_0(x_{test}) - \bar{f}(x_{test}))(f(x_{test}, \hat{\beta}) - \bar{f}(x_{test}))]$

$= (f_0(x_{test}) - \bar{f}(x_{test}))E[f(x_{test}, \hat{\beta}) - \bar{f}(x_{test})]$

$= (f_0(x_{test}) - \bar{f}(x_{test}))(f(x_{test}) - \bar{f}(x_{test})) = 0.$

IN THE USA

Blank Page- Continued Next Page

Please Turn Over

Blank Page- Continued Next Page

Please Turn Over

5(a)

Let Cancer volume be 'v'
Age be 'a'
Type of cancer be 'T'

Model - 1 : $y = \beta_0 + \beta_1 v$

Model - 2 : $y = \beta_0 + \beta_1 v + \beta_2 a$.

Model - 3 : using one hot encoding. for Type of cancer

~~Type — [0, 1]~~

Type - 1 = 001
Type - 2 = 010 → one hot encoding

For type 1

$$\boxed{y = \beta_0 + \beta_1 v + \beta_2 a + \beta_3 . T . v + \beta_4 . T . v}$$

For type - 2.

5(b) Model - 1 $y = \beta_0 + \beta_1 v$

So two parameters : β_0 → The intercept
β_1 → The coefficient of cancer volume.

Model - 2 $y = \beta_0 + \beta_1 v + \beta_2 a$

Blank Page- Continued Next Page

Please Turn Over

It has three parameters

$\beta_0 \rightarrow$ The intercept
$\beta_1 \rightarrow$ The coefficient of cancer volume
$\beta_2 \rightarrow$ The coefficient of age.

<u>Model-3</u> It has 4 parameters

$\beta_0 \rightarrow$ intercept
$\beta_1 \rightarrow$ coefficient of volume
$\beta_2 \rightarrow$ coefficient of age
$\beta_3 \rightarrow$ dependence on type of cancer I ~~on~~ and cancer volume
$\beta_4 \rightarrow$ depence of type of cancer II and cancer volume

<u>I think. model-3 is most complex</u>

5(c) Since $y = \beta_0 + \beta_1 a$ for Model-1

A for Model 1 $= \begin{bmatrix} 0.7 & 1 \\ 1.3 & 1 \\ 1.6 & 1 \end{bmatrix}$ since age is not considered in Model-1

A for Model 2 $= \begin{bmatrix} 0.7 & 55 & 1 \\ 1.3 & 65 & 1 \\ 1.6 & 70 & 1 \end{bmatrix}$ since age & volume are important for ch.

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

Type-1 Type-2

Blank Page- Continued Next Page

Please Turn Over

5(d) Based on "one standard error rule",

$\boxed{\text{Model -3}}$ is the best.

One standard error – Find model with minimum error, then select the simplest model whose mean falls within 1 std. deviation of minimum

one SE $\hat{p} = \min \{ p \mid \bar{S}[p] \leq S_{tgt} \}$

upper Bound of Model 1 = 2.01 + 0.03 = 2.04

upper bound of Model 2 = 0.72 + 0.04 = 0.76

upper bound of Model 3 = 0.70 + 0.05 = 0.75

2.04 >	0.76 >	0.75
Model 1	Model 2	Model -3

Hence according to one standard error rule Model 3 is best