```
%-------------------- Group 56 ---------------------
%-CH19B012-SUJITH KUMAR-----------------------------
%-CH19B052-CH GOVARDHANA SAI SRINIVAS---------------
%-CH19B055-G BHUVAN CHANDRA---------------------------

%-----------------Machine Learning-----------------

fprintf("---------------Machine Learning-----------\n\n")
%------------Question 1(Conceptual)----------------
for i=1:5
 fprintf('-----Q1 V%d (Question 1 , Variant %d)--------\n\n',i,i)
 fprintf("Which of the following statements are true ?\n\n")

 C= randi([2,4]);
 C1=C+1;
 C2=C-1;

 ic=randi([2,3]);
 B=ic-1;
 D=ic+1;


 Crct=["It is possible to design a Linear regression algorithm using a neural network."...
    ,"If the size of training data increases , Bias increases and Variance decreases."...
    "For the model with high variance the cost function or squared error function (Jθ) will be low."...
    "Traning neural network has the potential of overfitting the training data"...
    "If the modal bias and variance both are low,the modal will have higher accuracy "];

 Incrct=["Overfitting is more likely when you have huge amount of data to train"...
    ,"Logistic Regression is used for predicting continuous dependent variable"...
    "MLE estimates are often desirable because they have low variance"...
    'Using a model with high bias is always better than using a model with less bias'];

EIC1="False.With a small training dataset, it’s easier to find a hypothesis to fit the training data exactly i.e. overfitting.";
EIC2="False.Logistic Regression is used for classification problems.Hence it is not used for predicting continuous dependent variable.";
EIC3="False.Variance in Maximum likelihood Estimate(MLE) is high.High variance indicated measurement uncertainity hence they are undesirable";
EIC4="False.Bias is the difference between your model's expected predictions and the true values.Low bias algorithms trains model that are accurate on average";

EC1="True.Neural network can be used as a universal approximator, so it can definitely implement a linear regression algorithm.";
EC2="True.As we increase the size of the training data, the bias would increase while the variance would decrease.";
EC3="True.For model with high variance the hypothesis function fits the training data very well which causes the error to be low.";
EC4="True.Overfitting of the training data happens if neural network model is suffering from high variance.'It means the trained parameters fits the training set we
EC5="True.The statement is self explanatory";

CrctExp=[EC1,EC2,EC3,EC4,EC5];
IncrctExp=[EIC1,EIC2,EIC3,EIC4];

fprintf(['A) %s\n\n'...
         'B) %s \n\n'...
         'C) %s \n\n'...
         'D) %s \n\n'...
         'E)None of these\n\n'],Incrct(B),Crct(C1),Crct(C2),Incrct(D))
fprintf("Answer : B C\nExplanation : \n\n")
fprintf(['A) %s \n\n'...
         'B) %s \n\n'...
         'C) %s \n\n'...
         'D) %s \n\n'],IncrctExp(B),CrctExp(C1),CrctExp(C2),IncrctExp(D))
clear
end

%------------Question 2(Numerical)----------------
%--------Can produce any number of variants --------

for i=1:5
fprintf('-----Q2 V%d (Question 2 , Variant %d)-----\n\n',i,i)
p=randi([8,12]);
var=['c','a','b'];
v=randi([1,3]);
step=randi([1,2]);
t=0:step:p;
n=size(t,2);
h=rand(1,n);
fprintf(['The water level in the North sea is mainly determinedby so called M2 tide whose period is about %d hours.\n'...
    'The height H(t) thus roughly taken the form \n\t\t H(t)= c + a sin(2πt/%d) + b cos(2πt/%d)\nUse method of least squares to find %s\n'],p,p,p,var(v))
T = table(t(:),h(:),'VariableNames',{'t(hours)','H(t)(meters)'});
disp(T)
sine=sin(2*pi.*t/p);
cosine=cos(2*pi.*t/p);
X=[ones(n,1) sine' cosine'];
thetha= inv((X'*X))*X'*h';
Ans=thetha(v);
options=[Ans+0.25 Ans Ans+0.5 Ans-0.5];

    o=randi([1,4]);
    oa=o+1;
    if(oa >4)
        oa=oa-4;
    end
    ob=o-1;
```

```
    if(ob<1)
        ob=ob+4;
    end
    oc=o;
    od=o-2;
    if(od<1)
        od=od+4;
    end

    if(options(oa) == Ans)
        answer='A';
    elseif(options(ob) == Ans)
        answer='B';
    elseif(options(oc) == Ans)
        answer='C';
    else
        answer='D';
    end
fprintf(['A)%.4f\n'...
         'B)%.4f\n'...
         'C)%.4f\n'...
         'D)%.4f\n'...
         'E)None of these\n'],options(oa),options(ob),options(oc),options(od))
 fprintf('Answer: %s\nSolution :\n',answer)
 fprintf(['Given \n\t H(t)= c + a sin(2πt/%d) + b cos(2πt/%d)\n Using least square method Sr = Σ(hi - (c + a*sin(2πti/%d) + b*cos(2πti/%d)))^2 \n'...
          'To find best possible values of a,b,c equate\n\t del Sr/del a = del Sr/del b =del Sr/del c =0\nOn solving the three equations\n'...
          'we get a= %.4f , b = %.4f , c =%.4f\n\n'],p,p,p,p,thetha(2),thetha(3),thetha(1))
clear
end


%------------Question 3(Conceptual)-----------------
%---------------Machine Learning----------

for i=1:5
 fprintf('-----Q3 V%d (Question 3 , Variant %d)-----\n\n',i,i)
 fprintf("Which of the following statements are Incorerct about PCA(Principal Component Analysis)?\n\n")

 C= randi([2,4]);
 C1=C+1;
 C2=C-1;

 ic=randi([2,3]);
 B=ic-1;
 D=ic+1;


 Crct=["Even if all the input features are on very similar scales, we should still perform mean normalization (so that each feature has zero mean) before running PC
    ,"Given an input x ∈ R^n, PCA compresses it to a lower-dimensional vector z ∈ R^k."...
    "If the input features are on very different scales, it is a good idea to perform feature scaling before applying PCA."...
    "Given input data x ∈ R^n, it makes sense to run PCA only with values of k that satisfy k <= n. where k is the dimension to which the input data reduced"...
    "All principal components are orthogonal to each other and Maximum number of principal components <= number of features"];

 Incrct=[" PCA is susceptible to local optima; trying multiple random initializations may help."...
    ,"when the features reduces to lower dimensions using PCA,the features carries all information present in data"...
    "PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D)."...
    "PCA will perform outstandingly  when eigenvalues are roughly equal"];

EIC1="False.PCA is a deterministic algorithm which doesn’t have local minima problem like most of the machine learning algorithms has.";
EIC2="False.When the features reduces to lower dimensions ,most of the times some information of data will lose and won’t be able to interpret the lower dimension d
EIC3="False.PCA can be used to reduce the dimensionality of data to any dimensions less than the given dimension";
EIC4="False. When all eigen vectors are same in such case you won’t be able to select the principal components because in that case all principal components are equ


EC1="True.If you do not perform mean normalization, PCA will rotate the data in a possibly undesired way.";
EC2="True.PCA compresses given input to a lower dimensional vector by projecting it onto the learned principal components";
EC3="True.Feature scaling prevents one feature dimension from becoming a strongvprincipal component only because of the large magnitude of the feature values (as op
EC4="True.With k = n, there is no compression, so PCA has no use and k > n does not make sense.";
EC5="True.The statement is Self explanatory.";

CrctExp=[EC1,EC2,EC3,EC4,EC5];
IncrctExp=[EIC1,EIC2,EIC3,EIC4];

fprintf(['A) %s\n\n'...
         'B) %s \n\n'...
         'C) %s \n\n'...
         'D) %s \n\n'...
         'E)None of these\n\n'],Crct(C1),Incrct(B),Crct(C2),Incrct(D))
fprintf("Answer : A C\nExplanation : \n\n")
fprintf(['A) %s \n\n'...
         'B) %s \n\n'...
         'C) %s \n\n'...
         'D) %s \n\n'],CrctExp(C1),IncrctExp(B),CrctExp(C2),IncrctExp(D))
clear
end

%------------Question 4(Numerical)-----------------
%--------Can produce any number of variants --------
for i=1:5

s=10:30;
```

```
s1=1:20;
r=randi([1,20]);
n=s(r)*1000;

n1=s1(r)*1000;
n2=n-n1;

FP=randi([50,n1]);
TN=n1-FP;

TP=randi([50,n2]);
FN=n2-TP;

fprintf('-----Q4 V%d (Question 4 , Variant %d)-----\n\n',i,i)
fprintf(['Suppose %d patients get tested for flu out of them %d are actually healthy and %d are actually sick.\n'...
         'For the sick people a test was positive for %d and negative for %d.For healthy people, the same test was\n'...
         'positive for %d and negative for %d.Calculate the F1 score for the data\n'],n,n1,n2,TP,FN,FP,TN);

precision = TP/(TP+FP);
recall= TP/(FN+TP);
F1score= 2*(precision*recall)/(recall+precision);

options=[precision*recall 0.25*precision*recall F1score/4 F1score];

    o=randi([1,4]);
    oa=o+1;
    if(oa >4)
        oa=oa-4;
    end
    ob=o-1;
    if(ob<1)
        ob=ob+4;
    end
    oc=o;
    od=o-2;
    if(od<1)
        od=od+4;
    end

    if(options(oa) == F1score)
        answer='A';
    elseif(options(ob) == F1score)
        answer='B';
    elseif(options(oc) == F1score)
        answer='C';
    else
        answer='D';
    end
fprintf(['A)%.2f\n'...
         'B)%.2f\n'...
         'C)%.2f\n'...
         'D)%.2f\n'...
         'E)None of these\n'],options(oa),options(ob),options(oc),options(od))
fprintf('Answer: %c \nSolution :\n',answer)
T = table([TP ; FN ],[FP ; TN ],'VariableNames',{'No of Actual sick','No of Actual Healthy '},'RowName',{'No of predicted sick','No of predicted Healthy'});
disp(T)
fprintf('\t\tTrue Positives(TP)  : %d\t False Positives(FP) : %d\n\t\tFalse Negatives(FN) : %d\t True Negatives(TN)  : %d\n',TP,FP,FN,TN)
fprintf(['Precision quantifies the number of positive class predictions that actually belong to the positive class.\n'...
         'Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.\n'...
         'F1 score provides a single score that balances both the concerns of precision and recall in one number'])
 fprintf(['Therefore\n\t\tprecision = TP/(TP+FP) = %d/(%d + %d) = %.2f\n'...
          '\t\t\trecall= TP/(TP+FN) = %d/(%d + %d)= %.2f\n'],TP,TP,FP,precision,TP,TP,FN,recall)
 fprintf('\t\t\tF1score= 2*(precision*recall)/(recall+precision)\n\t\t\t\t   = %.2f\n\n',F1score)
 clear
 end
```

```
---------------Machine Learning-----------

-----Q1 V1 (Question 1 , Variant 1)--------

Which of the following statements are true ?

A) Overfitting is more likely when you have huge amount of data to train

B) For the model with high variance the cost function or squared error function (Jθ) will be low.

C) It is possible to design a Linear regression algorithm using a neural network.

D) MLE estimates are often desirable because they have low variance

E)None of these

Answer : B C
Explanation :

A) False.With a small training dataset, it’s easier to find a hypothesis to fit the training data exactly i.e. overfitting.

B) True.For model with high variance the hypothesis function fits the training data very well which causes the error to be low.

C) True.Neural network can be used as a universal approximator, so it can definitely implement a linear regression algorithm.
```

D) False.Variance in Maximum likelihood Estimate(MLE) is high.High variance indicated measurement uncertainity hence they are undesirable

-----Q1 V2 (Question 1 , Variant 2)--------

Which of the following statements are true ?

A) Logistic Regression is used for predicting continuous dependent variable

B) For the model with high variance the cost function or squared error function (Jθ) will be low.

C) It is possible to design a Linear regression algorithm using a neural network.

D) Using a model with high bias is always better than using a model with less bias

E)None of these

Answer : B C
Explanation :

A) False.Logistic Regression is used for classification problems.Hence it is not used for predicting continuous dependent variable.

B) True.For model with high variance the hypothesis function fits the training data very well which causes the error to be low.

C) True.Neural network can be used as a universal approximator, so it can definitely implement a linear regression algorithm.

D) False.Bias is the difference between your model's expected predictions and the true values.Low bias algorithms trains model that are accurate on average

-----Q1 V3 (Question 1 , Variant 3)--------

Which of the following statements are true ?

A) Logistic Regression is used for predicting continuous dependent variable

B) If the modal bias and variance both are low,the modal will have higher accuracy

C) For the model with high variance the cost function or squared error function (Jθ) will be low.

D) Using a model with high bias is always better than using a model with less bias

E)None of these

Answer : B C
Explanation :

A) False.Logistic Regression is used for classification problems.Hence it is not used for predicting continuous dependent variable.

B) True.The statement is self explanatory

C) True.For model with high variance the hypothesis function fits the training data very well which causes the error to be low.

D) False.Bias is the difference between your model's expected predictions and the true values.Low bias algorithms trains model that are accurate on average

-----Q1 V4 (Question 1 , Variant 4)--------

Which of the following statements are true ?

A) Logistic Regression is used for predicting continuous dependent variable

B) Traning neural network has the potential of overfitting the training data

C) If the size of training data increases , Bias increases and Variance decreases.

D) Using a model with high bias is always better than using a model with less bias

E)None of these

Answer : B C
Explanation :

A) False.Logistic Regression is used for classification problems.Hence it is not used for predicting continuous dependent variable.

B) True.Overfitting of the training data happens if neural network model is suffering from high variance.'It means the trained parameters fits the training set well

C) True.As we increase the size of the training data, the bias would increase while the variance would decrease.

D) False.Bias is the difference between your model's expected predictions and the true values.Low bias algorithms trains model that are accurate on average

-----Q1 V5 (Question 1 , Variant 5)--------

Which of the following statements are true ?

A) Overfitting is more likely when you have huge amount of data to train

B) For the model with high variance the cost function or squared error function (Jθ) will be low.

C) It is possible to design a Linear regression algorithm using a neural network.

D) MLE estimates are often desirable because they have low variance

E)None of these

Answer : B C
Explanation :

A) False.With a small training dataset, it's easier to find a hypothesis to fit the training data exactly i.e. overfitting.

B) True.For model with high variance the hypothesis function fits the training data very well which causes the error to be low.

C) True.Neural network can be used as a universal approximator, so it can definitely implement a linear regression algorithm.

D) False.Variance in Maximum likelihood Estimate(MLE) is high.High variance indicated measurement uncertainity hence they are undesirable

-----Q2 V1 (Question 2 , Variant 1)-----

The water level in the North sea is mainly determinedby so called M2 tide whose period is about 12 hours.
The height H(t) thus roughly taken the form

$$H(t)= c + a \sin(2\pi t/12) + b \cos(2\pi t/12)$$

Use method of least squares to find c

t(hours)	H(t)(meters)
0	0.74476
1	0.57641
2	0.068241
3	0.14423
4	0.39048
5	0.89642
6	0.84768
7	0.19338
8	0.7095
9	0.70867
10	0.30942
11	0.47339
12	0.35412

A)0.0007
B)0.5007
C)1.0007
D)0.7507
E)None of these
Answer: B
Solution :
Given

$$H(t)= c + a \sin(2\pi t/12) + b \cos(2\pi t/12)$$

Using least square method $Sr = \Sigma(hi - (c + a*\sin(2\pi ti/12) + b*\cos(2\pi ti/12)))^2$
To find best possible values of a,b,c equate
del Sr/del a = del Sr/del b =del Sr/del c =0
On solving the three equations
we get a= -0.1078 , b = -0.0922 , c =0.5007

-----Q2 V2 (Question 2 , Variant 2)-----

The water level in the North sea is mainly determinedby so called M2 tide whose period is about 8 hours.
The height H(t) thus roughly taken the form

$$H(t)= c + a \sin(2\pi t/8) + b \cos(2\pi t/8)$$

Use method of least squares to find b

t(hours)	H(t)(meters)
0	0.96257
1	0.75808
2	0.91236
3	0.85465
4	0.36637
5	0.66706
6	0.92096
7	0.14346
8	0.21698

A)-0.0552
B)-0.5552
C)0.1948
D)0.4448
E)None of these
Answer: A
Solution :
Given

$$H(t)= c + a \sin(2\pi t/8) + b \cos(2\pi t/8)$$

Using least square method $Sr = \Sigma(hi - (c + a*\sin(2\pi ti/8) + b*\cos(2\pi ti/8)))^2$
To find best possible values of a,b,c equate
del Sr/del a = del Sr/del b =del Sr/del c =0
On solving the three equations
we get a= 0.1397 , b = -0.0552 , c =0.6509

-----Q2 V3 (Question 2 , Variant 3)-----

The water level in the North sea is mainly determinedby so called M2 tide whose period is about 12 hours.
The height H(t) thus roughly taken the form

$$H(t)= c + a \sin(2\pi t/12) + b \cos(2\pi t/12)$$

Use method of least squares to find c

t(hours) H(t)(meters)

0	0.66365
2	0.27288
4	0.9565
6	0.9184
8	0.30042
10	0.96105
12	0.19934

A)0.6354
B)0.1354
C)0.8854
D)1.1354
E)None of these
Answer: A
Solution :
Given

H(t)= c + a sin(2πt/12) + b cos(2πt/12)

Using least square method Sr = Σ(hi - (c + a*sin(2πti/12) + b*cos(2πti/12)))^2
To find best possible values of a,b,c equate

del Sr/del a = del Sr/del b =del Sr/del c =0

On solving the three equations
we get a= -0.0093 , b = -0.1756 , c =0.6354

-----Q2 V4 (Question 2 , Variant 4)-----

The water level in the North sea is mainly determinedby so called M2 tide whose period is about 9 hours.
The height H(t) thus roughly taken the form

H(t)= c + a sin(2πt/9) + b cos(2πt/9)

Use method of least squares to find a

t(hours)	H(t)(meters)
0	0.99777
1	0.1372
2	0.70059
3	0.83735
4	0.093919
5	0.74015
6	0.68138
7	0.89261
8	0.43522
9	0.12724

A)-0.1037
B)-0.6037
C)0.1463
D)0.3963
E)None of these
Answer: A
Solution :
Given

H(t)= c + a sin(2πt/9) + b cos(2πt/9)

Using least square method Sr = Σ(hi - (c + a*sin(2πti/9) + b*cos(2πti/9)))^2
To find best possible values of a,b,c equate

del Sr/del a = del Sr/del b =del Sr/del c =0

On solving the three equations
we get a= -0.1037 , b = -0.0495 , c =0.5693

-----Q2 V5 (Question 2 , Variant 5)-----

The water level in the North sea is mainly determinedby so called M2 tide whose period is about 9 hours.
The height H(t) thus roughly taken the form

H(t)= c + a sin(2πt/9) + b cos(2πt/9)

Use method of least squares to find b

t(hours)	H(t)(meters)
0	0.90284
1	0.31006
2	0.64503
3	0.32871
4	0.21836
5	0.71593
6	0.35366
7	0.44538
8	0.81923
9	0.20842

A)0.0837
B)-0.4163
C)0.3337
D)0.5837
E)None of these
Answer: A
Solution :
Given

H(t)= c + a sin(2πt/9) + b cos(2πt/9)

Using least square method Sr = Σ(hi - (c + a*sin(2πti/9) + b*cos(2πti/9)))^2
To find best possible values of a,b,c equate

```
        del Sr/del a = del Sr/del b =del Sr/del c =0
On solving the three equations
we get a= -0.0717 , b = 0.0837 , c =0.4864
```

-----Q3 V1 (Question 3 , Variant 1)-----

Which of the following statements are Incorerct about PCA(Principal Component Analysis)?

A) All principal components are orthogonal to each other and Maximum number of principal components <= number of features

B) when the features reduces to lower dimensions using PCA,the features carries all information present in data

C) If the input features are on very different scales, it is a good idea to perform feature scaling before applying PCA.

D) PCA will perform outstandingly when eigenvalues are roughly equal

E)None of these

Answer : A C
Explanation :

A) True.The statement is Self explanatory.

B) False.When the features reduces to lower dimensions ,most of the times some information of data will lose and won't be able to interpret the lower dimension data

C) True.Feature scaling prevents one feature dimension from becoming a strongvprincipal component only because of the large magnitude of the feature values (as oppo

D) False. When all eigen vectors are same in such case you won't be able to select the principal components because in that case all principal components are equal.

-----Q3 V2 (Question 3 , Variant 2)-----

Which of the following statements are Incorerct about PCA(Principal Component Analysis)?

A) Given input data x ∈ R^n, it makes sense to run PCA only with values of k that satisfy k <= n. where k is the dimension to which the input data reduced

B) PCA is susceptible to local optima; trying multiple random initializations may help.

C) Given an input x ∈ R^n, PCA compresses it to a lower-dimensional vector z ∈ R^k.

D) PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D).

E)None of these

Answer : A C
Explanation :

A) True.With k = n, there is no compression, so PCA has no use and k > n does not make sense.

B) False.PCA is a deterministic algorithm which doesn't have local minima problem like most of the machine learning algorithms has.

C) True.PCA compresses given input to a lower dimensional vector by projecting it onto the learned principal components

D) False.PCA can be used to reduce the dimensionality of data to any dimensions less than the given dimension

-----Q3 V3 (Question 3 , Variant 3)-----

Which of the following statements are Incorerct about PCA(Principal Component Analysis)?

A) Given input data x ∈ R^n, it makes sense to run PCA only with values of k that satisfy k <= n. where k is the dimension to which the input data reduced

B) when the features reduces to lower dimensions using PCA,the features carries all information present in data

C) Given an input x ∈ R^n, PCA compresses it to a lower-dimensional vector z ∈ R^k.

D) PCA will perform outstandingly when eigenvalues are roughly equal

E)None of these

Answer : A C
Explanation :

A) True.With k = n, there is no compression, so PCA has no use and k > n does not make sense.

B) False.When the features reduces to lower dimensions ,most of the times some information of data will lose and won't be able to interpret the lower dimension data

C) True.PCA compresses given input to a lower dimensional vector by projecting it onto the learned principal components

D) False. When all eigen vectors are same in such case you won't be able to select the principal components because in that case all principal components are equal.

-----Q3 V4 (Question 3 , Variant 4)-----

Which of the following statements are Incorerct about PCA(Principal Component Analysis)?

A) Given input data x ∈ R^n, it makes sense to run PCA only with values of k that satisfy k <= n. where k is the dimension to which the input data reduced

B) when the features reduces to lower dimensions using PCA,the features carries all information present in data

C) Given an input x ∈ R^n, PCA compresses it to a lower-dimensional vector z ∈ R^k.

D) PCA will perform outstandingly when eigenvalues are roughly equal

E)None of these

Answer : A C
Explanation :

A) True.With k = n, there is no compression, so PCA has no use and k > n does not make sense.

B) False.When the features reduces to lower dimensions ,most of the times some information of data will lose and won't be able to interpret the lower dimension data

C) True.PCA compresses given input to a lower dimensional vector by projecting it onto the learned principal components

D) False. When all eigen vectors are same in such case you won't be able to select the principal components because in that case all principal components are equal.

-----Q3 V5 (Question 3 , Variant 5)-----

Which of the following statements are Incorerct about PCA(Principal Component Analysis)?

A) Given input data x ∈ R^n, it makes sense to run PCA only with values of k that satisfy k <= n. where k is the dimension to which the input data reduced

B) PCA is susceptible to local optima; trying multiple random initializations may help.

C) Given an input x ∈ R^n, PCA compresses it to a lower-dimensional vector z ∈ R^k.

D) PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D).

E)None of these

Answer : A C
Explanation :

A) True.With k = n, there is no compression, so PCA has no use and k > n does not make sense.

B) False.PCA is a deterministic algorithm which doesn't have local minima problem like most of the machine learning algorithms has.

C) True.PCA compresses given input to a lower dimensional vector by projecting it onto the learned principal components

D) False.PCA can be used to reduce the dimensionality of data to any dimensions less than the given dimension

-----Q4 V1 (Question 4 , Variant 1)-----

Suppose 10000 patients get tested for flu out of them 1000 are actually healthy and 9000 are actually sick.
For the sick people a test was positive for 8218 and negative for 782.For healthy people, the same test was positive for 438 and negative for 562.Calculate the F1 score for the data
A)0.22
B)0.93
C)0.87
D)0.23
E)None of these
Answer: B
Solution :

	No of Actual sick	No of Actual Healthy
No of predicted sick	8218	438
No of predicted Healthy	782	562

True Positives(TP) : 8218 False Positives(FP) : 438
False Negatives(FN) : 782 True Negatives(TN) : 562
Precision quantifies the number of positive class predictions that actually belong to the positive class.
Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.
F1 score provides a single score that balances both the concerns of precision and recall in one numberTherefore
precision = TP/(TP+FP) = 8218/(8218 + 438) = 0.95
recall= TP/(TP+FN) = 8218/(8218 + 782)= 0.91
F1score= 2*(precision*recall)/(recall+precision)
= 0.93

-----Q4 V2 (Question 4 , Variant 2)-----

Suppose 18000 patients get tested for flu out of them 9000 are actually healthy and 9000 are actually sick.
For the sick people a test was positive for 3854 and negative for 5146.For healthy people, the same test was positive for 1402 and negative for 7598.Calculate the F1 score for the data
A)0.14
B)0.31
C)0.08
D)0.54
E)None of these
Answer: D
Solution :

	No of Actual sick	No of Actual Healthy
No of predicted sick	3854	1402
No of predicted Healthy	5146	7598

True Positives(TP) : 3854 False Positives(FP) : 1402
False Negatives(FN) : 5146 True Negatives(TN) : 7598
Precision quantifies the number of positive class predictions that actually belong to the positive class.
Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.
F1 score provides a single score that balances both the concerns of precision and recall in one numberTherefore
precision = TP/(TP+FP) = 3854/(3854 + 1402) = 0.73
recall= TP/(TP+FN) = 3854/(3854 + 5146)= 0.43

```
                    F1score= 2*(precision*recall)/(recall+precision)
                               = 0.54
```

-----Q4 V3 (Question 4 , Variant 3)-----

Suppose 20000 patients get tested for flu out of them 11000 are actually healthy and 9000 are actually sick.
For the sick people a test was positive for 7186 and negative for 1814.For healthy people, the same test was positive for 1318 and negative for 9682.Calculate the F1 score for the data
A)0.17
B)0.82
C)0.67
D)0.21
E)None of these
Answer: B
Solution :

	No of Actual sick	No of Actual Healthy
No of predicted sick	7186	1318
No of predicted Healthy	1814	9682

```
            True Positives(TP)  : 7186       False Positives(FP) : 1318
            False Negatives(FN) : 1814       True Negatives(TN)  : 9682
```

Precision quantifies the number of positive class predictions that actually belong to the positive class.
Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.
F1 score provides a single score that balances both the concerns of precision and recall in one numberTherefore

```
            precision = TP/(TP+FP) = 7186/(7186 + 1318) = 0.85
                    recall= TP/(TP+FN) = 7186/(7186 + 1814)= 0.80
                    F1score= 2*(precision*recall)/(recall+precision)
                               = 0.82
```

-----Q4 V4 (Question 4 , Variant 4)-----

Suppose 18000 patients get tested for flu out of them 9000 are actually healthy and 9000 are actually sick.
For the sick people a test was positive for 6070 and negative for 2930.For healthy people, the same test was positive for 3083 and negative for 5917.Calculate the F1 score for the data
A)0.17
B)0.45
C)0.11
D)0.67
E)None of these
Answer: D
Solution :

	No of Actual sick	No of Actual Healthy
No of predicted sick	6070	3083
No of predicted Healthy	2930	5917

```
            True Positives(TP)  : 6070       False Positives(FP) : 3083
            False Negatives(FN) : 2930       True Negatives(TN)  : 5917
```

Precision quantifies the number of positive class predictions that actually belong to the positive class.
Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.
F1 score provides a single score that balances both the concerns of precision and recall in one numberTherefore

```
            precision = TP/(TP+FP) = 6070/(6070 + 3083) = 0.66
                    recall= TP/(TP+FN) = 6070/(6070 + 2930)= 0.67
                    F1score= 2*(precision*recall)/(recall+precision)
                               = 0.67
```

-----Q4 V5 (Question 4 , Variant 5)-----

Suppose 27000 patients get tested for flu out of them 18000 are actually healthy and 9000 are actually sick.
For the sick people a test was positive for 5378 and negative for 3622.For healthy people, the same test was positive for 54 and negative for 17946.Calculate the F1 score for the data
A)0.59
B)0.19
C)0.75
D)0.15
E)None of these
Answer: C
Solution :

	No of Actual sick	No of Actual Healthy
No of predicted sick	5378	54
No of predicted Healthy	3622	17946

```
            True Positives(TP)  : 5378       False Positives(FP) : 54
            False Negatives(FN) : 3622       True Negatives(TN)  : 17946
```

Precision quantifies the number of positive class predictions that actually belong to the positive class.
Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.
F1 score provides a single score that balances both the concerns of precision and recall in one numberTherefore

```
            precision = TP/(TP+FP) = 5378/(5378 + 54) = 0.99
                    recall= TP/(TP+FN) = 5378/(5378 + 3622)= 0.60
                    F1score= 2*(precision*recall)/(recall+precision)
                               = 0.75
```