EXAMENS EN LIGNE ? Vérifiez bien préalablement votre accès aux cours concernés.
Naviguez avec Google Chrome pour éviter les soucis d'affichage.

INFO-F422 - Statistical foundations of machine learning - 202122

Commencé le	samedi 4 juin 2022, 23:50
État	Terminé
Terminé le	samedi 4 juin 2022, 23:52
Temps mis	2 min 7 s
Note	**0,57** sur 20,00 (**3**%)

Question 1

Partiellement correct

Note de 0,57 sur 1,00

PCA is a [dimensionality reduction] ✔ algorithm whose aim is to [select the features most correlated with the target] ✖ . It is [unsupervised] ✔ and relies on [SVD] ✔ to [create linear combination of the original features] ✔ and [recursive least-squares] ✖ to [select the shrinkage parameter] ✖

Votre réponse est partiellement correcte.

Vous en avez sélectionné correctement 4.
La réponse correcte est :

PCA is a [dimensionality reduction] algorithm whose aim is to [transform the original input space]. It is [unsupervised] and relies on [SVD] to [create linear combination of the original features] and [cross-validation] to [select the optimal number of components]

Question **2**

Non répondue

Noté sur 1,00

Let us consider a supervised learning problem with two input variables x1 and x2 and one output y. Is it possible to have I(x1;y)>0 , I(x2;y)>0 and I(x1;x2)=0

- a. YES
- b. NON
- c. only is it is a classification problem
- d. only is it is a regression problem
- e. only is all inputs are irrelevant

Votre réponse est incorrecte.

Yes: everytime that the inputs are independent

La réponse correcte est :
YES

Question **3**

Non répondue

Noté sur 1,00

In a supervised learning problem

- a. there is at least a strongly relevant variable
- b. there is at least a weakly relevant variable
- c. there is at least an irrelevant variable
- d. there is at least a relevant variable only if the conditional entropy of the target is smaller than its marginal entropy
- e. the conditional entropy of the target is always larger than the marginal entropy
- f. a variable x is relevant only if the mutual information I(x;y)>0

Votre réponse est incorrecte.

La réponse correcte est :
there is at least a relevant variable only if the conditional entropy of the target is smaller than its marginal entropy

Question **4**

Non répondue

Noté sur 1,00

Let us consider a supervised learning problem with two input variables x1 and x2 and one output y. Is it possible to have I(x1;y)>0 , I(x2;y)>0 and I((x1,x2); y)=0

- a. YES
- b. NON
- c. only is it is a classification problem
- d. only is it is a regression problem
- e. only is all inputs are irrelevant

Votre réponse est incorrecte.

No: since I((x1,x2);y)=I(x1;y)+I(x2;y| x1) and the second term cannot be negative

La réponse correcte est :
NON

Question **5**

Non répondue

Noté sur 1,00

Let us consider 4 random binary variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{y}$ where $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are independent and uniform and the conditional distribution of $\mathbf{y} = 1$ given the vector $x = (x_1, x_2, x_3)$ is

x1	x2	x3	P(y=1|x)
0	0	1	0.2
0	0	1	0.8
0	1	0	0.8
0	1	1	0.2
1	0	0	0.2
1	0	1	0.8
1	1	0	0.8
1	1	1	0.2

Compute by using the base-2 logarithm

- the entropy $H(\mathbf{y})$: ✖
- the entropy $H(\mathbf{y}|\mathbf{x})$: ✖
- the entropy $H(\mathbf{y}|\mathbf{x}_1)$: ✖
- the entropy $H(\mathbf{y}|\mathbf{x}_2)$: : ✖

- the entropy $H(\mathbf{y}|\mathbf{x}_3)$: : ✖
- the entropy $H(\mathbf{y}|\mathbf{x}_2, \mathbf{x}_3)$: ✖
- the information $I(\mathbf{x}_1; \mathbf{x}_2)$: ✖
- the information $I(\mathbf{x}_2; \mathbf{y}|\mathbf{x}_3)$: ✖

Let us first compute the marginal distribution of $\mathbf{y}$:

$P(\mathbf{y}=1) = 0.5 = P(\mathbf{y}=0)$

Then $H(\mathbf{y}) = 1$

The entropy of $H(\mathbf{y}|\mathbf{x})$ is the average of all conditional entropies.

Since all conditional entropies are identical and equal to $H(\mathbf{y}|\mathbf{x}=[0,0,0]) = 0.7219281$ we have $H(\mathbf{y}|\mathbf{x}) = 0.7219281$

To compute $H(\mathbf{y}|\mathbf{x}_1)$ let us first derive

$P(\mathbf{y}=1|\mathbf{x}_1=0) = 0.5$ and $P(\mathbf{y}=1|\mathbf{x}_1=1) = 0.5$

Then $H(\mathbf{y}|\mathbf{x}_1) = 1$

To compute $H(\mathbf{y}|\mathbf{x}_2)$ let us first derive

$P(\mathbf{y}=1|\mathbf{x}_2=0) = 0.5$ and $P(\mathbf{y}=1|\mathbf{x}_2=1) = 0.5$

Then $H(\mathbf{y}|\mathbf{x}_2) = 1$

Since $H(\mathbf{y}|\mathbf{x}_2=0,\mathbf{x}_3=0) = H(\mathbf{y}|\mathbf{x}_2=0,\mathbf{x}_3=1) = H(\mathbf{y}|\mathbf{x}_2=1,\mathbf{x}_3=0) = H(\mathbf{y}|\mathbf{x}_2=1,\mathbf{x}_3=1) = 0.7219281$ we have $H(\mathbf{y}|\mathbf{x}_2,\mathbf{x}_3) = 0.7219281$

$I(\mathbf{x}_1; \mathbf{x}_2) = 0$ since the two variables are independent

$$I(\mathbf{x}_2; \mathbf{y}|\mathbf{x}_3) = H(\mathbf{y}|\mathbf{x}_3) - H(\mathbf{y}|\mathbf{x}_2, \mathbf{x}_3) = 1 - 0.7219281$$

Question 6

Non répondue

Noté sur 1,00

Let us consider 4 random binary variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{y}$ where $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are independent and uniform and the conditional distribution of $\mathbf{y} = 1$ given the vector $x = (x_1, x_2, x_3)$ is

x1	x2	x3	P(y=1|x)
0	0	1	0.2
0	0	1	0.8
0	1	0	0.8
0	1	1	0.2
1	0	0	0.2
1	0	1	0.8
1	1	0	0.8
1	1	1	0.2

Compute by using the base-2 logarithm

- the entropy $H(\mathbf{x}_1)$: ✖
- the information $I(\mathbf{y}; \mathbf{x})$: ✖
- the information $I(\mathbf{y}; \mathbf{x}_1 | \mathbf{x}_2)$: ✖
- the information $I(\mathbf{y}; \mathbf{x}_2)$: ✖

- the information $I(\mathbf{y};\mathbf{x}_3)$: ✖
- the information $I(\mathbf{y};(\mathbf{x}_2,\mathbf{x}_3))$: ✖
- the information $I(\mathbf{x}_1;\mathbf{x}_3)$: ✖
- the information $I(\mathbf{x}_2;\mathbf{x}_3)$: ✖

Since $\mathbf{x}_1$ is uniform then $H(\mathbf{x}_1) = 1$

Let us first compute the marginal distribution of $\mathbf{y}$:

$P(\mathbf{y}=1) = 0.5 = P(\mathbf{y}=0)$

Then $H(\mathbf{y}) = 1$

Since all conditional entropy terms $H(\mathbf{y}|\mathbf{x}=x)$ are identical and equal to 0.7219281 we have

$H(\mathbf{y}|\mathbf{x} = 0.7219281$ and $I(\mathbf{y};\mathbf{x}) = 1 - 0.7219281 = 0.2780719$

Since

$P(\mathbf{y}=1|\mathbf{x}_1=0,\mathbf{x}_2=0) = P(\mathbf{y}=1|\mathbf{x}_1=0,\mathbf{x}_2=1) = P(\mathbf{y}=1|\mathbf{x}_1=1,\mathbf{x}_2=0) = P(\mathbf{y}=1|\mathbf{x}_1=1,\mathbf{x}_2=1) = 0.5$

we have $H(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) = 1$ and then

$I(\mathbf{y};\mathbf{x}_1|\mathbf{x}_2) = H(\mathbf{y}|\mathbf{x}_2) - H(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) = 1 - 1 = 0$

To compute $H(\mathbf{y}|\mathbf{x}_2)$ let us first derive

$P(\mathbf{y}=1|\mathbf{x}_2=0)=0.5$ and $P(\mathbf{y}=1|\mathbf{x}_2=1)=0.5$

Then $H(\mathbf{y}|\mathbf{x}_2)=1$ and $I(\mathbf{y};\mathbf{x}_2)=0$

Since $H(\mathbf{y}|\mathbf{x}_2=0,\mathbf{x}_3=0)=H(\mathbf{y}|\mathbf{x}_2=0,\mathbf{x}_3=1)=H(\mathbf{y}|\mathbf{x}_2=1,\mathbf{x}_3=0)=H(\mathbf{y}|\mathbf{x}_2=1,\mathbf{x}_3=1)=0.7219281$ we have $H(\mathbf{y}|\mathbf{x}_2,\mathbf{x}_3)=0.7219281$ and $I(\mathbf{y};\mathbf{x}_2,\mathbf{x}_3)=0.2780719$

$I(\mathbf{x}_1;\mathbf{x}_3)=0$ since the two variables are independent

$I(\mathbf{x}_2;\mathbf{x}_3)=0$ since the two variables are independent

Question **7**

Non répondue

Noté sur 2,00

Let us consider a regression task with $n = 50$ inputs and 1 output
whose training set is contained in the data matrices X and Y and the test data set
is contained in the data matrices Xts and Yts of this .Rdata file

By using R, compute the 4 most relevant eigen-features by using PCA.
Then compare the MISE error (mean of the squared prediction error) for the test set with the original set of features and the set of eigen-features. Use as learning algorithm a linear least squares

- MISE test (all features) : ✖
- MISE test (subset of 5 best ranked features) : ✖

Was feature selection useful?

○ NO ○ YES

Note de 0,00 sur 1,00

La réponse correcte est : YES

```
load("FS2.Rdata")

n=NCOL(X)
N=NROW(X)

fselected<-4
Xhat=scale(X)
S=svd(Xhat)
Z=Xhat%*%S$v

Xhats=scale(Xts)
Zts=Xhats%*%S$v

Z=Z[,1:fselected]
Zts=Zts[,1:fselected]


X=cbind(numeric(N)+1,X)
Xts=cbind(numeric(NROW(Xts))+1,Xts)
betahat=solve(t(X)%*%X)%*%t(X)%*%Y
Yhats=Xts%*%betahat
MISEts=mean((Yts-Yhats)^2)


Z=cbind(numeric(N)+1,Z)
Zts=cbind(numeric(NROW(Zts))+1,Zts)
```

```
betahat=solve(t(Z)%*%Z)%*%t(Z)%*%Y
Yhats=Zts%*%betahat
MISEts2=mean((Yts-Yhats)^2)
```

Question **8**

Non répondue

Noté sur 2,00

Let us consider a regression task with $n = 20$ inputs and 1 output
whose training set is contained in the data matrices X and Y and the test data set
is contained in the data matrices Xts and Yts of this .Rdata file

By using R, find the 5 most relevant features by using a forward selection algorithm based on linear least-squares and leave-one-out. Then compare the MISE error (mean of the squared prediction error) for the test set with the original set of features and the selected set of features. Use as learning algorithm a linear least squares

- MISE test (all features) : ✖
- MISE test (subset of 5 best ranked features) : ✖

Was feature selection useful?

○ NO ○ YES

Note de 0,00 sur 1,00

La réponse correcte est : YES

```
rm(list=ls())

load("FS.Rdata")
n=NCOL(X)
N=NROW(X)

fselected<-NULL
nmax=5
for (f in 1:nmax){
  MSEloo=numeric(n)+Inf
  for (j in setdiff(1:n,fselected)){
    subs<-c(fselected,j)
    eloo=numeric(N)
    for (i in 1:N){
      Xi=cbind(numeric(N-1)+1,X[-i,subs])
      Yi=Y[-i]
      betai=solve(t(Xi)%*%Xi)%*%t(Xi)%*%Yi
      yhati=c(1,X[i,subs])%*%betai
      eloo[i]=Y[i]-yhati
    }
    MSEloo[j]=mean(eloo^2)
  }
  fselected=c(fselected,which.min(MSEloo))

}
fsel=fselected
```

```
X2=X[,fsel]
Xts2=Xts[,fsel]
X=cbind(numeric(N)+1,X)
Xts=cbind(numeric(NROW(Xts))+1,Xts)
betahat=solve(t(X)%*%X)%*%t(X)%*%Y
Yhats=Xts%*%betahat
MISEts=mean((Yts-Yhats)^2)


X2=cbind(numeric(N)+1,X2)
Xts2=cbind(numeric(NROW(Xts))+1,Xts2)
betahat2=solve(t(X2)%*%X2)%*%t(X2)%*%Y
Yhats2=Xts2%*%betahat2
MISEts2=mean((Yts-Yhats2)^2)
```

Question 9

Non répondue

Noté sur 2,00

Let us consider a regression task with $n = 20$ inputs and 1 output
whose training set is contained in the data matrices X and Y of this .Rdata file.

By using R, find the 5 most relevant features by using a ranking algorithm based on correlation

- First most important feature : ✖
- Second most important feature : ✖
- Third most important feature : ✖
- Fourth most important feature : ✖
- Fifth most important feature : ✖

```
load("FS.Rdata")
n=NCOL(X)
corXY=NULL
for (j in 1:n){
```

```
  corXY=c(corXY,abs(cor(X[,j],Y)))
}
print(sort(corXY,decre=TRUE,index=TRUE)$ix[1:5])
```

Question 10

Non répondue

Noté sur 2,00

Let us consider a regression task with $n = 20$ inputs and 1 output
whose training set is contained in the data matrices X and Y of this .Rdata file

By using R, find the 4 most relevant according to a mRMR filter strategy where the mutual information is estimated on the basis of the Pearson correlation.

- a. 1
- b. 2
- c. 3
- d. 4
- e. 5
- f. 6
- g. 7
- h. 8
- i. 9
- j. 10
- k. 11
- l. 12
- m. 13
- n. 14
- o. 15

p. 16

q. 17

r. 18

s. 19

t. 20

u. 21

v. 22

w. 23

Votre réponse est incorrecte.

```
rm(list=ls())
load("FS.Rdata")
n=NCOL(X)
N=NROW(X)
XY<-cbind(X,Y)
CC=cor(XY)
InfM=-1/2*log(1-CC^2)
subset=which.max(InfM[-(n+1),n+1])

for (s in 1:3){
 mRMR<-numeric(n)-Inf
 for (j in setdiff(1:n,subset)){
  mRMR[j]=InfM[j,n+1]-mean(InfM[j,subset])
 }
 subset<-c(subset,which.max(mRMR))
}
```

print(subset)

Les réponses correctes sont :
3,

4,

5,

19

Question **11**

Non répondue

Noté sur 2,00

Let us consider a regression task with $n = 20$ inputs and 1 output
whose training set is contained in the data matrices X and Y and the test data set
is contained in the data matrices Xts and Yts of this .Rdata file

By using R, find the 3 most relevant according to an embedded strategy based on a balanced decision tree (of depth 2) with constant models in the leaves.

- a. 1
- b. 2
- c. 3
- d. 4
- e. 5
- f. 6
- g. 7
- h. 8
- i. 9
- j. 10
- k. 11
- l. 12
- m. 13
- n. 14
- o. 15

- p. 16
- q. 17
- r. 18
- s. 19
- t. 20

Votre réponse est incorrecte.

```
rm(list=ls())


splitRT<-function(X,Y, splits){
 ## X [N,n]
 ##splits [S,1]
 n<-NCOL(X)
 S<-length(splits)
 SSE<-numeric(n)
 bests<-numeric(n)
 for (f in 1:n){
  SSEs<-numeric(S)
  for (s in 1:S){
   I1<-which(X[,f]<splits[s])
   I2<-which(X[,f]>=splits[s])
   SSE1=sum((Y[I1]-mean(Y[I1])^2))
   SSE2=sum((Y[I2]-mean(Y[I2])^2))
   SSEs[s]<-SSE1+SSE2
  }
```

```
    SSE[f]=min(SSEs)
    bests[f]<-which.min(SSEs)
  }
  list(bestf=which.min(SSE),bestsplit=splits[bests[which.min(SSE)]])
}



load("FS.Rdata")

n=NCOL(X)
N=NROW(X)
splits=seq(-2,2,by=0.5)

Spl1<-splitRT(X,Y,splits)
fs1<-Spl1$bestf

I1<-which(X[,fs1]<Spl1$bestsplit)
I2<-which(X[,fs1]>=Spl1$bestsplit)


Spl2<-splitRT(X[I1,],Y[I1],splits)
fs2<-Spl2$bestf

Spl3<-splitRT(X[I2,],Y[I2],splits)
fs3<-Spl3$bestf
```

Les réponses correctes sont :

1,

3,

4

Question **12**

Non répondue

Noté sur 2,00

Let us consider a regression task with $n = 50$ inputs and 1 output
whose training set is contained in the data matrices X and Y of this .Rdata file

By using R, compute the optimal λ shrinkage parameter for a ridge-regression approach by using a leave-one-out assessment strategy.

◯ 0 ◯ 1 ◯ 10 ◯ 100 ◯ 1000 ◯ 10000 ◯ 100000

Note de 0,00 sur 1,00

La réponse correcte est : 10

```
rm(list=ls())
```

```
load("FS.Rdata")

n=NCOL(X)
N=NROW(X)

X<-cbind(numeric(N)+1,X)


LAM=c(0,1,10,100,1000,10000,100000)
MSEloo=numeric(length(LAM))
for (l  in 1:length(LAM)){
  Eloo<-NULL
  lam=LAM[l]
  for (i in 1:N){
    Xi=X[-i,]
    Yi=Y[-i]
    betahat= solve(t(Xi)%*%Xi+lam*diag(n+1))%*%t(Xi)%*%Yi
    Eloo<-c(Eloo,Y[i]-X[i,]%*%betahat)
  }
  MSEloo[l]=mean(Eloo^2)
}
bestlam=LAM[which.min(MSEloo)]

cat("best lam=",bestlam,"\n")
```

Question **13**

Non répondue

Noté sur 2,00

Let us consider a regression task with $n = 50$ inputs and 1 output
whose training set is contained in the data matrices X and Y of this .Rdata file

By using R, compute the optimal λ shrinkage parameter for a ridge-regression approach by using a leave-one-out assessment strategy.

○ 0 ○ 1 ○ 10 ○ 100 ○ 1000 ○ 10000 ○ 100000

Note de 0,00 sur 1,00

La réponse correcte est : 100

```
rm(list=ls())
load("FS2.Rdata")
n=NCOL(X)
```

```
N=NROW(X)

X<-cbind(numeric(N)+1,X)


LAM=c(0,1,10,100,1000,10000,100000)
MSEloo=numeric(length(LAM))
for (l  in 1:length(LAM)){
  Eloo<-NULL
  lam=LAM[l]
  for (i in 1:N){
    Xi=X[-i,]
    Yi=Y[-i]
    betahat= solve(t(Xi)%*%Xi+lam*diag(n+1))%*%t(Xi)%*%Yi
    Eloo<-c(Eloo,Y[i]-X[i,]%*%betahat)
  }
  MSEloo[l]=mean(Eloo^2)
}
bestlam=LAM[which.min(MSEloo)]

cat("best lam=",bestlam,"\n")
```

«

»