

Subject Name (Gujarati)

4341603 -- Summer 2025

Semester 1 Study Material

Detailed Solutions and Explanations

પ્રશ્ન 1(અ) [3 ગુણ]

મશીન લર્નિંગની વ્યાખ્યા આપો. મશીન લર્નિંગની કોઈપણ બે ઉપયોગીતાઓ આપો.

જવાબ

મશીન લર્નિંગ એ આર્ટિફિશિયલ ઇન્ટેલિજન્સનો એક ભાગ છે જે કમ્પ્યુટરને ડેટામાંથી શીખવા અને દરેક કાર્ય માટે સ્પષ્ટ પ્રોગ્રામિંગ વિના નિર્ણયો લેવાની ક્ષમતા આપે છે.

ઉપયોગીતાઓ:

- ઇમેઇલ સ્પામ ડિટેક્શન: આપોઆપ સ્પામ ઇમેઇલ ઓળખે અને ફિલ્ટર કરે છે
- સુઝાવ સિસ્ટમ: Amazon જેવી ઈ-કોમર્સ સાઇટ્સ પર પ્રોડક્ટ સુઝાવે છે

ટેબલ: ML વિટ્રેડિશનલ પ્રોગ્રામિંગ

| પરંપરાગત પ્રોગ્રામિંગ | મશીન લર્નિંગ |
|--|---|
| ઇનપુટ ડેટા + પ્રોગ્રામ → નિયમો સ્પષ્ટપણે કોડ કરવામાં આવે છે | ઇનપુટ ડેટા + આર્ટિફિશિયલ નિયમો ડેટામાંથી શીખવામાં આવે છે |

મેમરી ટ્રીક

"ML = ડેટામાંથી શીખવું બનાવો"

પ્રશ્ન 1(બ) [4 ગુણ]

વ્યાખ્યા આપો: અંડર ફિલ્ટરિંગ અને ઓવર ફિલ્ટરિંગ.

જવાબ

અંડરફિલ્ટરિંગ ત્યારે થાય છે જ્યારે મોડલ ડેટામાં છુપાયેલા પેર્ટન કેચર કરવા માટે ખૂબ સાંદું હોય છે, જેના પરિણામે ટ્રેનિંગ અને ટેસ્ટ બંને ડેટા પર નબળી કામગીરી થાય છે.

ઓવરફિલ્ટરિંગ ત્યારે થાય છે જ્યારે મોડલ ટ્રેનિંગ ડેટાને અવાજ સહિત ખૂબ સારી રીતે શીખે છે, જેના કારણે નવા અદ્ધર્થ ડેટા પર નબળી કામગીરી થાય છે.

ટેબલ: સરખામણી

| પાસું | અંડરફિલ્ટરિંગ | ઓવરફિલ્ટરિંગ |
|----------------------|---------------------|---------------------|
| ટ્રેનિંગ એક્ષ્યુરેસી | ઓછી | વધારે |
| ટેસ્ટ એક્ષ્યુરેસી | ઓછી | ઓછી |
| મોડલ કોમ્પ્લેક્સટી | ખૂબ સાંદું | ખૂબ જટિલ |
| સોલ્યુશન | કોમ્પ્લેક્સટી વધારો | કોમ્પ્લેક્સટી ઘટાડો |

મેમરી ટ્રીક

"અંડર = ઓછું કામ, ઓવર = વધુ પડતું શીખવું"

પ્રશ્ન 1(ક) [7 ગુણ]

મશીન લર્નિંગના વિવિધ પ્રકારો યોગ્ય ઉદાહરણની મદદથી વર્ણવો.

જવાબ

ટેબલ: મશીન લર્નિંગના પ્રકારો

| પ્રકાર | વર્ણન | ઉદાહરણ |
|----------------|----------------------------------|--------------------|
| સુપરવાઇઝડ | લેબલ કરેલ ટ્રેનિંગ ડેટા વાપરે છે | ઇમેઇલ વર્ગીકરણ |
| અન્સુપરવાઇઝડ | લેબલ કરેલ ડેટા નથી, પેટન શોધે છે | કુટ્ટમર સેંગમેટેશન |
| રિઝન્ફોર્મેન્ટ | પુરસ્કાર/દંડ દ્વારા શીખે છે | ગેમ રમતું AI |

સુપરવાઇઝડ લર્નિંગ ઇનપુટ-આઉટપુટ જોડીઓ વાપરીને મોડલ ટ્રેન કરે છે. અન્નોરિધમ ઉદાહરણોમાંથી શીખીને નવા ડેટા માટે પરિણામોની આગાહી કરે છે.

અન્સુપરવાઇઝડ લર્નિંગ ટાર્ગેટ લેબલ વિના ડેટામાં છુપાયેલા પેટન શોધે છે. તે સમાન ડેટા પોઇન્ટ્સને એક્સાથે જૂથબદ્ધ કરે છે.

રિઝન્ફોર્મેન્ટ લર્નિંગ સારા કાર્યો માટે પુરસ્કાર અને ખરાબ કાર્યો માટે દંડ આપીને એજન્ટને નિર્ણય લેવાનું શીખવે છે.

ડાયગ્રામ:

Mermaid Diagram (Code)

```
{Shaded}
{Highlighting} []
graph TD
    A[ ] --- B[ ]
    A --- C[ ]
    A --- D[ ]
    B --- E[ ]
    B --- F[ ]
    C --- G[ ]
    C --- H[ ]
{Highlighting}
{Shaded}
```

મેમરી ટ્રીક

"સુપર અન-સુપરવાઇઝડ રિઝન્ફોર્સ શીખવું"

પ્રશ્ન 1(ક) અથવા [7 ગુણ]

મશીન લર્નિંગમાં ઉપયોગ થતી વિવિધ ટૂલ્સ અને ટેકનોલોજી વર્ણવો.

જવાબ

ટેબલ: ML ટૂલ્સ અને ટેકનોલોજીઓ

| કેટેગોરી | ટૂલ્સ | હેતુ |
|-----------------|--------------------------|------------------|
| પ્રોગ્રામિંગ | Python, R | મુખ્ય ડેવલપમેન્ટ |
| લાઇબ્રેરીઓ | Scikit-learn, TensorFlow | મોડલ બિલ્ડિંગ |
| ડેટા પ્રોસેસિંગ | Pandas, NumPy | ડેટા મેનેજમેન્ટ |
| વિજ્યુલાઇઝન | Matplotlib, Seaborn | ડેટા પ્લોટિંગ |

Python તેની સરળતા અને વ્યાપક લાઇબ્રેરીઓને કારણે સૌથી લોકપ્રિય ભાષા છે.
Scikit-learn ડેટા માઇનિંગ અને વિશ્લેષણ માટે સરળ ટૂલ્સ પ્રદાન કરે છે, જે શરૂઆતીઓ માટે પરફેક્ટ છે.
TensorFlow અને **PyTorch** ડીપ લર્નિંગ એપ્લિકેશન માટે એડવાન્સ ફેમર્વર્ક છે.
Jupyter Notebook પ્રોગ્રામ માટે ઇન્ટરેક્ટિવ ડેવલપમેન્ટ એન્વાર્ન્મેન્ટ ઓફર કરે છે.

ડાયાગ્રામ:

Mermaid Diagram (Code)

```
{Shaded}
{Highlighting} []
graph LR
    A[ ] --> B[Pandas/NumPy]
    B --> C[Scikit{-}learn]
    C --> D[ ]
    D --> E[Matplotlib]
{Highlighting}
{Shaded}
```

મેમરી ટ્રીક

“Python Pandas Scikit Tensor Jupyter”

પ્રશ્ન 2(અ) [3 ગુણ]

Qualitative ડેટા અને Quantitative ડેટા વચ્ચેનો તફાવત આપો.

જવાબ

ટેબલ: Qualitative વિષિ Quantitative ડેટા

| Qualitative ડેટા | Quantitative ડેટા |
|---|--|
| બિન-સંખ્યાત્મક કેટેગરીઓ રૂંગો, નામો, ગ્રેડ્સ માપી શકતું નથી | સંખ્યાત્મક મૂલ્યો ઉંચાઈ, વજન, કિંમત માપી શકાય છે |

Qualitative ડેટા એવા ગુણો અથવા લક્ષણોનું વર્ણન કરે છે જે સંખ્યાત્મક રીતે માપી શકતા નથી.
Quantitative ડેટા સંખ્યાઓ તરીકે વ્યક્ત કરેલા માપી શકાય તેવા જથ્થાઓનું પ્રતિનિધિત્વ કરે છે.

મેમરી ટ્રીક

“Quality = કેટેગરીઓ, Quantity = સંખ્યાઓ”

પ્રશ્ન 2(બ) [4 ગુણ]

નીચે આપેલા ડેટાનું mean અને median શોધો: 3,4,5,5,7,8,9,11,12,14.

જવાબ

આપેલ ડેટા: 3, 4, 5, 5, 7, 8, 9, 11, 12, 14

Mean ગણતરી:

- સરવાળો = $3+4+5+5+7+8+9+11+12+14 = 78$
- સંખ્યાઓની ઊંઠતી = 10
- Mean = $78/10 = 7.8$

Median ગણતરી:

- ડેટા પહેલેથી જ સોર્ટ થયેલ છે
- 10 સંખ્યાઓ માટે: Median = $(5\text{મી} + 6\text{ઠી})/2$
- Median = $(7+8)/2 = 7.5$

ટેબલ: પરિણામો

| માપદંડ | મૂલ્ય |
|--------|-------|
| Mean | 7.8 |
| Median | 7.5 |

મેમરી ટ્રીક

“Mean = સરેરાશ, Median = મધ્યક”

પ્રશ્ન 2(ક) [7 ગુણ]

મશીન લર્નિંગની એક્ટિવિટી વિગતવાર વર્ણવો.

જવાબ

ટેબલ: મશીન લર્નિંગ એક્ટિવિટીઓ

| એક્ટિવિટી | વર્ણન | ઉદાહરણ |
|-------------------|----------------------------------|------------------------|
| ડેટા કલેક્શન | સંબંધિત ડેટા એક્સ્ટ્રાક્શન કરવું | સર્વે પ્રતિભાવો |
| ડેટા પ્રીપોર્ચિંગ | ડેટા સાફ્ અને તૈયાર કરવું | ડુલ્યિક્ટ્ડ્સ દૂર કરવા |
| ફીચર સિલેક્શન | મહત્વપૂર્ણ વેરિયેબલ્સ પસંદ કરવા | લોન માટે ઉમર, આવક |
| મોડલ ટ્રેનિંગ | અલ્ગોરિધમને પેટર્ન શીખવવું | ટ્રેનિંગ ડેટા ખવડાવવો |
| મોડલ ઇવેલ્યુએશન | મોડલની કામગીરી પરીક્ષણ | એક્યુરેસી મેઝરમેન્ટ |

ડેટા કલેક્શન ડેટાબેસ, સેન્સર્સ અથવા સર્વે જેવા વિવિધ સોતોમાંથી માહિતી એક્સ્ટ્રાક્શન કરવાનો સમાવેશ કરે છે.

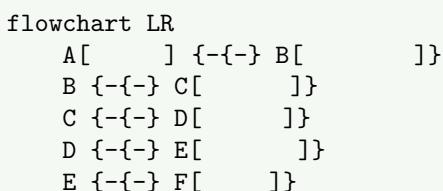
ડેટા પ્રીપોર્ચિંગ વિશ્લેષણ માટે કાચા ડેટાને સાફ્, રૂપાંતર અને ગોઠવવાનો સમાવેશ કરે છે.

ફીચર સિલેક્શન આગાહીઓમાં યોગદાન આપતા સૌથી સંબંધિત વેરિયેબલ્સ ઓળખે છે.

મોડલ ટ્રેનિંગ તૈયાર કરેલા ટ્રેનિંગ ડેટામાંથી પેટર્ન શીખવા માટે અલ્ગોરિધમસનો ઉપયોગ કરે છે.

મોડલ ઇવેલ્યુએશન ટ્રેન કરેલ મોડલ નવા, અદ્રશ્ય ડેટા પર કેટલી સારી કામગીરી કરે છે તેનું પરીક્ષણ કરે છે.

ડાયાગ્રામ:



મેમરી ટ્રીક

“કલેક્ટ પ્રોસેસ ફીચર ટ્રેન ઇવેલ્યુએટ ડિપ્લોય”

પ્રશ્ન 2(અ) અથવા [3 ગુણ]

Predictive મોડલ અને Descriptive મોડલ વચ્ચેનો તફાવત આપો.

જવાબ

ટેબલ: Predictive વિ Descriptive મોડલ્સ

| Predictive મોડલ | Descriptive મોડલ |
|---|--|
| ભવિષ્યના પરિણામોની આગાહી કરે છે સુપરવાઈઝ લર્નિંગ વાપરે છે સ્ટોક પ્રાઇસ પ્રિડિક્શન | વર્તમાન પેટર્નનું સમજૂતી આપે છે અનસુપરવાઈઝ લર્નિંગ વાપરે છે કસ્ટમર સેગમેન્ટેશન |

Predictive મોડલ્સ ભવિષ્યની ઘટનાઓ અથવા અજાણાત્મક આગાહી કરવા માટે ઐતિહાસિક ડેટાનો ઉપયોગ કરે છે.
Descriptive મોડલ્સ વર્તમાન પેટર્ન અને સંબંધોને સમજવા માટે હાલના ડેટાનું વિશ્લેષણ કરે છે.

મેમરી ટ્રીક

“Predict = ભવિષ્ય, Describe = વર્તમાન”

પ્રશ્ન 2(બ) અથવા [4 ગુણ]

નીચે આપેલા ડેટાને ચોંચ ડેટા ટાઇપની મદદથી classify કરો: hair color, gender, blood group type, time of day.

જવાબ

ટેબલ: ડેટા ટાઇપ કલાસિફિકેશન

| ડેટા | પ્રકાર | કારણ |
|-------------|------------|-----------------------|
| Hair color | Nominal | કોઈ કમ વિના કેટેગરીઓ |
| Gender | Nominal | કોઈ કમ વિના કેટેગરીઓ |
| Blood group | Nominal | કોઈ કમ વિના કેટેગરીઓ |
| Time of day | Continuous | માપી શકાય તેવી માત્રા |

Nominal ડેટા કોઈ કુદરતી કમ વિના કેટેગરીઓનું પ્રતિનિધિત્વ કરે છે.

Continuous ડેટા શૈખીમાં કોઈપણ મૂલ્ય લઈ શકે છે અને માપી શકાય છે.

મેમરી ટ્રીક

“નામો = Nominal, સંખ્યાઓ = Numerical”

પ્રશ્ન 2(ક) અથવા [7 ગુણ]

ડેટા પ્રી-પ્રોસેસિંગમાં ઉપયોગ થતી વિવિધ મેથ્ડડ્સ વર્ણવો.

જવાબ

ટેબલ: ડેટા પ્રીપ્રોસેસિંગ મેથ્ડડ્સ

| મેથ્ડ | હેતુ | ઉદાહરણ |
|---------------------|-------------------------------|------------------------------------|
| ડેટા કલીનિંગ | ભૂલો અને અસંગતતાઓ દૂર કરવી | ટાઇપોઝ સીક કરવા, ડુલિકેટસ દૂર કરવા |
| ડેટા ઇન્ટીગ્રેશન | બહુવિધ સોતો એક્સાથે જોડવા | કસ્ટમર ડેટાબેસ મર્જ કરવા |
| ડેટા ટ્રાન્સફર્મેશન | ચોંચ ફોર્મેટમાં કન્વર્ટ કરવું | 0-1 મૂલ્યો નોર્મલાઇઝ કરવા |
| ડેટા રિડક્ષન | ડેટાસેટનું કદ ઘટાડવું | મહત્વપૂર્ણ ફીચર્સ પરસ્પંદ કરવા |

ડેટા કલીનિંગ ભૂલભરેલ, અધૂરા અથવા અપ્રસ્તુત ડેટાને દૂર કરે છે અથવા સુધારે છે.

ડેટા ઇન્ટીગ્રેશન બહુવિધ સોતોમાંથી ડેટાને એક્સીક્યુટિવ ડેટાસેટમાં જોડે છે.

ડેટા ટ્રાન્સફર્મેશન વિશ્લેષણ માટે ડેટાને ચોંચ ફોર્મેટમાં કન્વર્ટ કરે છે.

ડેટા રિડક્ષન માહિતીની ગુણવત્તા જાળવીને ડેટાસેટનું કદ ઘટાડે છે.

ડાયાગ્રામ:

Mermaid Diagram (Code)

```

{Shaded}
{Highlighting} []
graph LR
    A[ ] --{-{-}{}} B[ ]
    B --{-{-}{}} C[ ]
    C --{-{-}{}} D[ ]
    D --{-{-}{}} E[ ]
    E --{-{-}{}} F[ ]
{Highlighting}
{Shaded}

```

મેમરી ટ્રીક

“કલીન ઇન્ટીગ્રેટ ટ્રાન્સફોર્મ રિડ્યુસ”

પ્રશ્ન 3(અ) [3 ગુણ]

Classification અને Regression વચ્ચેનો તફાવત આપો.

જવાબ

ટેબલ: Classification વિશે Regression

| Classification | Regression |
|---|--|
| ડિસ્ક્રીટ આઉટપુટ કેટેગરીઓની આગાહી કરે છે ઈમેઇલ: સ્પામ/બિન-સ્પામ | કન્ટિન્યુઅસ આઉટપુટ સંખ્યાત્મક મૂલ્યોની આગાહી કરે છે ઘરની કિંમત આગાહી |

Classification ઇનપુટ ડેટામાંથી ડિસ્ક્રીટ કેટેગરીઓ અથવા કલાસની આગાહી કરે છે.

Regression ઇનપુટ ડેટામાંથી કન્ટિન્યુઅસ સંખ્યાત્મક મૂલ્યોની આગાહી કરે છે.

મેમરી ટ્રીક

“Class = કેટેગરીઓ, Regress = વાસ્તવિક સંખ્યાઓ”

પ્રશ્ન 3(બ) [4 ગુણ]

યોગ્ય ઉદાહરણ લઈને confusion matrix લખો. તેના માટે accuracy અને error rate ગણો.

જવાબ

ઉદાહરણ: ઈમેઇલ કલાસિફિકેશન

ટેબલ: Confusion Matrix

| | પ્રિડિક્ટેડ સ્પામ | પ્રિડિક્ટેડ નોટ સ્પામ |
|--------------------|-------------------|-----------------------|
| વાસ્તવિક સ્પામ | 85 (TP) | 15 (FN) |
| વાસ્તવિક નોટ સ્પામ | 10 (FP) | 90 (TN) |

ગણતરીઓ:

- Accuracy = $(TP+TN)/(TP+TN+FP+FN) = (85+90)/200 = 87.5\%$
- Error Rate = $(FP+FN)/(TP+TN+FP+FN) = (10+15)/200 = 12.5\%$

મુખ્ય શાબ્દો:

- TP: True Positive - યોગ્ય રીતે સ્પામ આગાહી
- TN: True Negative - યોગ્ય રીતે નોટ સ્પામ આગાહી

મેમરી ટ્રીક

“True Positive True Negative = યોગ્ય આગાહીઓ”

પ્રશ્ન 3(ક) [7 ગુણ]

KNN અલોરિધમ વિગતવાર વર્ણવો.

જવાબ

K-Nearest Neighbors (KNN) એક સરળ કલાસિફિકેશન અલોરિધમ છે જે તેમના K નજીકના પડોશીઓના મેજોરિટી કલાસના આધારે ડેટા પોઇન્ટ્સને કલાસિફાઇ કરે છે.

ટેબલ: KNN અલોરિધમ સ્ટેપ્સ

| સ્ટેપ | વર્ણન | ઉદાહરણ |
|-------------|-----------------------------|---------------------|
| K પસંદ કરો | પડોશીઓની સંખ્યા પસંદ કરો | K=3 |
| અંતર ગણો | બધા પોઇન્ટ્સનો અંતર શોધો | Euclidean અંતર |
| પડોશીઓ શોધો | K સૌથી નજીકના પોઇન્ટ્સ ઓળખો | 3 નજીકના પોઇન્ટ્સ |
| વોટ કરો | મેજોરિટી કલાસ જીતે છે | 2 બિલાડી, 1 ફૂતરો → |

કામગીરી પ્રક્રિયા:

- અંતર ગણો ટેસ્ટ પોઇન્ટ અને બધા ટ્રેનિંગ પોઇન્ટ્સ વચ્ચે
- અંતર સોર્ટ કરો અને K નજીકના પડોશીઓ પસંદ કરો
- વોટ ગણો પડોશીઓ વચ્ચે દરેક કલાસમાંથી
- કલાસ અસાઇન કરો મેજોરિટી વોટ સાથે

ડાયગ્રામ:

Mermaid Diagram (Code)

```
{Shaded}
{Highlighting} []
graph LR
    A[ ] --{-{-}{}}--> B[ ]
    B --{-{-}{}}--> C[K ]
    C --{-{-}{}}--> D[ ]
    D --{-{-}{}}--> E[ ]
{Highlighting}
{Shaded}
```

ફાયદાઓ:

- લાગુ કરવામાં સરળ અને સમજવામાં આસાન
- ટ્રેનિંગની જરૂર નથી - આળસુ લર્નિંગ અલોરિધમ

મેમરી ટ્રીક

"K નજીકના પડોશીઓ કલાસિફિકેશન માટે વોટ કરે છે"

પ્રશ્ન 3(અ) અથવા [3 ગુણ]

Multiple linear regression ની કોઈપણ ત્રણ ઉપયોગીતાઓ આપો.

જવાબ

Multiple Linear Regression ની ઉપયોગીતાઓ:

ટેબલ: ઉપયોગીતાઓ

| ઉપયોગીતા | વેરિયેબલ્સ | હેતુ |
|---------------------|------------------------|-----------------------|
| ઘરની કિંમત આગાહી | કદ, સ્થાન, ઉંમર | પ્રોપરીની કિંમત અંદાજ |
| સેલ્સ ફોર્કાસ્ટિંગ | જાહેરાત, સીરિઝન, કિંમત | આવકની આગાહી કરવી |
| મેડિકલ ડાયગ્નોસ્ટિસ | લક્ષણો, ઉંમર, ઇતિહાસ | જોખમ આકારણી |

Multiple Linear Regression એક કન્ટિન્યુઅસ આઉટપુટ વેરિયેબલની આગાહી કરવા માટે બહુવિધ ઇનપુટ વેરિયેબલ્સનો ઉપયોગ કરે છે.

મેમરી ટ્રીક

"બહુવિધ ઇનપુટ્સ, એક આઉટપુટ"

પ્રશ્ન 3(બ) અથવા [4 ગુણ]

Bagging, boosting અને stacking વિગતવાર વાર્ષિકો.

જવાબ

ટેબલ: Ensemble મેથ્ડ્સ

| મેથ્ડ | અભિગમ | ઉદાહરણ |
|----------|---------------------------------------|-------------------------|
| Bagging | પેરેલલ ટ્રેનિંગ, સરેરાશ પરિણામો | Random Forest |
| Boosting | સિક્વેન્શિયલ ટ્રેનિંગ, ભૂલોમાંથી શીખે | AdaBoost |
| Stacking | મેટા-લર્નિંગ મોડલ્સ કન્બાઇન કરે | Neural network combiner |

Bagging વિવિધ ડેટા સબસ્ટેટ્સ પર બહુવિધ મોડલ્સને ટ્રેન કરે છે અને આગાહીઓની સરેરાશ કાઢે છે.

Boosting મોડલ્સને કમિક રીતે ટ્રેન કરે છે, દરેક અગાઉના મોડલની ભૂલોમાંથી શીખે છે.

Stacking બેઝ મોડલ્સની આગાહીઓને કેવી રીતે કન્બાઇન કરવી તે શીખવા માટે મેટા-મોડલનો ઉપયોગ કરે છે.

મેમરી ટ્રીક

"Bag પેરેલલ, Boost સિક્વેન્શિયલ, Stack મેટા"

પ્રશ્ન 3(ક) અથવા [7 ગુણ]

Single linear regression તેની ઉપયોગીતાઓ સાથે વાર્ષિકો.

જવાબ

Single Linear Regression એક ઇનપુટ વેરિયેબલ (X) અને એક આઉટપુટ વેરિયેબલ (Y) વચ્ચે શ્રેષ્ઠ સીધો રેખા સંબંધ શોધે છે.

ફોર્મ્યુલા: $Y = a + bX$

- a: Y-intercept
- b: લાઈનનો Slope

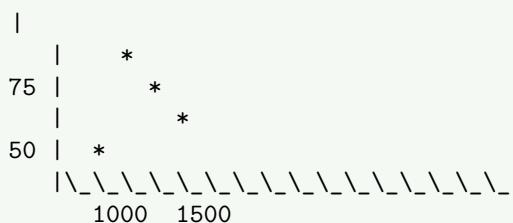
ટેબલ: ઉપયોગ ઉદાહરણ - ઘરની કિંમત વિ કંડ

| ઘરનું કંડ (sq ft) | કિંમત (લાખ) |
|-------------------|-------------|
| 1000 | 50 |
| 1500 | 75 |
| 2000 | 100 |

કામકાજની પ્રક્રિયા:

- ડેટા એક્સિસ કરો ઇનપુટ-આઉટપુટ જોડીઓ સાથે
- પોઇન્ટ્સ પ્લોટ કરો સ્ક્રિટર ગ્રાફ પર
- શ્રેષ્ઠ લાઈન શીધો જે ભૂલ ન્યૂનતમ કરે
- આગાહીઓ કરો લાઈન સમીકરણ વાપરીને

ડાયાગ્રામ:



ઉપયોગીતાઓ:

- સેલ્સ વિ જાહેરાત: વધુ જાહેરાત →
- તાપમાન વિ આઇસક્રીમ સેલ્સ: ગરમ હવામાન →

મેમરી ટ્રીક

“એક X એક Y ની લાઇન સાથે આગાહી કરે છે”

પ્રશ્ન 4(અ) [3 ગુણ]

વ્યાખ્યા આપો: (1) support (2) confidence.

જવાબ

Support માપે છે કે આઇટમસેટ ડેટાસેટમાં કેટલી વાર દેખાય છે.

Confidence માપે છે કે જ્યારે antecedent હાજર હોય ત્યારે consequent માં આઇટમ્સ કેટલી વાર દેખાય છે.

ટેબલ: વ્યાખ્યાઓ

| માપદંડ | ફોર્મ્યુલા | ઉદાહરણ |
|-------------------|---------------------------------|---|
| Support | Count(itemset)/કુલ transactions | બ્રેડ 60% transactions માં દેખાય છે |
| Confidence | Support(A)/Support(A) | બ્રેડ ખરીદનારા 80% લોકો બટર પણ ખરીદે છે |

Support = આવૃત્તિની આવર્તન **Confidence** = નિયમની વિશ્વસનીયતા

મેમરી ટ્રીક

“Support = કેટલી વાર, Confidence = કેટલું વિશ્વસનીય”

પ્રશ્ન 4(બ) [4 ગુણ]

Unsupervised learning ની ઉપયોગીતાઓ વર્ણવો.

જવાબ

ટેબલ: Unsupervised Learning ઉપયોગીતાઓ

| ઉપયોગીતા | હેતુ | ઉદાહરણ |
|----------------------|-------------------------------|------------------------|
| કસ્ટમર સેગમેન્ટેશન | સમાન કસ્ટમર્સને જૂથબદ્ધ કરવા. | માર્કેટિંગ ક્રમેન્ટન્સ |
| ડેટા કમ્પ્યુટેશન | ડેટાનું કદ ઘટાડનું | ઇમેજ કમ્પ્યુટેશન |
| અનોમલી ડિટેક્ષન | અસામાન્ય પેર્ટન શોધવા | ફોડ ડિટેક્ષન |
| રેકમેન્ડેશન સિસ્ટમ્સ | સમાન આઇટમ્સ સુઝાવવા | મ્યુઝિક રેકમેન્ડેશન્સ |

કસ્ટમર સેગમેન્ટેશન લક્ષિત માર્કેટિંગ માટે સમાન ખરીદી વર્તાશૂક ધરાવતા કસ્ટમર્સને જૂથબદ્ધ કરે છે.

ડેટા કમ્પ્યુટેશન પેર્ટન શોધીને અને રિન્ડ-ન્સી દૂર કરીને સ્ટોરેજ સ્પેસ ઘટાડે છે.

અનોમલી ડિટેક્ષન અસામાન્ય પેર્ટન ઓળખે છે જે ફોડ અથવા ભૂલો સૂચવી શકે છે.

મેમરી ટ્રીક

“સેગમેન્ટ કમ્પ્યુટેક્ટ રેકમેન્ડ”

પ્રશ્ન 4(ક) [7 ગુણ]

Apriori અલોરિધમ યોગ્ય ઉદાહરણ સાથે વર્ણવો.

જવાબ

Apriori Algorithm માર્કેટ બાસ્કેટ એનાલિસિસ માટે ફીકવન્ટ આઇટમસેટ્સ શોધે છે અને એસોસિએશન રૂલ્સ જનરેટ કરે છે.

ટેબલ: અલોરિધમ સ્ટેપ્સ

| સ્ટેપ | વર્ણન | ઉદાહરણ |
|-------------------------|-----------------------------|---------------------|
| ફીકવન્ટ 1-itemsets શોધો | વ્યક્તિગત આઇટમ્સ ગણો | {બ્રેડ}:4, {દૂધ}:3 |
| 2-itemsets જનરેટ કરો | ફીકવન્ટ આઇટમ્સ કન્બાઇન કરો | {બ્રેડ, દૂધ}:2 |
| મિનિમમ સપોર્ટ લાગુ કરો | ઇન્ફીકવન્ટ સેટ્સ ફિલ્ટર કરો | support \geq 50% |
| રૂલ્સ જનરેટ કરો | if-then રૂલ્સ બનાવો | બ્રેડ \rightarrow |

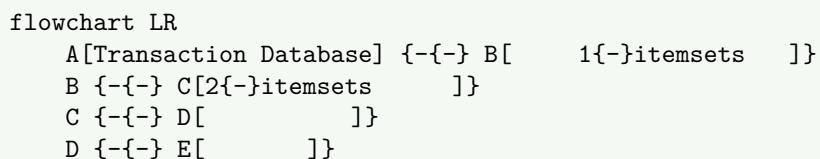
ઉદાહરણ ડેટાસેટ:

- Transaction 1: {બ્રેડ, દૂધ, ઈંડા}
- Transaction 2: {બ્રેડ, દૂધ}
- Transaction 3: {બ્રેડ, ઈંડા}
- Transaction 4: {દૂધ, ઈંડા}

કામકાજની પ્રક્રિયા:

- ડેટાબેઝ સ્કેન કરો આઇટમ ફીકવન્સીઝ ગણાવા માટે
- કેન્ડિક્ટ આઇટમસેટ્સ જનરેટ કરો વધતા કદની
- ઇન્ફીકવન્ટ આઇટમસેટ્સ પૂન કરો મિનિમમ સપોર્ટથી નીચે
- એસોસિએશન રૂલ્સ જનરેટ કરો ફીકવન્ટ આઇટમસેટ્સમાંથી

ડાયાગ્રામ:



મેમરી ટ્રીક

“A-priori જ્ઞાન ફીકવન્ટ પેટર્ન શોધવામાં મદદ કરે છે”

પ્રશ્ન 4(અ) અથવા [3 ગુણ]

Clustering અને Classification ના તફાવતની ચાદી આપો.

જવાબ

ટેબલ: Clustering વિ Classification

| Clustering | Classification |
|--|--|
| અનસુપરવાઈઝ લન્નિંગ લેબલ કરેલ ડેટા નથી સમાન ડેટાને જૂથબદ્ધ કરે છે | સુપરવાઈઝ લન્નિંગ લેબલ કરેલ ટ્રેનિંગ ડેટા વાપરે છે પૂર્વનિર્ધારિત લેબલ્સ અસાઇન કરે છે |

Clustering અનલેબલ ડેટામાં છુપાયેલા જૂથો શોધે છે.

Classification ટ્રેન કરેલા મોડલ્સ વાપરીને નવા ડેટાને જાણીતી કેટેગરીઓમાં અસાઇન કરે છે.

મેમરી ટ્રીક

“Cluster = અજાણ્યા જૂથો, Classify = જાણીતા લેબલ્સ”

પ્રશ્ન 4(બ) અથવા [4 ગુણ]

Clustering ની પ્રોસેસ વિગતવાર વર્ણવો.

જવાબ

ટેબલ: Clustering પ્રોસેસ સ્ટેપ્સ

| સ્ટેપ | વર્ણન | હેતુ |
|--------------------|----------------------------------|-------------------------------------|
| ડેટા પ્રિપેરેશન | ડેટા સાફ અને નોર્મલાઇઝ કરો | ગુણવત્તાપૂર્ણ ઇનપુટ સુનિશ્ચિત કરવું |
| ડિસ્ટન્સ મેટ્રિક | સમાનતાનું માપ પસંદ કરો | Euclidean, Manhattan |
| અલ્ગોરિધમ સિલેક્શન | કલસ્ટરિંગ મેથડ પસંદ કરો | K-means, Hierarchical |
| કલસ્ટર વેલિડેશન | કલસ્ટર ગુણવત્તાનું મૂલ્યાંકન કરો | Silhouette score |

Clustering પ્રોસેસ તેમની લાક્ષણિકતાઓના આધારે સમાન ડેટા પોઇન્ટ્સને એકસાથે જૂથબદ્ધ કરે છે.
મુખ્ય નિર્ણયોમાં કલસ્ટરસની સંખ્યા અને યોગ્ય ડિસ્ટન્સ મેટ્રિક્સ પસંદ કરવાનો સમાવેશ થાય છે.
વેલિડેશન સુનિશ્ચિત કરે છે કે કલસ્ટર્સ અર્થપૂર્ણ અને સારી રીતે અલગ છે.

મેમરી ટ્રીક

"પ્રિપેર ડિસ્ટન્સ અલ્ગોરિધમ વેલિડેટ"

પ્રશ્ન 4(ક) અથવા [7 ગુણ]

K-means clustering અલ્ગોરિધમ યોગ્ય ઉદાહરણ સાથે વર્ણવો.

જવાબ

K-means વિધિન-કલસ્ટર સમ ઓફ સ્કવેર્સ ન્યૂનતમ કરીને ડેટાને K કલસ્ટર્સમાં વિભાજિત કરે છે.

ટેબલ: અલ્ગોરિધમ સ્ટેપ્સ

| સ્ટેપ | વર્ણન | ઉદાહરણ |
|----------------------------|---|-------------------|
| સેન્ટ્રોઇઝસ ઇનિશિયલાઇઝ કરો | રેન્ડમ K સેન્ટર પોઇન્ટ્સ | C1(2,3), C2(8,7) |
| પોઇન્ટ્સ અસાઇન કરો | દરેક પોઇન્ટ નજીકના સેન્ટ્રોઇઝને | Point(1,2) → C1 |
| સેન્ટ્રોઇઝસ અપડેટ કરો | અસાઇન થયેલા પોઇન્ટ્સનો મીન | નવું C1(1.5, 2.5) |
| રિપીટ કરો | સેન્ટ્રોઇઝસ હલનયલન બંધ ન થાય ત્યાં સુધી | કન્વર્જન્સ |

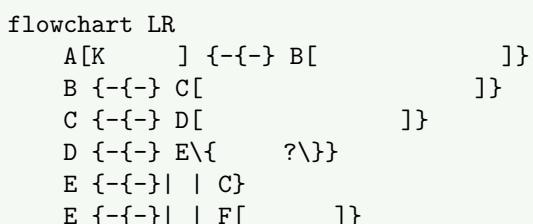
ઉદાહરણ: કસ્ટમર આવક વિ ઉમર

- કસ્ટમર 1: (આવક=30k, ઉમર=25)
- કસ્ટમર 2: (આવક=35k, ઉમર=30)
- કસ્ટમર 3: (આવક=70k, ઉમર=45)
- કસ્ટમર 4: (આવક=75k, ઉમર=50)

કામકાજની પ્રક્રિયા:

- K=2 પસંદ કરો યુવા/વૃદ્ધ કસ્ટમર્સ માટે કલસ્ટર્સ
- સેન્ટ્રોઇઝસ ઇનિશિયલાઇઝ કરો રેન્ડમ રીતે
- અંતર ગણો દરેક કસ્ટમરથી સેન્ટ્રોઇઝસ સુધી
- કસ્ટમર્સ અસાઇન કરો નજીકના સેન્ટ્રોઇઝને
- સેન્ટ્રોઇઝ પોઝિશન્સ અપડેટ કરો અસાઇન થયેલા કસ્ટમર્સના કેન્દ્રમાં
- સ્થિર ન થાય ત્યાં સુધી રિપીટ કરો

ડાયાગ્રામ:



ફાયદાઓ:

- સરળ અને જડપી મોટા ડેટાસેટ્સ માટે
- ગોળાકાર કલસ્ટર્સ સાથે સારું કામ કરે છે

મેમરી ટ્રીક

"K સેન્ટ્રોઇઝસ તેમના અસાઇન થયેલા પોઇન્ટ્સનો મીન કરે છે"

પ્રશ્ન 5(અ) [3 ગુણ]

Matplotlib ની ઉપયોગીતાઓની યાદી આપો.

જવાબ

ટેબલ: Matplotlib ઉપયોગીતાઓ

| ઉપયોગીતા | હેતુ | ઉદાહરણ |
|------------------------|----------------------------|----------------------------|
| ડેટા વિઝ્યુલાઇઝન | ચાર્ટ્સ અને ગ્રાફ્સ બનાવવા | બાર ચાર્ટ્સ, હિસ્ટોગ્રામ્સ |
| સાયાન્ટિક્ષિક પ્લોટિંગ | સંશોધન પ્રોન્ટેશન્સ | ગાણિતિક ફંક્શન્સ |
| ડેશબોર્ડ કિએશન | ઇન્ટરેક્ટિવ ડિસ્પ્લે | બિજનેસ મેટ્રિક્સ |

Matplotlib સ્ટેટિક, એનિમેટેડ અને ઇન્ટરેક્ટિવ વિઝ્યુલાઇઝનન્સ બનાવવા માટે Python ની પ્રાથમિક પ્લોટિંગ લાઇબ્રેરી છે. મુખ્ય ફ્રીચર્સમાં બહુવિધ પ્લોટ ટાઇપ્સ માટેનું સપોર્ટ અને કસ્ટમાઇઝેબલ સ્ટાઇલિંગનો સમાવેશ થાય છે.

મેમરી ટ્રીક

“Mat-plot-lib = ગાણિત પ્લોટિંગ લાઇબ્રેરી”

પ્રશ્ન 5(બ) [4 ગુણ]

હોરિડોન્ટલ અને વર્ટિકલ લાઇન પ્લોટ કરવાનો કોડ matplotlib ની મદદથી લખો.

જવાબ

કોડ પ્લોક:

```
import matplotlib.pyplot as plt

#plt.figure(figsize=(8, 6))

#plt.axvline(x=3, color={red}, linestyle={{-}{-}}, label={      })

#plt.axhline(y=2, color={blue}, linestyle={{-}}, label={      })

plt.xlabel({X{-} })
plt.ylabel({Y{-} })
plt.title({          })
plt.legend()
plt.grid(True)
plt.show()
```

મુખ્ય ફંક્શન્સ:

- axvline(): વર્ટિકલ લાઇન બનાવે છે
- axhline(): હોરિડોન્ટલ લાઇન બનાવે છે

મેમરી ટ્રીક

“axvline = વર્ટિકલ, axhline = હોરિડોન્ટલ”

પ્રશ્ન 5(ક) [7 ગુણ]

Scikit-Learn ની વિશેષતાઓ અને ઉપયોગીતાઓ સમજાવો.

જવાબ

ટેબલ: Scikit-Learn વિશેષતાઓ

| વિશેષતા | વર્ણન | ઉદાહરણ |
|---------------------|-----------------------|--------------------|
| સરળ API | ઉપયોગમાં સરળ ઇન્ટરફેસ | fit(), predict() |
| બહુવિધ અલ્ગોરિધમ્સ | વિવિધ ML મેથડ્સ | SVM, Random Forest |
| ડેટા પ્રીપ્રોસેસિંગ | બિલ્ટ-ઇન ડેટા ટૂલ્સ | StandardScaler |
| મોડલ ઇવેલ્યુઅશન | પરફોર્મન્સ મેટ્રિક્સ | accuracy_score |

Scikit-Learn ડેટા એનાલિસિસ માટે સરળ ટૂલ્સ પ્રદાન કરતી Python ની સૌથી લોકપ્રિય મશીન લર્નિંગ લાઇબ્રેરી છે.

મુખ્ય શક્તિઓ:

- સુસંગત ઇન્ટરફેસ બધા અલ્ગોરિધમ્સમાં
- ઉત્કૃષ્ટ દસ્તાવેજુકરણ ઉદાહરણો સાથે
- સંક્ષિય કમ્પ્યુનિટી સપોર્ટ અને ડેવલપમેન્ટ

ઉપયોગીતાઓ:

- કલાસિફિકેશન: ઈમેઇલ સ્પામ ડિટેક્શન
- રિઝ્રેશન: ઘરની કિંમત આગાહી
- કલસ્ટરિંગ: કસ્ટમર સેગમેટેશન
- ડાયમેન્શનાલિટી રિડક્શન: ડેટા વિજ્યુલાઇઝન

ડાયાગ્રામ:

Mermaid Diagram (Code)

```

{Shaded}
{Highlighting} []
graph TD
    A[Scikit{-Learn}] --> B[      ]
    A --> C[      ]
    A --> D[      ]
    A --> E[      ]
    B --> F[SVM, Decision Trees]
    C --> G[Linear, Polynomial]
    D --> H[K{-}means, DBSCAN]
    E --> I[Scaling, Encoding]
{Highlighting}
{Shaded}

```

મેમરી ટ્રીક

“Scikit = મશીન લર્નિંગ માટે સાયન્સ કિટ”

પ્રશ્ન 5(અ) અથવા [૩ ગુણ]

NumPy નો મશીન લર્નિંગના સંદર્ભમાં ઉપયોગ આપો.

જવાબ

ટેબલ: ML માં NumPy નો હેતુ

| હેતુ | વર્ણન | ફાયદો |
|------------------------|---------------------------|--|
| ન્યુમેરિકલ કમ્પ્યુટિંગ | અડ્પી array ઓપરેશન્સ | કાર્યક્ષમ ગણતરીઓ |
| ફાઉન્ડેશન લાઇબ્રેરી | અન્ય લાઇબ્રેરીઓ માટે આધાર | Pandas, Scikit-learn તેનો ઉપયોગ કરે છે |
| ગાણિતિક ફંક્શન્સ | બિલ્ટ-ઇન મેથ ઓપરેશન્સ | સ્ટેટિસ્ટિક્સ, લિનિયર આલજીબ્રા |

NumPy Python મશીન લર્નિંગ એપ્લિકેશન્સમાં ન્યુમેરિકલ કમ્પ્યુટિંગ માટે પાયો પ્રદાન કરે છે. મોટા ડેટાસેટ્સ હન્ડલ કરવા અને ગાણિતિક ઓપરેશન્સ કાર્યક્ષમ રીતે કરવા માટે જરૂરી છે.

પ્રશ્ન 5(બ) અથવા [4 ગુણ]

csv ફાઈલને pandas માં ઇમ્પોર્ટ કરવાના સ્ટેપ લખો.

જવાબ

કોડ બ્લોક:

```
import pandas as pd

# 1: pandas
# 2: read\_csv()
data = pd.read\_csv({filename.csv})

# 3:
print(data.head())

# :
data = pd.read\_csv({file.csv},
                    delimiter={,},
                    header=0,
                    index\_col=0)
```

સ્ટેપ્સ:

1. pandas ઇમ્પોર્ટ કરો લાઇબ્રેરી
2. read_csv() વાપરો ફાઈલનેમ સાથે
3. ડેટા વેરિફાઇ કરો head() મેથડ સાથે

પ્રશ્ન 5(ક) અથવા [7 ગુણ]

Pandas ની વિશેષતાઓ અને ઉપયોગીતાઓ સમજાવો.

જવાબ

ટેબલ: Pandas વિશેષતાઓ

| વિશેષતા | વર્ણન | ઉદાહરણ |
|------------------|----------------------------|-------------------------|
| ડેટા સ્ટ્રક્ચર્સ | DataFrame અને Series | ટેબ્યુલર ડેટા હેન્ડલિંગ |
| ડેટા I/O | બહુવિધ ફોર્મેટ્સ રીડ/રાઇટ | CSV, Excel, JSON |
| ડેટા કલીનિંગ | મિસિંગ વેલ્ચુઝ હેન્ડલ કરવા | dropna(), fillna() |
| ડેટા એનાલિસિસ | સ્ટેટિસ્ટિકલ ઓપરેશન્સ | groupby(), describe() |

Pandas મશીન લર્નિંગ પ્રોજેક્ટ્સમાં Python માં પ્રાથમિક ડેટા મેનિપ્યુલેશન લાઇબ્રેરી છે.

મુખ્ય ક્ષમતાઓ:

- ડેટા લોડિંગ વિવિધ ફાઇલ ફોર્મેટ્સમાંથી
- ડેટા કલીનિંગ અને પ્રીપ્રોસેસિંગ ઓપરેશન્સ
- ડેટા ટ્રાન્સફોર્મેશન અને રીશેપિંગ
- સ્ટેટિસ્ટિકલ એનાલિસિસ અને એગ્રિગેશન

ઉપયોગીતાઓ:

- ડેટા પ્રીપ્રોસેસિંગ: ML પહેલાં ડેટાસેટ્સ સાફ્ કરવા
- એક્સ્પ્લોરેટરી એનાલિસિસ: ડેટા પેર્ટન સમજવા
- ફીચર એન્જિનિયરિંગ: નવા વેરિયેબલ્સ બનાવવા
- ડેટા ઇન્ટીગ્રેશન: બહુવિધ ડેટા સોટો મર્જ કરવા

ડાયાગ્રામ:

Mermaid Diagram (Code)

```
{Shaded}
{Highlighting} []
graph LR
    A[ ] --> B[Pandas DataFrame]
    B --> C[ ]
    C --> D[ ]
    D --> E[ ]
    E --> F[ML]
{Highlighting}
{Shaded}
```

ફાયદાઓ:

- સાહજિક સિન્ટેક્સ ડેટા ઓપરેશન્સ માટે
- હાઇ પરફોર્મન્સ ઓપ્ટિમાઇઝ ઓપરેશન્સ સાથે
- ઇન્ટીગ્રેશન અન્ય ML લાઇબ્રેરીઓ સાથે

મેમરી ટ્રીક

“Pandas = એનાલિસિસ માટે પેનલ ડેટા”