

Fundamentals of Machine Learning (4341603) - Summer 2023 Solution

Milav Dabgar

July 18, 2023

પ્રશ્ન 1(a) [3 ગુણ]

હૃમન લર્નિંગને વ્યાખ્યાયિત કરો. હૃમન લર્નિંગના પ્રકારોની ચાદી બનાવો.

જવાબ

હૃમન લર્નિંગ એ પ્રક્રિયા છે જેના દ્વારા માનવીઓ અનુભવ, અભ્યાસ અથવા સૂચનાઓ દ્વારા નવા જ્ઞાન, કૌશળ્યો, વર્તન મેળવે છે અથવા હાલનાઓમાં ફેરફાર કરે છે.

કોષ્ટક 1. હૃમન લર્નિંગના પ્રકારો

પ્રકાર	વર્ણન
સુપરવાઇઝ લર્નિંગ	શિક્ષક/માર્ગદર્શકની મદદથી શીખવું
અનસુપરવાઇઝ લર્નિંગ	બાહ્ય માર્ગદર્શન વિના સ્વ-નિર્દશિત શીખવું
રિઝનેફર્મેન્ટ લર્નિંગ	ફીડબેક સાથે પ્રયાસ અને ભૂલ દ્વારા શીખવું

મેમરી ટ્રીક

SUR - Supervised, Unsupervised, Reinforcement

પ્રશ્ન 1(b) [4 ગુણ]

કવાલિટેટિવ ડેટા અને કવોન્ટિટેટિવ ડેટા વર્ચ્યે તફાવત આપો.

જવાબ

કોષ્ટક 2. કવાલિટેટિવ vs કવોન્ટિટેટિવ ડેટા

લક્ષણ	કવાલિટેટિવ ડેટા	કવોન્ટિટેટિવ ડેટા
પ્રકૃતિ	વર્ણનાત્મક, કેટેગોરિકલ	સંઘાતક, માપી શકાય તેવું
વિશ્લેષણ	વ્યક્તિગત અર્થધટન	આંકડાકીય વિશ્લેષણ
ઉદાહરણો	રંગો, નામો, લિંગ	ઉંચાઈ, વજન, ઉંમર
પ્રતિનિધિત્વ	શબ્દો, કેટેગરીઓ	સંઘાઓ, ગ્રાફ્સ

મેમરી ટ્રીક

QUAN-Numbers, QUAL-Words

પ્રશ્ન 1(c) [7 ગુણ]

મશીન લર્નિંગના વિવિધ પ્રકારોની સરખામણી કરો.

જવાબ

કોષ્ટક 3. મશીન લર્નિંગના પ્રકારોની સરખામણી

પ્રકાર	ટ્રેનિંગ ડેટા	ધોય	ઉદાહરણો
સુપરવાઈઝડ	લેબલ ડેટા	પરિણામોની આગાહી	કલાસિફિકેશન, રિગ્રેશન
અનસુપરવાઈઝડ	અનલેબલ ડેટા	પેટન શોધવા	કલસ્ટરિંગ, એસોસિએશન
રિદનફોર્મેન્ટ	રિવોર્ડ/પેનલ્ટી	રિવોર્ડ મેક્સિમાઇઝ કરવા	ગેમિંગ, રોબોટિક્સ

મુખ્ય તક્ષાવતો:

- સુપરવાઈઝડ: ટ્રેનિંગ માટે ઇનપુટ-આઉટપુટ જોડીનો ઉપયોગ કરે છે
- અનસુપરવાઈઝડ: ડેટામાં છુપાયેલા પેટન શોધે છે
- રિદનફોર્મેન્ટ: પર્યાવરણ સાથે કિયાપ્રતિક્રિયા દ્વારા શીખે છે

મેમરી ટ્રીક

SUR-LAP: Supervised-Labeled, Unsupervised-Reveal, Reinforcement-Action

પ્રશ્ન 1(c OR) [7 ગુણ]

મશીન લર્નિંગ વ્યાખ્યાયિત કરો. મશીન લર્નિંગની કોઈપણ ચાર એપ્લિકેશનને ટૂકમાં સમજાવો.

જવાબ

મશીન લર્નિંગ આર્ટિફિશિયલ ઇન્ટેલિજન્સનો ઉપબિભાગ છે જે કમ્પ્યુટરોને સ્પષ્ટ પ્રોગ્રામિંગ વિના ડેટામાંથી શીખવા અને નિર્ણયો લેવા સક્ષમ બનાવે છે.

કોષ્ટક 4. ચાર એપ્લિકેશનો

એપ્લિકેશન	વર્ણન
ઇમેઇલ સ્પામ ડિટેક્શન	ઇમેઇલને સ્પામ અથવા વૈધ તરીકે વર્ગીકૃત કરે છે
ઇમેજ રેકૉર્ડિંગ	ફોટોમાં ઓળખેકટ્સ ઓળખે છે
રેકમેડેશન સિસ્ટમ	યુઝરને પ્રોડક્ટ્સ/કટેન સૂચવે છે
મેડિકલ ડાયગ્નોસ્ટિસ	રોગોની શોધમાં ડોક્ટરોની મદદ કરે છે

મેમરી ટ્રીક

SIRM - Spam, Image, Recommendation, Medical

પ્રશ્ન 2(a) [3 ગુણ]

નીચેના ઉદાહરણોનો યોગ્ય ડેટા પ્રકાર જણાવો.

જવાબ**કોષ્ટક 5. ડેટા પ્રકાર વર્ગીકરણ**

ઉદાહરણ	ડેટા પ્રકાર
વિદ્યાર્થીઓની રાષ્ટ્રીયતા	ક્રેટોરિકલ (નોમિનલ)
વિદ્યાર્થીઓની શિક્ષણ સ્થિતિ	ક્રેટોરિકલ (ઓડિનલ)
વિદ્યાર્થીઓની ઉંચાઈ	ન્યુમેરિકલ (કન્ટિન્યુઅસ)

મેમરી ટ્રીક

NCN - Nominal, Categorical, Numerical

પ્રશ્ન 2(b) [4 ગુણ]

ડેટા પ્રી-પ્રોસેસિંગ ટૂકમાં સમજાવો.

જવાબ

ડેટા પ્રી-પ્રોસેસિંગ એ મશીન લર્નિંગ અલ્ગોરિધમ માટે કાચા ડેટાને તૈયાર કરવાની તકનીક છે.

કોષ્ટક 6. મુખ્ય સ્ટેપ્સ

સ્ટેપ	હેતુ
ડેટા કલીનિંગ	ભૂલો અને અસંગતતાઓ દૂર કરવી
ડેટા ઇન્ટીગ્રેશન	બહુવિધ સ્ત્રોતોમાંથી ડેટાને જોડવો
ડેટા ટ્રાન્સફોર્મેશન	ડેટાને યોગ્ય ફોર્મેટમાં બદલવો
ડેટા રિડક્ષન	માહિતી જાળવીને ડેટાનું કદ ઘટાડવું

મેમરી ટ્રીક

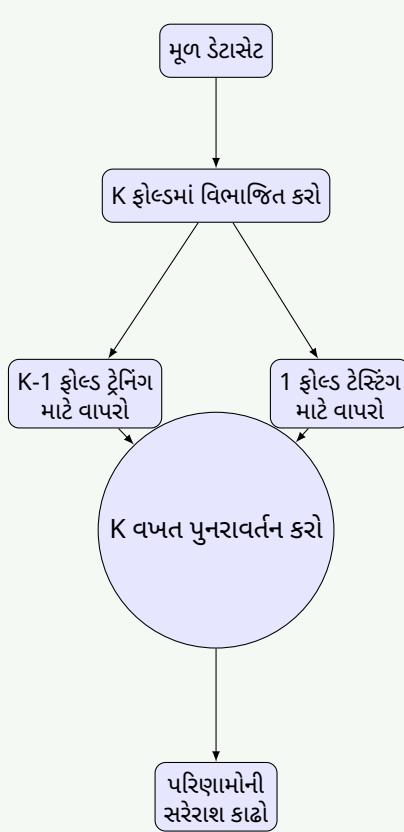
CITR - Clean, Integrate, Transform, Reduce

પ્રશ્ન 2(c) [7 ગુણ]

K-ફોલ્ડ કોસ વેલિડેશન વિગતવાર સમજાવો.

જવાબ

K-ફોલ્ડ કોસ વેલિડેશન એ ડેટાને K સમાન ભાગોમાં વિભાજિત કરીને મોડેલ પરફોર્મન્સ મૂલ્યાંકનની તકનીક છે.



આકૃતિ 1. K-Fold કોસ વેલિડેશન પ્રક્રિયા

સ્ટેપ્સ:

- વિભાજન: ડેટાસેટને K સમાન ભાગોમાં વહેંચો
- ટ્રેનિંગ: K-1 ફોલ્ડનો ઉપયોગ ટ્રેનિંગ માટે કરો
- ટેસ્ટ: બાકીના ફોલ્ડનો ઉપયોગ વેલિડેશન માટે કરો
- પુનરાવર્તન: K વખત પ્રક્રિયા કરો
- સરેરાશ: સરેરાશ પરફોર્મન્સ કાઢો

ફાયદા:

- ઓવરફિલ્ડિંગ ઘટાડે છે
- મર્યાદિત ડેટાનો બહેતર ઉપયોગ
- વધુ વિશ્વસનીય પરફોર્મન્સ અંદાજ

મેમરી ટ્રીક

DTRA - Divide, Train, Repeat, Average

પ્રશ્ન 2(a OR) [3 ગુણ]

નીચેના શબ્દો વ્યાખ્યાપિત કરો: i) Mean, ii) Outliers, iii) Interquartile range

જવાબ

કોષ્ટક 7. અંકડાકીય શબ્દો

શબ્દ	વ્યાખ્યા
Mean	ડેટાસેટમાં બધી વેલ્યુઝની સરેરાશ
Outliers	અન્ય ડેટા પોઇન્ટ્સથી નોંધપાત્ર રીતે અલગ ડેટા પોઇન્ટ્સ
Interquartile Range	75મા અને 25મા પર્સોનાઇલ વર્ચ્યેનો તફાવત

પ્રેરણ ટ્રીક

MOI - Mean, Outliers, Interquartile

પ્રશ્ન 2(b OR) [4 ગુણ]

કન્ફ્યુશન મેટ્રિક્સની રચના સમજાવો.

જવાબ

કન્ફ્યુશન મેટ્રિક્સ સ્ક્રૂચર:

કોષ્ટક 8. કન્ફ્યુશન મેટ્રિક્સ

	આગાહી પોઝિટિવ	આગાહી નેગેટિવ
વાસ્તવિક પોઝિટિવ	True Positive (TP)	False Negative (FN)
વાસ્તવિક નેગેટિવ	False Positive (FP)	True Negative (TN)

ઘટકો:

- TP: સાચી રીતે આગાહી કરેલા પોઝિટિવ કેસો
- TN: સાચી રીતે આગાહી કરેલા નેગેટિવ કેસો
- FP: ખોટી રીતે પોઝિટિવ તરીકે આગાહી કરેલા
- FN: ખોટી રીતે નેગેટિવ તરીકે આગાહી કરેલા

પ્રેરણ ટ્રીક

TTFF - True True, False False

પ્રશ્ન 2(c OR) [7 ગુણ]

ફીચર સબસેટની પસંદગી પર ટૂંકી નોંધ લખો.

જવાબ

ફીચર સબસેટ સિલેક્શન એ મૂળ ફીચર સેટમાંથી સંબંધિત ફીચર્સ પસંદ કરવાની પ્રક્રિયા છે.

કોષ્ટક 9. મેથ્ડ્સ

મેથ્ડ	વર્ણન
ફિલ્ટર મેથ્ડ્સ	ફીચર્સ રેન્ક કરવા અંકડાકીય માપદંડોનો ઉપયોગ
રેપર મેથ્ડ્સ	ફીચર સબસેટ્સ મૂલ્યાંકન માટે ML અલ્ગોરિધમનો ઉપયોગ
ઓફ્લોડ મેથ્ડ્સ	મોડેલ ટ્રેનિંગ દરમિયાન ફીચર સિલેક્શન

ફાયદા:

- **ઘટાડેલી જટિલતા**: ઓછા ફીચર્સ, સરળ મોડેલ્સ
- **સુધારેલ પરફોર્મન્સ**: નોઈજ અને અપ્રસ્તુત ફીચર્સ દૂર કરે છે

- **જડપી ટ્રેનિંગ:**: ઓછો કમ્પ્યુટેશનલ ઓવરહેડ

લોકપ્રિય તકનીકો:

- Chi-square ટેસ્ટ
- Recursive Feature Elimination
- LASSO રેગ્યુલરાઇઝેશન

મેમરી ટ્રીક

FWE - Filter, Wrapper, Embedded

પ્રશ્ન 3(a) [3 ગુણા]

પ્રેડિક્ટિવ મોડેલ અને ડીસ્ક્રિપ્ટિવ મોડેલ વચ્ચેનો તફાવત આપો.

જવાબ

કોષ્ટક 10. પ્રેડિક્ટિવ vs ડીસ્ક્રિપ્ટિવ મોડેલ

લક્ષણ	પ્રેડિક્ટિવ મોડેલ	ડીસ્ક્રિપ્ટિવ મોડેલ
હેતુ	ભાવિ પરિણામોની આગાહી	વર્તમાન પેર્ટન સમજવા
આઉટપુટ	આગાહીઓ/વર્ગીકરણ	અંતર્દૃષ્ટિ/સારાંશ
ઉદાહરણો	રિંગ્રેશન, કલાસિફિકેશન	કલસ્ટરિંગ, એસોસિએશન રૂલ્સ

મેમરી ટ્રીક

PF-DC: Predictive-Future, Descriptive-Current

પ્રશ્ન 3(b) [4 ગુણા]

કલાસિફિકેશન અને રિંગ્રેશન વચ્ચેના તફાવતની ચર્ચા કરો.

જવાબ

કોષ્ટક 11. કલાસિફિકેશન vs રિંગ્રેશન

પાસું	કલાસિફિકેશન	રિંગ્રેશન
આઉટપુટ	ડિસ્ક્રીટ કેટેગરીઓ	કાન્ટિન્યુઅસ વેલ્યુઝ
ધ્યાય	કલાસ લેબલ્સની આગાહી	ન્યૂમેરિકલ વેલ્યુઝની આગાહી
ઉદાહરણો	સ્પામ ડિટેક્શન, ઇમેજ રેક્ગ્રિશન	કિંમત આગાહી, તાપમાન
મૂલ્યાંકન	Accuracy, precision, recall	MSE, RMSE, R-squared

મેમરી ટ્રીક

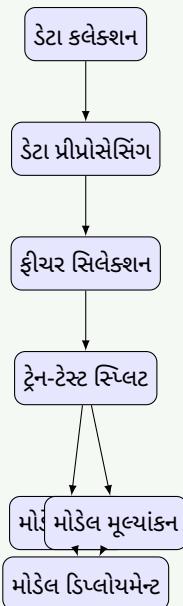
CCNM - Classification-Categories, Regression-Numbers

પ્રશ્ન 3(c) [7 ગુણા]

કલાસિફિકેશનને વ્યાખ્યાયિત કરો. કલાસિફિકેશન લર્નિંગના સ્ટેપને વિગતોમાં સમજાવો.

જવાબ

કલાસિફિકેશન એ સુપરવાઇઝ લર્નિંગ તકનીક છે જે ઇનપુટ ડેટા માટે ડિસ્ક્રીટ કલાસ લેબલ્સની આગાહી કરે છે.



આકૃતિ 2. કલાસિફિકેશન લર્નિંગ સ્ટેપ્સ

વિગતવાર સ્ટેપ્સ:

- ડેટા કલેક્શન: લેબલ્ડ ટ્રેનિંગ ડેટા એકત્ર કરવો
- પ્રીપ્રોસેસિંગ: ડેટાને સાફ કરવો અને તૈયાર કરવો
- ફીચર સિલેક્શન: સંબંધિત લક્ષણો પરસંદ કરવા
- ડેટા સ્પલિટ: ટ્રેનિંગ અને ટેસ્ટિંગ સેટમાં વિભાજન
- ટ્રેનિંગ: ટ્રેનિંગ ડેટાનો ઉપયોગ કરીને મોડેલ બનાવવું
- મૂલ્યાંકન: મોડેલ પરફોર્મન્સ ચકાસવી
- ડિપ્લોયમેન્ટ: આગાહીઓ માટે મોડેલનો ઉપયોગ

મેમરી ટ્રીક

DCFSTED - Data, Clean, Features, Split, Train, Evaluate, Deploy

પ્રશ્ન 3(a OR) [3 ગુણ]

બેન્ચિંગ અને બૂસ્ટિંગ વચ્ચેનો તફાવત આપો.

જવાબ

કોષ્ટક 12. બેન્ચિંગ vs બૂસ્ટિંગ

લક્ષણ	બેન્ચિંગ	બૂસ્ટિંગ
સેમ્પલિંગ	બૂટસ્ટ્રેપ સેમ્પલિંગ	ક્રમાનુગત વેઈટેડ સેમ્પલિંગ
ટ્રેનિંગ	પેરેલલ ટ્રેનિંગ	ક્રમાનુગત ટ્રેનિંગ
ફોકસ	વેરિયન્સ ઘટાડવું	બાયસ ઘટાડવું

મેમરી ટ્રીક

BPV-BSB: Bagging-Parallel-Variance, Boosting-Sequential-Bias

પ્રશ્ન 3(b OR) [4 ગુણ]

લોજિસ્ટિક રિગ્રેશનના વિવિધ પ્રકારો સંક્ષિપ્તમાં સમજાવો.

જવાબ

કોષ્ટક 13. લોજિસ્ટિક રિગ્રેશનના પ્રકારો

પ્રકાર	ક્લાસો	ઉપયોગ
બાઇનરી	2 ક્લાસો	હા/ના, પાસ/ફેલ
મલ્ટિનોમિયલ	3+ ક્લાસો (અવ્યવસ્થિત)	રંગ જાતિ
ઓર્ડિનલ	3+ ક્લાસો (ક્રમાંકિત)	રેટિંગ સ્કેલ

મેમરી ટ્રીક

BMO - Binary, Multinomial, Ordinal

પ્રશ્ન 3(c OR) [7 ગુણ]

k-NN અલ્ગોરિધમ લખો અને તેના ઉપયોગ બતાવો.

જવાબ

K-નિયરેસ્ટ નેઇબર્સ (k-NN) એ લેઝી લર્નિંગ અલ્ગોરિધમ છે જે k નજીકના પડોશીઓના બહુમતી ક્લાસના આધારે ડેટા પોઇન્ટ્સને વર્ગીકૃત કરે છે. અલ્ગોરિધમ સ્ટેપ્સ:

1. k ની વેલ્યુ પસંદ કરો
2. બધા ટ્રેનિંગ પોઇન્ટ્સ સાથે અંતર કાઢો
3. k નજીકના પડોશીઓ પસંદ કરો
4. ક્લાસિફિકેશન માટે: બહુમતી મત; રિગ્રેશન માટે: k પડોશીઓની સરેરાશ
5. ટેસ્ટ પોઇન્ટને ક્લાસ/વેલ્યુ અસાઇન કરો

અંતર ગણતરી:

- યુક્લિડિયન ડિસ્ટન્સ: $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

એપ્લિકેશનો:

- એકમેન્ડેશન સિસ્ટમ્સ: સમાન યુઝર પ્રાધાન્યો
- ઇમેજ રેકૉર્ડિંગ: પેર્ટન મેર્ચિંગ
- મેડિકલ ડાયગ્નોસ્ટિસ: લક્ષણોની સમાનતા

ફાયદા:

- અમલમાં મૂકવામાં સરળ
- ટ્રેનિંગ-નથી
- નાના ડેટાસેટ સાથે સારં કામ કરે છે

મેમરી ટ્રીક

CDSA - Choose, Distance, Select, Assign

પ્રશ્ન 4(a) [3 ગુણ]

સપોર્ટ વેક્ટર મશીનની એપ્લિકેશનોની યાદી બનાવો.

જવાબ

કોષ્ટક 14. SVM એપ્લિકેશનો

એપ્લિકેશન	ડોમેન
ટેક્સ્ટ કલાસિફિકેશન	ડોક્યુમેન્ટ કેટેગોરિઝેશન
ઇમેજ રેકૉર્ડિંગ	ફિલ્મ ડિટેક્શન
બાયોઇન્ફોર્મેટિક્સ	જીન કલાસિફિકેશન

મેમરી ટ્રીક

TIB - Text, Image, Bio

પ્રશ્ન 4(b) [4 ગુણ]

k-means અલોરિધમ માટે સ્યુડો કોડ બનાવો.

જવાબ

K-means સ્યુડો કોડ:

```

1 BEGIN K-means
2   1. Initialize k cluster centroids randomly
3   2. REPEAT
4     a. Assign each point to nearest centroid
5     b. Update centroids to mean of assigned points
6     c. Calculate total within-cluster sum of squares
7   3. UNTIL convergence or max iterations
8   4. RETURN final clusters and centroids
9 END

```

મેમરી ટ્રીક

IAUC - Initialize, Assign, Update, Check

પ્રશ્ન 4(c) [7 ગુણ]

અનસુપરવાઈઝ લર્નિંગની એપ્લિકેશનો લખો અને સમજાવો.

જવાબ

અનસુપરવાઈઝ લર્નિંગ લેબલ ઉદાહરણો વિના ડેટામાં છુપાયેલા પેટર્ન શોધે છે.

કોષ્ટક 15. મુખ્ય એપ્લિકેશનો

એપ્લિકેશન	વર્ણન	ઉદાહરણ
કસ્ટમર સેગ્મેન્ટેશન	વર્તન પ્રમાણે ગ્રાહકોનું ગુપ્ચિંગ	માર્કેટ રિસર્ચ
અનોમેલી ડિટેક્શન	અસામાન્ય પેટર્ન ઓળખવા	ફોડ ડિટેક્શન
ડેટા કમ્પ્રેશન	ડાયમેન્શનાલિટી ઘટાડવી	ઇમેજ કમ્પ્રેશન
એસોસિએશન રૂલ્સ	આઇટમ સંબંધો શોધવા	માર્કેટ બાર્સ્કેટ વિશ્લેષણ

કલસ્ટરિંગ એપ્લિકેશનો:

- માર્કેટ રિસર્ચ: કસ્ટમર ગુપ્ચિંગ
- સોશિયલ નેટવર્ક વિશ્લેષણ: કમ્પ્યુનિટી ડિટેક્શન
- જીન સીકવેન્સિંગ: બાયોલોજિકલ કલાસિફિકેશન

ડાયમેન્શનાલિટી રિડક્શન:

- વિશ્યુઅલાઇઝેન્સ: હાઇ-ડાયમેન્શનલ ડેટા પ્લોટિંગ
- ફ્રીગ્રાર એક્સ્પ્રેક્શન: નોઇજ રિડક્શન

મેમરી ટ્રીક

CADA - Customer, Anomaly, Data, Association

પ્રશ્ન 4(a OR) [3 ગુણ]

રિશેશનની એપ્લિકેશનોની યાદી બનાવો.

જવાબ

કોષ્ટક 16. રિશેશન એપ્લિકેશનો

એપ્લિકેશન	હેતુ
સ્ટોક પ્રાઇસ પ્રેડિક્શન	ફાઈનાન્શિયલ ફોરકાસ્ટિંગ
સેલ્સ ફોરકાસ્ટિંગ	બિઝનેસ પ્લાનિંગ
મેડિકલ ડાયગ્નોસ્ટિસ	રિસ્ક એસેસમેન્ટ

મેમરી ટ્રીક

SSM - Stock, Sales, Medical

પ્રશ્ન 4(b OR) [4 ગુણ]

નીચેના શબ્દો વ્યાખ્યાયિત કરો: i) Support ii) Confidence

જવાબ

કોષ્ટક 17. એસોસિએશન રૂલ શબ્દો

શબ્દ	વ્યાખ્યા	ફોર્મ્યુલા
Support	ડેટાબેઝમાં આઇટમસેટની આવર્તન	$Support(A) = \frac{ A }{ D }$
Confidence	રૂલની શરતી સંભાવના	$Confidence(A \rightarrow B) = \frac{Support(A \cup B)}{Support(A)}$

ઉદાહરણ:

- જો 30% ટ્રાન્జેક્શનમાં બ્રેડ અને દૂધ હોય: $Support = 0.3$
- જો 80% બ્રેડ ખરીદનારાઓ દૂધ પણ ખરીદે: $Confidence = 0.8$

મેમરી ટ્રીક

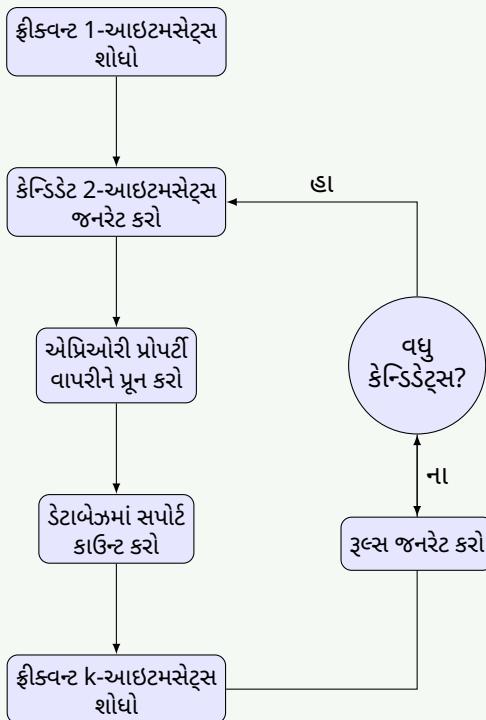
SF-CP: Support-Frequency, Confidence-Probability

પ્રશ્ન 4(c OR) [7 ગુણ]

apriori algorithm ને વિગતવાર સમજાવો.

જવાબ

Apriori અલોરિધમ એપ્રિયોરી પ્રોપર્ટીનો ઉપયોગ કરીને ટ્રોન્-ઝેક્શનલ ડેટામાં ફીકવન્ટ આઇટમસેટ્સ શોધે છે.



આકૃતિ 3. Apriori અલોરિધમ પ્રક્રિયા

એપ્રિયોરી પ્રોપર્ટી:

- જો આઇટમસેટ ફીકવન્ટ છે, તો તેના બધા સબસેટ્સ ફીકવન્ટ છે
- જો આઇટમસેટ ઇનફીકવન્ટ છે, તો તેના બધા સુપરસેટ્સ ઇનફીકવન્ટ છે

સ્ટેપ્સ:

- **ડેટાબેઝ સ્કેન**: 1-આઇટમ સપોર્ટ કાઉન્ટ કરો
- **કેન્ડિક્ટ્ટેસ જનરેટ**: ફીકવન્ટ k-આઇટમસેટ્સમાંથી k+1 આઇટમસેટ્સ બનાવો
- **પ્રૂન**: ઇનફીકવન્ટ સબસેટ્સ સાથેના કેન્ડિક્ટ્ટેસ દૂર કરો
- **સપોર્ટ કાઉન્ટ**: કેન્ડિક્ટ ફીકવન્સી માટે ડેટાબેઝ સ્કેન કરો
- **પુનરાવર્તન**: નવા ફીકવન્ટ આઇટમસેટ્સ ન મળે ત્યાં સુધી

એપ્લિકેશનો:

- માર્કટ બાસ્કેટ વિશ્લેષણ
- વેબ યુઝર પેર્ટન
- પ્રોટીન સીકવન્સ

મેમરી ટ્રીક

SGPCR - Scan, Generate, Prune, Count, Repeat

પ્રશ્ન 5(a) [3 ગુણ]

matplotlib ના મુખ્ય ફીચર્સની યાદી બનાવો.

જવાબ

કોષ્ટક 18. Matplotlib ફીચર્સ

ફીચર	વર્ણન
મલિટપલ પ્લોટ ટાઇપ્સ	લાઇન, બાર, સ્કેટર, હિસ્ટોગ્રામ
કસ્ટમાઇઝેશન	કલર્સ, સ્ટાઇલ્સ, લેબલ્સ
એક્સપોર્ટ ઓપ્શન્સ	PNG, PDF, SVG ફાર્મેટ્સ

મેમરી ટ્રીક

MCE - Multiple, Customization, Export

પ્રશ્ન 5(b) [4 ગુણ]

Numpy ના પ્રોગ્રામમાં iris ડેટાસેટ કેવી રીતે લોડ કરવો? સમજાવો.

જવાબ

NumPy માં Iris ડેટાસેટ લોડ કરવું:

```

1 import numpy as np
2 from sklearn.datasets import load_iris
3
4 # iris ડેટાસેટ લોડ કરો
5 iris = load_iris()
6 data = iris.data # ફીચર્સ
7 target = iris.target # લેબલ્સ

```

સ્ટેપ્સ:

- Import: જરૂરી લાઇબ્રેરીઓ import કરો
- Load: sklearn ના load_iris() ફંક્શનનો ઉપયોગ કરો
- Extract: ફીચર્સ અને ટાર્ગેટ એરે મેળવો
- Access: .data અને .target એટ્રિબ્યુટ્સનો ઉપયોગ કરો

મેમરી ટ્રીક

ILEA - Import, Load, Extract, Access

પ્રશ્ન 5(c) [7 ગુણ]

Pandas ની વિશેષતાઓ અને એપ્લિકેશનો સમજાવો.

જવાબ

Pandas એ Python માટે શક્તિશાળી ડેટા મેનિપ્યુલેશન અને વિશ્લેષણ લાઇબ્રેરી છે.

કોષ્ટક 19. મુખ્ય ફીચર્સ

ફીચર	વર્ણન
DataFrame	2D લેબલ ડેટા સ્ટ્રક્ચર
Series	1D લેબલ એરે
Data I/O	વિવિધ ફાઇલ ફોર્મેટ્સ વાંચવા/લખવા
Data Cleaning	મિસિંગ વેલ્ચુઝ હેન્ડલ કરવા
Grouping	ગુપ અને એગ્રીગેટ ઓપરેશન્સ

એપ્લિકેશનો:

- ડેટા એનાલિસિસ: અંકડાકીય વિશ્લેષણ.
- ડેટા કલીનિંગ: ML માટે પ્રીપ્રોસેસિંગ
- ફાઇનાન્શિયલ એનાલિસિસ: સ્ટોક માર્કેટ ડેટા
- વેબ સ્કેપિંગ: HTML ટેબલ્સ પાર્સ કરવા

સામાન્ય ઓપરેશન્સ:

- ડેટા વાંચવો: pd.read_csv(), pd.read_excel()
- ફિલ્ટરિંગ: df[df['column'] > value]
- ગ્રુપિંગ: df.groupby(column).mean()

મેમરી ટ્રીક

DSDCG - DataFrame, Series, Data I/O, Cleaning, Grouping

પ્રશ્ન 5(a OR) [3 ગુણ]

matplotlib ની એપ્લિકેશનોની યાદી બનાવો.

જવાબ

કોષ્ટક 20. Matplotlib એપ્લિકેશનો

એપ્લિકેશન	હેતુ
સાયાન્ટિફિક વિઝ્યુઅલાઇઝેશન	રિસર્ચ ડેટા પ્લોટિંગ
બિજનેસ એનાલિટિક્સ	ડેશબોર્ડ બનાવવું
એજ્યુકેશનલ કન્ટેન્ટ	શિક્ષણ સામગ્રી

મેમરી ટ્રીક

SBE - Scientific, Business, Educational

પ્રશ્ન 5(b OR) [4 ગુણ]

Pandas માં csv ફાઇલ ઇમ્પોર્ટ કરવાના સ્ટેપ્સ લખો અને સમજાવો.

જવાબ

Pandas માં CSV ઇમ્પોર્ટ કરવાના સ્ટેપ્સ:

```

1 import pandas as pd
2
3 # સ્ટેપ 1: pandas લાઇબ્રેરી import કરો
4 # સ્ટેપ 2: read_csv() ફંક્શનનો ઉપયોગ કરો

```

```

5 df = pd.read_csv('filename.csv')
6
7 # વૈકલ્પિક પેરામીટ્ર્સ
8 df = pd.read_csv('file.csv',
9     header=0,    # પ્રથમ પંક્તિહેડર તરીકે
10    sep=',',    # કોમા સેપરેટર
11    index_col=0) # પ્રથમ કોલમ ઇન્ડેક્સ તરીકે

```

પ્રક્રિયા:

- Import:** pandas લાઇબ્રેરી import કરો
- Read:** pd.read_csv() ફંક્શનનો ઉપયોગ કરો
- Specify:** ફાઇલ પાથ અને પેરામીટર્સ ઉમેરો
- Store:** DataFrame વેરિએબલમાં અસાઇન કરો

મેમરી ટ્રીક

IRSS - Import, Read, Specify, Store

પ્રશ્ન 5(c OR) [7 ગુણ]

Scikit-Learn ની વિશેષતાઓ અને એપ્લિકેશનો સમજાવો.

જવાબ

Scikit-Learn એ Python માટે વ્યાપક મશીન લર્નિંગ લાઇબ્રેરી છે.

કોષ્ટક 21. મુખ્ય ફીચર્સ

ફીચર	વર્ણન
અલ્ગોરિધમ્સ	કલાસિફિકેશન, રિગ્રેશન, કલસ્ટરિંગ
પ્રીપ્રોસેસિંગ	ડેટા સ્કેલિંગ અને ટ્રાન્સફોર્મેશન
મોડેલ સિલેક્શન	કોસ-વેલિડેશન અને ગ્રિડ સર્ચ
મેટ્રિક્સ	પરફોર્મન્સ મૂલ્યાંકન ટૂલ્સ

એપ્લિકેશનો:

- હેલ્પાર્કર: રોગ આગાહી
- ફાઇનાન્સ: કેર્ડિટ સ્કોરિંગ
- માર્કાટિંગ: કસ્ટમર સેગમેન્ટેશન
- ટેકનોલોજી: રેકમેન્ડેશન સિસ્ટમ્સ

અલ્ગોરિધમ કેટેગરીઓ:

- સુપરવાઈઝ: SVM, Random Forest, Linear Regression
- અનસુપરવાઈઝ: K-means, DBSCAN, PCA
- એ-સેમ્બલ: Bagging, Boosting

વર્કફુલો:

- ડેટા તૈયારી: પ્રીપ્રોસેસિંગ
- મોડેલ સિલેક્શન: અલ્ગોરિધમ પસંદ કરો
- ટ્રેનિંગ: ડેટા પર મોડેલ ફિટ કરો
- મૂલ્યાંકન: પરફોર્મન્સ આકારો
- આગાહી: ફોરકારટ બનાવો

મેમરી ટ્રીક

APME - Algorithms, Preprocessing, Metrics, Evaluation