

Fundamentals of Machine Learning (4341603) - Winter 2023 Solution

Milav Dabgar

February 2, 2024

પ્રશ્ન 1(a) [3 ગુણ]

Human learning વ્યાખ્યાયિત કરો અને સમજાવો કે machine learning human learning થી કેવી રીતે અલગ છે?

જવાબ

Human Learning એ અનુભવ, અવલોકન અને તર્ક દ્વારા જ્ઞાન મેળવવાની પ્રક્રિયા છે.

કોષ્ટક 1. Human Learning વિ Machine Learning

પાસાં	Human Learning	Machine Learning
પદ્ધતિ	અનુભવ, પ્રયાસ અને ભૂલ	ડેટા અને અલ્ગોરિધમ
ઝડપ	ધીમી, ક્રમશઃ	ઝડપી પ્રોસેસિંગ
ડેટા જરૂરિયાત	મર્યાદિત ઉદાહરણો જોઈએ	મોટા ડેટાસેટ જરૂરી

Machine Learning: ડેટામાં પેટર્ન ઓળખવા માટે અલ્ગોરિધમનો ઉપયોગ કરીને આપોઆપ શીખવાની પ્રક્રિયા.

મેમરી ટ્રીક

“Humans Experience, Machines Analyze Data (HEMAD)”

પ્રશ્ન 1(b) [4 ગુણ]

ફાઇનાન્સ અને બેંકિંગમાં મશીન લર્નિંગના ઉપયોગનું વર્ણન કરો.

જવાબ

કોષ્ટક 2. ફાઇનાન્સ અને બેંકિંગમાં ઉપયોગો

ઉપયોગ	હેતુ	ફાયદો
Fraud Detection	શંકાસ્પદ ટ્રાન્ઝેક્શન ઓળખવા	નાણાકીય નુકસાન ઘટાડવું
Credit Scoring	લોન ડિફોલ્ટ રિસ્ક આંકવી	વધુ સારા લેન્ડિંગ નિર્ણયો
Algorithmic Trading	આપોઆપ ટ્રેડિંગ નિર્ણયો	ઝડપી માર્કેટ રિસ્પોન્સ

Risk Assessment: ગ્રાહકની ક્રેડિટવર્થીનેસની આગાહી માટે ML ડેટાનું વિશ્લેષણ કરે છે. Customer Service: NLP નો ઉપયોગ કરીને ચેટબોટ્સ 24/7 સપોર્ટ આપે છે. Regulatory Compliance: શંકાસ્પદ પ્રવૃત્તિઓ માટે આપોઆપ મોનિટરિંગ.

મેમરી ટ્રીક

“Finance Needs Smart Analysis (FNSA)”

પ્રશ્ન 1(c) [7 ગુણ]

સુપરવાઇઝ્ડ લર્નિંગ, અનસુપરવાઇઝ્ડ લર્નિંગ અને રિઇન્ફોર્સમેન્ટ લર્નિંગ વચ્ચે તફાવત આપો.

જવાબ

કોષ્ટક 3. તુલનાત્મક કોષ્ટક

લક્ષણ	Supervised	Unsupervised	Reinforcement
ડેટા પ્રકાર	લેબલ્ડ ડેટા	અનલેબલ્ડ ડેટા	પર્યાવરણ ઇન્ટરેક્શન
લક્ષ્ય	આઉટપુટની આગાહી	પેટર્નો શોધવા	રિવોર્ડ વધારવા
ઉદાહરણો	Classification	Clustering	Game playing
Feedback	તાત્કાલિક	કંઈ નહીં	વિલંબિત પુરસ્કારો

Supervised Learning: સાચા જવાબો સાથે શિક્ષક દ્વારા માર્ગદર્શિત શીખવું.

Unsupervised Learning: ડેટામાં છુપાયેલા પેટર્નોની સ્વ-શોધ.

Reinforcement Learning: પુરસ્કાર/દંડ સાથે ટ્રાયલ અને એરર દ્વારા શીખવું.

મેમરી ટ્રીક

“Supervised Teachers, Unsupervised Explores, Reinforcement Rewards (STUER)”

પ્રશ્ન 1(c) OR [7 ગુણ]

મશીન લર્નિંગમાં વપરાતા વિવિધ TOOLS અને ટેકનોલોજી સમજાવો.

જવાબ

કોષ્ટક 4. ML Tools અને Technologies

કેટેગરી	Tools	હેતુ
Programming	Python, R, Java	અલ્ગોરિથમ ઇમ્પ્લિમેન્ટેશન
Libraries	Scikit-learn, TensorFlow	તૈયાર અલ્ગોરિથમ
Visualization	Matplotlib, Seaborn	ડેટા વિઝ્યુઅલાઇઝેશન
Data Processing	Pandas, NumPy	ડેટા મેનિપ્યુલેશન

મુખ્ય ટેકનોલોજીઓ:

- **Cloud Platforms:** AWS, Google Cloud સ્કેલેબલ કમ્યુટિંગ માટે
- **Development Environments:** Jupyter Notebook, Google Colab
- **Big Data Tools:** મોટા ડેટાસેટ માટે Spark, Hadoop

મેમરી ટ્રીક

“Python Libraries Visualize Data Effectively (PLVDE)”

પ્રશ્ન 2(a) [3 ગુણ]

એક ઉદાહરણ સાથે outliers ને વ્યાખ્યાયિત કરો.

જવાબ

વ્યાખ્યા: Outliers એવા ડેટા પોઇન્ટ્સ છે જે ડેટાસેટમાં અન્ય અવલોકનોથી નોંધપાત્ર રીતે અલગ હોય છે.

કોષ્ટક 5. ઉદાહરણ: વિદ્યાર્થીઓની ઊંચાઈ

ઊંચાઈ (cm)	વર્ગીકરણ
165, 170, 168, 172	સામાન્ય મૂલ્યો
195	Outlier (ખૂબ ઊંચું)
140	Outlier (ખૂબ નીચું)

શોધ: Quartiles થી $1.5 \times \text{IQR}$ થી વધુ મૂલ્યો. અસર: આંકડાકીય વિશ્લેષણ અને મોડલ પર્ફોર્મન્સને અસર કરી શકે.

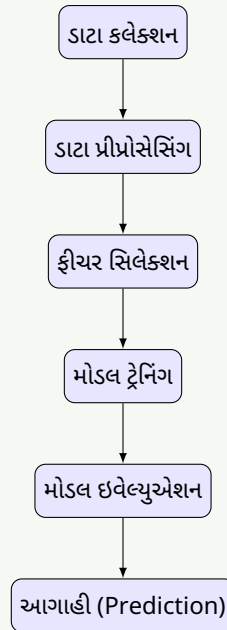
મેમરી ટ્રીક

“Outliers Stand Apart (OSA)”

પ્રશ્ન 2(b) [4 ગુણ]

રીગ્રેશન સ્ટેપ્સ વિગતવાર સમજાવો.

જવાબ



આકૃતિ 1. રીગ્રેશન પ્રોસેસ સ્ટેપ્સ

વિગતવાર સ્ટેપ્સ:

- **Data Collection:** ઇનપુટ-આઉટપુટ જોડી સાથે સંબંધિત ડેટાસેટ એકત્રિત કરવું.
- **Preprocessing:** ડેટા સાફ કરવું, ખોવાયેલા મૂલ્યો સંભાળવા, features ને normalize કરવા.
- **Feature Selection:** લક્ષ્યને અસર કરતા સંબંધિત variables પસંદ કરવા.
- **Model Training:** આગાહીની ભૂલો ન્યૂનતમ કરવા માટે રીગ્રેશન લાઇન ફિટ કરવી.

મેમરી ટ્રીક

“Data Preprocessing Features Train Evaluation Predicts (DPFTEP)”

પ્રશ્ન 2(c) [7 ગુણ]

ચોક્કસાઈ વ્યાખ્યાયિત કરો અને નીચેના binary classifier ની confusion matrix માટે વિવિધ માપન પરિમાણો શોધો જેમ કે 1. Accuracy 2. Precision.

જવાબ

Confusion Matrix વિશ્લેષણ:

કોષ્ટક 6. આપેલ Confusion Matrix		
	અનુમાનિત ના	અનુમાનિત હા
વાસ્તવિક ના	10 (TN)	3 (FP)
વાસ્તવિક હા	2 (FN)	15 (TP)

ગણતરીઓ:

- **Accuracy** = $(TP + TN) / (TP + TN + FP + FN) = (15 + 10) / (15 + 10 + 3 + 2) = 25/30 = 0.8333$
- પરિણામ: 83.33%
- **Precision** = $TP / (TP + FP) = 15 / (15 + 3) = 15/18 = 0.8333$
- પરિણામ: 83.33%

વ્યાખ્યાઓ:

- **Accuracy**: કુલ આગાહીઓમાંથી સાચી આગાહીઓનું પ્રમાણ.
- **Precision**: બધી positive આગાહીઓમાંથી true positive આગાહીઓનું પ્રમાણ.

મેમરી ટ્રીક

"Accuracy Counts All, Precision Picks Positives (ACAPP)"

પ્રશ્ન 2(a) OR [3 ગુણ]

Feature સબસેટ પસંદગીના મૂળભૂત પગલાઓને ઓળખો.

જવાબ



આકૃતિ 2. Feature Subset Selection Process

મૂળભૂત પગલાઓ:

1. **Generation**: Features ના વિવિધ સંયોજનો બનાવવા.
2. **Evaluation**: પ્રત્યેક સબસેટને પર્ફોર્મન્સ મેટ્રિક્સ વાપરીને ટેસ્ટ કરવા.
3. **Selection**: માપદંડોના આધારે શ્રેષ્ઠ સબસેટ પસંદ કરવા.

મેમરી ટ્રીક

"Generate, Evaluate, Select (GES)"

પ્રશ્ન 2(b) OR [4 ગુણ]

KNN અલ્ગોરિથમની તાકાત અને નબળાઈની ચર્ચા કરો.

જવાબ

કોષ્ટક 7. KNN અલ્ગોરિધમ વિશ્લેષણ

તાકાતો	નબળાઈઓ
સમજવામાં સરળ	કમ્પ્યુટેશનલી મોંઘું
Training ની જરૂર નથી (Lazy)	અપ્રસ્તુત features ને સંવેદનશીલ
Non-linear ડેટા સાથે કામ કરે	High dimensions સાથે performance ઘટે
નાના ડેટાસેટ માટે અસરકારક	શ્રેષ્ઠ K value પસંદગી જરૂરી

મુખ્ય મુદ્દાઓ:

- **Lazy Learning:** સ્પષ્ટ training phase ની જરૂર નથી.
- **Distance-Based:** પડોશીની નજીકતા આધારિત વર્ગીકરણ.

મેમરી ટ્રીક

“Simple but Slow, Effective but Expensive (SBSEBE)”

પ્રશ્ન 2(c) OR [7 ગુણ]

ભૂલ-દર વ્યાખ્યાયિત કરો અને નીચેના binary classifier ની confusion matrix માટે વિવિધ માપન પરિમાણો શોધો જેમ કે 1. Error value 2. Recall.

જવાબ

Confusion Matrix વિશ્લેષણ:

કોષ્ટક 8. આપેલ Confusion Matrix		
	અનુમાનિત ના	અનુમાનિત હા
વાસ્તવિક ના	20 (TN)	3 (FP)
વાસ્તવિક હા	2 (FN)	15 (TP)

ગણતરીઓ:

- **Error Rate** = $(FP + FN) / (Total) = (3 + 2) / (15 + 20 + 3 + 2) = 5 / 40 = 0.125$
- પરિણામ: **12.5%**
- **Recall** = $TP / (TP + FN) = 15 / (15 + 2) = 15 / 17 = 0.8824$
- પરિણામ: **88.24%**

વ્યાખ્યાઓ:

- **Error Rate:** કુલ આગાહીઓમાંથી ખોટી આગાહીઓનું પ્રમાણ.
- **Recall:** વાસ્તવિક positives માંથી સાચી રીતે ઓળખાયેલાનું પ્રમાણ.

મેમરી ટ્રીક

“Error Excludes, Recall Retrieves (EERR)”

પ્રશ્ન 3(a) [3 ગુણ]

Unsupervised learning ના કોઈ પણ ત્રણ ઉદાહરણો આપો.

જવાબ

કોષ્ટક 9. Unsupervised Learning ઉદાહરણો

ઉદાહરણ	વર્ણન	ઉપયોગ
Customer Segmentation	વર્તન દ્વારા ગ્રાહકોને જૂથબદ્ધ કરવા	માર્કેટિંગ વ્યૂહરચના
Document Classification	વિષયો દ્વારા દસ્તાવેજો ગોઠવવા	માહિતી પુનઃપ્રાપ્તિ
Gene Sequencing	સમાન DNA પેટર્ન જૂથબદ્ધ કરવા	તબીબી સંશોધન

અન્ય ઉદાહરણો:

- **Market Basket Analysis:** ઉત્પાદન ખરીદીના પેટર્ન શોધવા.
- **Anomaly Detection:** ડેટામાં અસામાન્ય પેટર્ન શોધવા.

મેમરી ટ્રીક

“Customers, Documents, Genes Group Automatically (CDGGA)”

પ્રશ્ન 3(b) [4 ગુણ]

નીચેના ડેટા માટે સરેરાશ અને મધ્યક શોધો: 4,6,7,8,9,12,14,15,20

જવાબ

ડેટા: 4, 6, 7, 8, 9, 12, 14, 15, 20 (પહેલેથી જ સોર્ટ થયેલ)

સરેરાશ (Mean) ગણતરી:

- સરવાળો = $4 + 6 + 7 + 8 + 9 + 12 + 14 + 15 + 20 = 95$
- ગણતરી = 9
- **Mean** = $95/9 = 10.56$

મધ્યક (Median) ગણતરી:

- $N = 9$ (એકી સંખ્યા)
- મધ્ય મૂલ્ય = $(N + 1)/2 = 5$ મી સ્થિતિ
- 5મી સ્થિતિએ મૂલ્ય 9 છે
- **Median** = 9

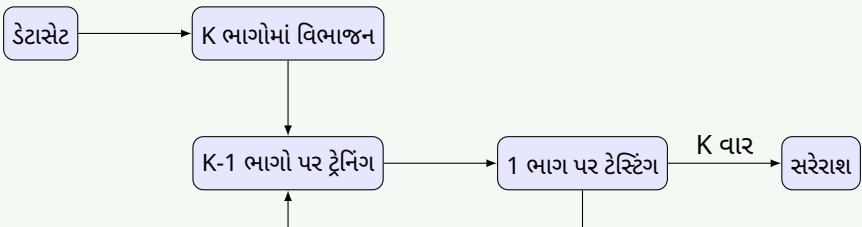
મેમરી ટ્રીક

“Mean Averages All, Median Middle Value (MAAMV)”

પ્રશ્ન 3(c) [7 ગુણ]

k-ફોલ્ડ ક્રોસ વેલિડેશન પદ્ધતિનું વિગતવાર વર્ણન કરો.

જવાબ



આકૃતિ 3. K-Fold Cross Validation

પ્રોસેસ સ્ટેપ્સ:

- 1. ડેટા વિભાજન: ડેટાને K સમાન ભાગોમાં વહેંચવું.
- 2. પુનરાવર્તિત Training: મોડલ ટ્રેનિંગ માટે K-1 folds નો ઉપયોગ કરવો.
- 3. Validation: મોડલને બાકીના fold પર ટેસ્ટ કરવું.
- 4. સરેરાશ: K વાર પુનરાવર્તન કરો અને performance મેટ્રિક્સની સરેરાશ કાઢો.

ફાયદાઓ:

- નિષ્પક્ષ અંદાજ: દરેક ડેટા પોઇન્ટ training અને testing બંને માટે વપરાય છે.
- Overfitting ઘટાડવું: અનેક validation રાઉન્ડ વિશ્વસનીયતા વધારે છે.

મેમરી ટ્રીક

“K-fold Keeps Keen Knowledge (KKKK)”

પ્રશ્ન 3(a) OR [3 ગુણ]

Multiple linear રીગ્રેશનની કોઈ પણ ત્રણ એપ્લિકેશન આપો.

જવાબ

કોષ્ટક 10. Multiple Linear Regression એપ્લિકેશન

એપ્લિકેશન	Variables	હેતુ
House Price Prediction	Size, location, age	રિયલ એસ્ટેટ વેલ્યુએશન
Sales Forecasting	Marketing spend, season	બિઝનેસ પ્લાનિંગ
Medical Diagnosis	Symptoms, age, history	રોગની આગાહી

મેમરી ટ્રીક

“Houses, Sales, Medicine Predict Multiple Variables (HSMPV)”

પ્રશ્ન 3(b) OR [4 ગુણ]

નીચેના ડેટા માટે માનક વિચલન શોધો: 4,15,20,28,35,45

જવાબ

ડેટા: 4, 15, 20, 28, 35, 45 (N=6)

Step 1: સરેરાશ ગણતરી (Mean)

- સરવાળો = 147
- Mean (\bar{x}) = $147/6 = 24.5$

Step 2: વર્ગ વિચલન (Squared Deviations)

- $(4 - 24.5)^2 = 420.25$, $(15 - 24.5)^2 = 90.25$, ...

Step 3: Variance અને Std Dev

- વર્ગ વિચલનોનો સરવાળો = 1073.5
- Variance (σ^2) = $1073.5/6 = 178.92$
- માનક વિચલન (σ) = $\sqrt{178.92} = 13.376$

મેમરી ટ્રીક

“Deviation Measures Data Spread (DMDS)”

પ્રશ્ન 3(c) OR [7 ગુણ]

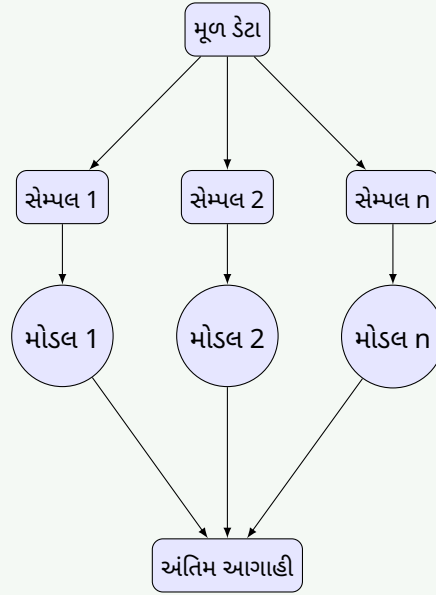
બેગિંગ અને બૂસ્ટિંગને વિગતવાર સમજાવો.

જવાબ

કોષ્ટક 11. Bagging વિ Boosting

પાસું	Bagging	Boosting
વ્યૂહરચના	સમાંતર training	ક્રમિક training
ડેટા સેમ્પલિંગ	રેન્ડમ (replacement સાથે)	વેઇટેડ સેમ્પલિંગ
લક્ષ્ય	Variance ઘટાડે	Bias ઘટાડે

Bagging (Bootstrap Aggregating): ડેટાના રેન્ડમ સબસેટ્સનો ઉપયોગ કરીને સમાંતરમાં બહુવિધ સ્વતંત્ર મોડેલ્સને ટ્રેન કરે છે અને તેમની આગાહીઓની સરેરાશ કાઢે છે.



આકૃતિ 4. Bagging Process

Boosting: મોડેલ્સને ક્રમિક રીતે ટ્રેન કરે છે, જ્યાં દરેક નવું મોડેલ અગાઉના મોડેલ્સ દ્વારા કરવામાં આવેલી ભૂલો પર ધ્યાન કેન્દ્રિત કરે છે.

મેમરી ટ્રીક

“Bagging Builds Parallel, Boosting Builds Sequential (BBPBS)”

પ્રશ્ન 4(a) [3 ગુણ]

વ્યાખ્યાયિત કરો: Support, Confidence.

જવાબ

કોષ્ટક 12. Association Rule મેટ્રિક્સ

મેટ્રિક	વ્યાખ્યા અને ફોર્મ્યુલા
Support	ટ્રાન્ઝેક્શનમાં itemset ની આવર્તન. $Support(A) = Count(A)/Total$
Confidence	નિયમની શરતી સંભાવના. $Confidence(A \rightarrow B) = Support(A \cup B)/Support(A)$

ઉદાહરણ:

- **Support:** 60% ટ્રાન્ઝેક્શનમાં બ્રેડ છે.
- **Confidence:** 80% બ્રેડ ખરીદનારા લોકો બટર પણ ખરીદે છે.

મેમરી ટ્રીક

“Support Shows Frequency, Confidence Shows Connection (SSFC)”

પ્રશ્ન 4(b) [4 ગુણ]

લોજિસ્ટિક રીગ્રેશનની કોઈ પણ બે એપ્લિકેશનને સમજાવો.

જવાબ

કોષ્ટક 13. Logistic Regression એપ્લિકેશન

એપ્લિકેશન	વર્ણન	પરિણામ
Email Spam	સ્પામ ડિટેક્શન	Spam/Not Spam
Medical Diagnosis	લક્ષણો પરથી રોગની આગાહી	રોગ/સ્વસ્થ
Credit Approval	લોન રિસ્ક આકારણી	મંજૂર/નામંજૂર

મુખ્ય લાક્ષણિકતાઓ:

- **Binary Classification:** સંભાવના (0 થી 1) આગાહી કરે છે.
- **Sigmoid Function:** આઉટપુટને S-shaped વળાંકમાં મેપ કરે છે.

મેમરી ટ્રીક

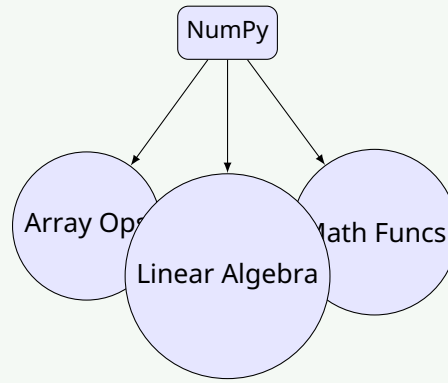
“Logistic Limits Linear Logic (LLLL)”

પ્રશ્ન 4(c) [7 ગુણ]

Machine learning માં Numpy અને Pandas ના મુખ્ય હેતુની ચર્ચા કરો.

જવાબ

NumPy મોટા, મલ્ટી-ડાયમેન્શનલ એરે અને મેટ્રિક્સિસ માટે સપોર્ટ પૂરો પાડે છે. **Pandas** ન્યૂમેરિકલ ટેબલ્સ અને ટાઇમ સીરિઝના મેનિપ્યુલેશન માટે ડેટા સ્ટ્રક્ચર્સ પૂરા પાડે છે.



આકૃતિ 5. NumPy Features

કોષ્ટક 14. તુલના

Library	પ્રાથમિક હેતુ	મુખ્ય Features
NumPy	Numerical Computing	N-dim arrays, broadcasting
Pandas	Data Manipulation	DataFrames, cleaning

મેમરી ટ્રીક

“NumPy Numbers, Pandas Processes Data (NNPD)”

પ્રશ્ન 4(a) OR [3 ગુણ]

સુપરવાઇઝ્ડ લર્નિંગના કોઈ પણ ત્રણ ઉદાહરણો આપો.

જવાબ

કોષ્ટક 15. Supervised Learning ઉદાહરણો

ઉદાહરણ	પ્રકાર	Input → Output
Email Classification	Classification	Email features → Spam/Not
House Price Prediction	Regression	House features → કિંમત
Image Recognition	Classification	Pixels → Object Class

મેમરી ટ્રીક

“Emails, Houses, Images Learn Supervised (EHILS)”

પ્રશ્ન 4(b) OR [4 ગુણ]

એપ્રિઓરી અલ્ગોરિધમના કોઈ પણ બે એપ્લિકેશનો સમજાવો.

જવાબ

કોષ્ટક 16. Apriori એપ્લિકેશન

એપ્લિકેશન	વર્ણન
Market Basket Analysis	એકસાથે ખરીદાતી વસ્તુઓ શોધવી (દા.ત., બ્રેડ અને બટર).
Web Usage Mining	વેબસાઇટ UX સુધારવા માટે વિશ્વસનીય નેવિગેશન પેટર્ન શોધવી.

મેમરી ટ્રીક

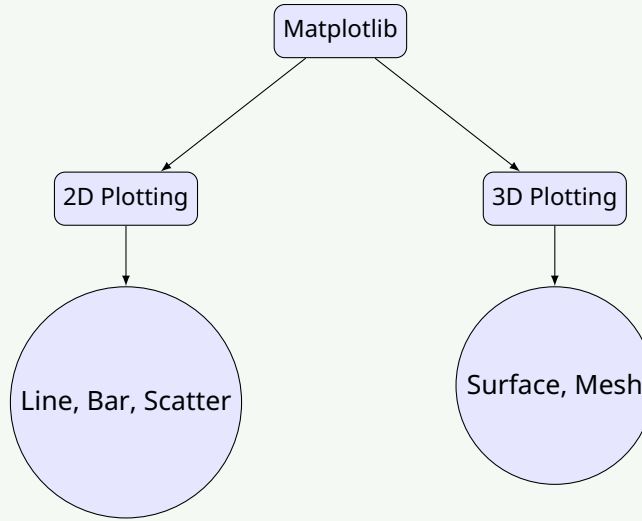
“Apriori Analyzes Associations Automatically (AAAA)”

પ્રશ્ન 4(c) OR [7 ગુણ]

Matplotlib ની વિશેષતાઓ અને એપ્લિકેશનો સમજાવો.

જવાબ

Matplotlib Python માં સ્ટેટિક, એનિમેટેડ અને ઇન્ટરેક્ટિવ વિઝ્યુલાઇઝેશન બનાવવા માટેની વ્યાપક લાઇબ્રેરી છે.



આકૃતિ 6. Matplotlib ક્ષમતાઓ

એપ્લિકેશન્સ:

- **Data Exploration:** ડેટા સમજવા માટે હિસ્ટોગ્રામ, સ્કેટર પ્લોટ.
- **Model Performance:** લોસ કર્વ્સ અને એક્ચ્યુરસી પ્લોટ કરવા.
- **Result Presentation:** પ્રકાશન-ગુણવત્તાવાળા ગ્રાફ્સ.

મેમરી ટ્રીક

“Matplotlib Makes Meaningful Visual Displays (MMVD)”

પ્રશ્ન 5(a) [3 ગુણ]

Numpy ના મુખ્ય features ની યાદી બનાવો.

જવાબ

NumPy ના Features:

- **N-dimensional Arrays:** ઝડપી અને કાર્યક્ષમ મલ્ટી-ડાયમેન્શનલ એરે ઓબ્જેક્ટ (ndarray).
- **Broadcasting:** વિવિધ આકારોના એરે પર ઓપરેશન્સ કરવા માટેના ફંક્શન્સ.
- **Linear Algebra:** મેટ્રિક્સ ઓપરેશન્સ અને Fourier ટ્રાન્સફોર્મ્સ માટે બિલ્ટ-ઇન સપોર્ટ.
- **C/C++ Integration:** C/C++ અને Fortran કોડને ઇન્ટિગ્રેટ કરવા માટેના સાધનો.

મેમરી ટ્રીક

“NumPy Numbers Need Neat Operations (NNNNO)”

પ્રશ્ન 5(b) [4 ગુણ]

પ્રોગ્રામમાં iris ડેટાસેટ Pandas Dataframe કેવી રીતે લોડ કરવો? ઉદાહરણ સાથે સમજાવો.

જવાબ

```

1 import pandas as pd
2
3 # પદ્ધતિ 1: CSV ફાઇલમાંથી લોડ કરવું
4 df = pd.read_csv('iris.csv')
5
6 # પદ્ધતિ 2: sklearn માંથી (ML માં સામાન્ય)
7 from sklearn.datasets import load_iris
8 iris = load_iris()
9 df_iris = pd.DataFrame(iris.data, columns=iris.feature_names)
10
11 # પ્રથમ 5 પંક્તિઓ બતાવો
12 print(df.head())

```

સમજૂતી:

- `pd.read_csv()`: CSV ફાઇલો વાંચવા માટેનું ફંક્શન.
- `df.head()`: પ્રથમ n પંક્તિઓ પાછી આપે છે (ડિફોલ્ટ 5).

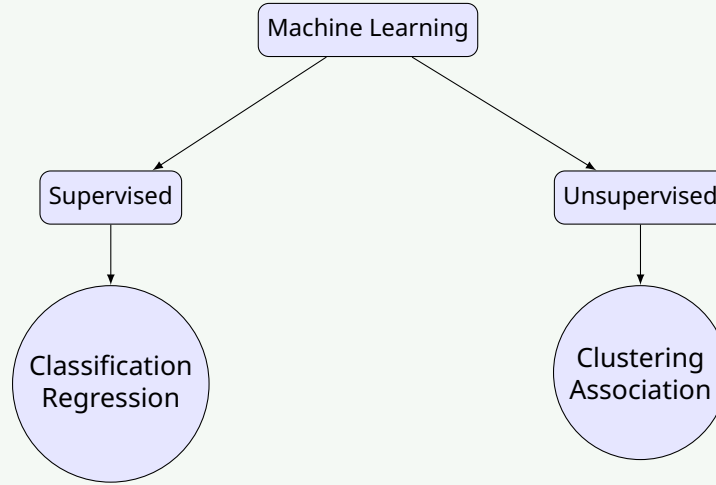
મેમરી ટ્રીક

“Pandas Reads CSV Files Easily (PRCFE)”

પ્રશ્ન 5(c) [7 ગુણ]

સુપરવાઇઝ્ડ લર્નિંગ અને અનસુપરવાઇઝ્ડ લર્નિંગની સરખામણી કરો અને કોન્ટ્રાસ્ટ કરો.

જવાબ



આકૃતિ 7. ML Learning Types

કોષ્ટક 17. તુલના

પાસું	Supervised	Unsupervised
ડેટા	લેબલ્ડ	અનલેબલ્ડ
લક્ષ્ય	આઉટપુટની આગાહી	પેટર્ન શોધવી
Feedback	સીધો પ્રતિસાદ	કોઈ પ્રતિસાદ નહીં
જટિલતા	વેલિડેશન સરળ છે	વેલિડેશન અઘરું છે

મેમરી ટ્રીક

“Supervised Seeks Specific Solutions, Unsupervised Uncovers Unknown (SSUU)”

પ્રશ્ન 5(a) OR [3 ગુણ]

Pandas ની એપ્લિકેશન્સની યાદી બનાવો.

જવાબ

કોષ્ટક 18. Pandas એપ્લિકેશન

એપ્લિકેશન	વર્ણન	ક્ષેત્ર
Data Cleaning	ખોવાયેલા ડેટાને સંભાળવું	General ML
Financial Analysis	શેરબજારના વલણો	Finance
Recommendation	વપરાશકર્તા વર્તન વિશ્લેષણ	E-commerce

મેમરી ટ્રીક

“Pandas Processes Data Perfectly (PPDP)”

પ્રશ્ન 5(b) OR [4 ગુણ]

Matplotlib લાઇબ્રેરીનો ઉપયોગ કરીને આકૃતિ કેવી રીતે બનાવવી? ઉદાહરણ સાથે સમજાવો.

જવાબ

```

1 import matplotlib.pyplot as plt
2
3 # સરળ લાઇન પ્લોટ
4 plt.plot([1, 2, 3], [1, 4, 9])
5
6 # x = 2 પર વર્ટિકલ લાઇન લાલ( રેડ)
7 plt.axvline(x=2, color='red', linestyle='--')
8
9 # y = 4 પર હોરિઝોન્ટલ લાઇન લીલી( ગ્રીન)
10 plt.axhline(y=4, color='green', linestyle='--')
11
12 plt.show()

```

ફંક્શનસ:

- axvline(x): અક્ષો પર ઊભી રેખા ઉમેરે છે.
- axhline(y): અક્ષો પર આડી રેખા ઉમેરે છે.

પ્રશ્ન 5(c) OR [7 ગુણ]

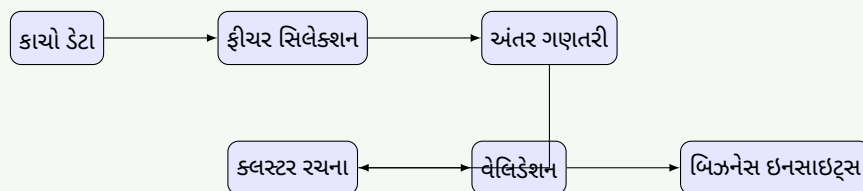
યોગ્ય વાસ્તવિક વિશ્વ ઉદાહરણોનો ઉપયોગ કરીને clustering ના concept નું વર્ણન કરો.

જવાબ

Clustering એ એક unsupervised learning તકનીક છે જે સમાન ડેટા પોઇન્ટ્સને જૂથબદ્ધ કરે છે જેથી એક જ જૂથમાંના બિંદુઓ અન્ય જૂથોના મુકાબલે એકબીજા સાથે વધુ સમાન હોય.

કોષ્ટક 19. Clustering એપ્લિકેશન

પ્રકાર	ઉદાહરણ	અસર
Customer Seg.	ખરીદી વર્તન દ્વારા જૂથ	Targeted માર્કેટિંગ
Image Seg.	MRI માં ગાંઠ શોધવી	સુધારેલ નિદાન
Gene Analysis	expression દ્વારા genes જૂથ	દવા શોધ



આકૃતિ 8. Clustering પ્રોસેસ

વાસ્તવિક ઉદાહરણો:

1. **Customer Segmentation:** ઉચ્ચ-મૂલ્યના ગ્રાહકો વિ. મોસમી ખરીદદારોને ઓળખવા.
2. **Social Media Analysis:** વપરાશકર્તાઓને રુચિઓ દ્વારા જૂથબદ્ધ કરવા (દા.ત. સ્પોર્ટ્સ, ટેક).
3. **Market Research:** સમાન ઉત્પાદન જરૂરિયાતો ધરાવતા સેગમેન્ટ્સ શોધવા.

મેમરી ટ્રીક

“Clustering Creates Clear Categories (CCCC)”