Chrome File Edit View History Bookmarks Profiles Tab Window Help 7:43

Search Google or type a URL

Update

Simplify Data-science Interviews

scaler.com

1 / 4

Chrome File Edit View History Bookmarks Profiles Tab Window Help 6:43

Agenda:

1. Roles
2. Skills for each role
3. Interview rounds
4. Sample Interview Questions
5. How to prepare & ace them
6. Projects
7. Non-tech & Non-cs backgrounds

2 / 2

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:47

Freshers (approx..)

12-15LPA (1) Data Analyst → BA; Op Analyst; PA; RA; FA ..... SCA

Roles

20-28L (2) Data scientists

20-28L (3) Machine learning Engineer → MLops ← Devops; SDE(ML); CV-Engineer

25-35L (4) Applied Scientist

— (5) Research Scientist

[Data Engineer → SDE + Data pipelines

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:46

Skills:

1. Python & libraries → Pandas; Numpy; Scipy; Matplotlib; Seaborn
   (Numpy: inbuilt; basic DS)
2. SQL
3. Data Analysis
   - ↳ Probability
   - → Applied stats
   - → Plotting

Data Analyst (must)

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:27

4: Dashboarding/Viz Tools — DA (good to have)
(not a must)

5: Math: Optimization;
Linear Algebra,
High-dim geom

6: Classical ML
↳ Supervised: GBDT; RF; ...
↳ Unsup: Clustering
→ RecSys; Time series

DS

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:32

7: Deep learning

↳ NN

→ CNN (CV)

→ NLP (RNN, Transformer)

8: MLops

→ API; Docker / kubernetes

→ Spark (how to use)

→ Aws / Azure / GCP (not must)

DS

→ ML System design

→ basics of NoSQL

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:40

MLE

↳ part SDE → DS & algo (some companies)

↳ part DS ✓

Applied Scientist

↳ DS + Cutting edge research

PhD → { Research Scientist → GB / DeepMind / FAIR / OpenAI / MSR

↳ Designs new algo from scratch

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:44

Interview rounds :

Data Analyst

1. SQL → subqueries ; Joins
2. Python → Libraries ; nested loops ; Simple recursion
3. Data analysis → scenario / data based

LeetCode (Medi some hard)

Data scientist

1. Programming → Python + SQL
2. Math → Prob, stats, optimization, LA ; Data Analysis ; high dim geom

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:41

3. breadth → ML & DL
4. depth → internals of techniques
5. Scenario based → MLops / CoreML/DL.....

MLE:

1. Programing — DSA }
2. SQL / programing
3. Breadth } scenario-based
4. depth }

9 / 9

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:53

ML breadth:

scenario

Train Data : $\mathcal{D}_{Tr}$

Test Data : $\mathcal{D}_{Te}$

Interview questions:

(Q) Determine if $\mathcal{D}_{Tr}$ is similar $\mathcal{D}_{Te}$

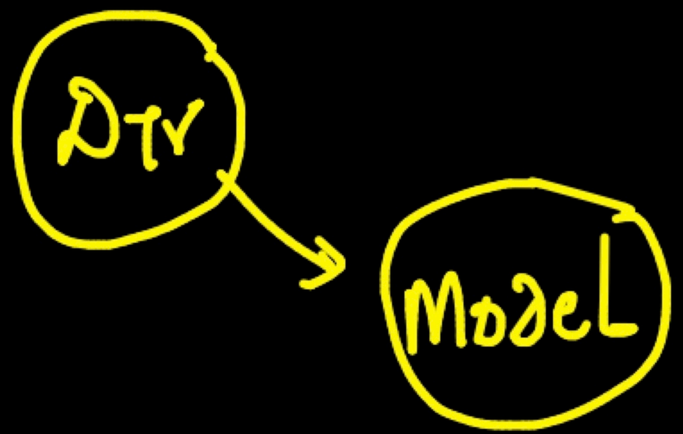
Classification

Chrome File Edit View History Bookmarks Profiles Tab Window Help 8:57

EDA :

range check y_i in D_{Tr} & D_{Te} → C1 & C2

↳ C1, C2



Tsain : F1 score Tr

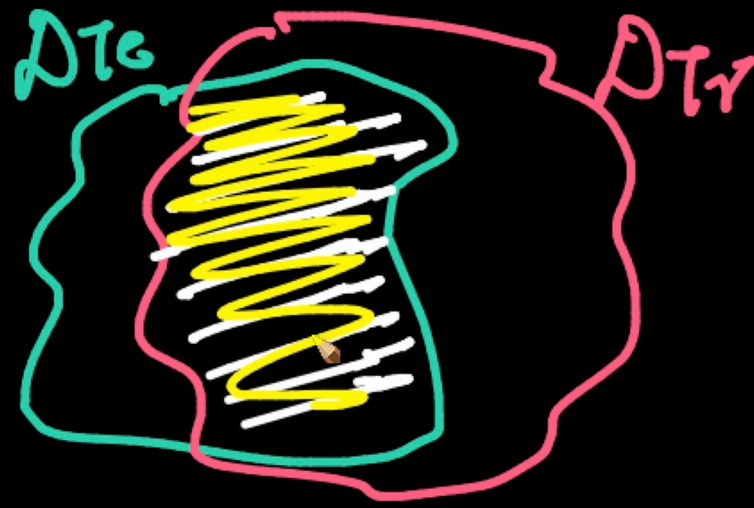
D_{Te} → Model → $F1_{Te}$

VI [if $F1_{Tr} \approx F1_{Te}$ — then D_{Tr} & D_{Te} are similar

else — dissimilar]

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:01

Follw-up:



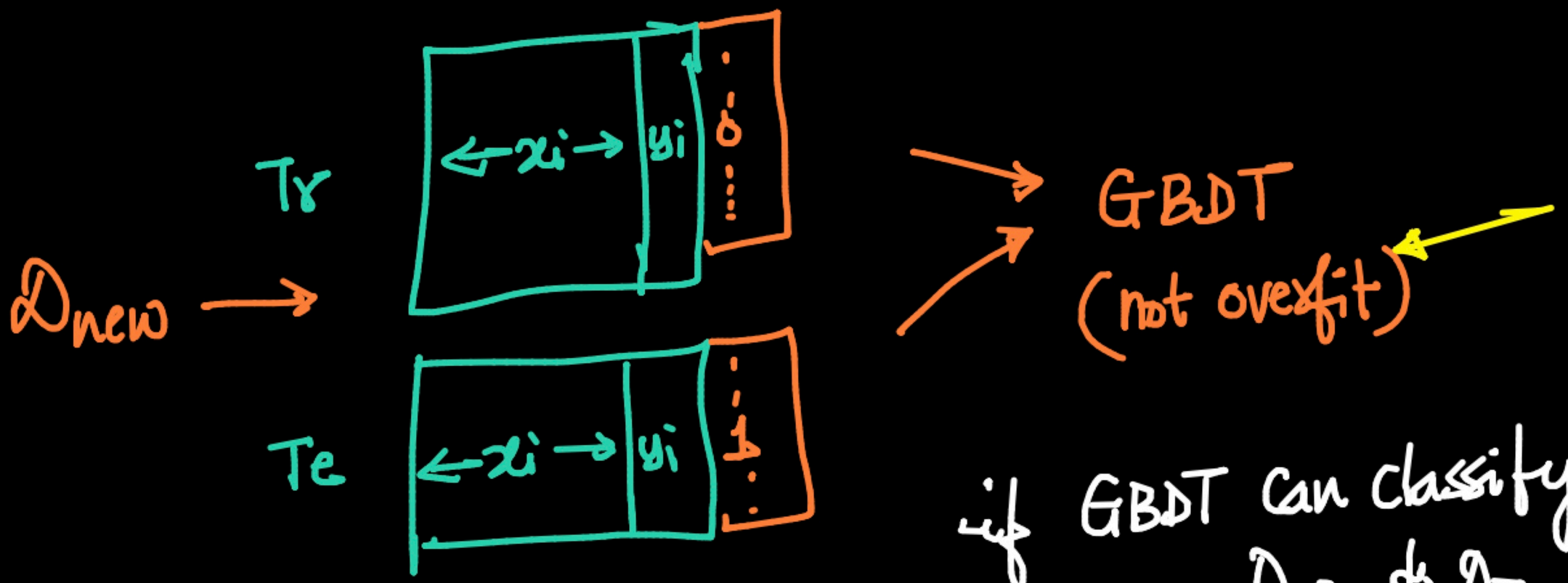
d-dim

$Fl_{Tr} \approx Fl_{Te}$

{ Is there a better technique to see if distb of D_{Tr} ≠ distb of D_{Te}

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:01

D or 1



Complex but generic

if GBDT can classify
then $D_{Tr} \not\perp D_{Te}$
else similar

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:10

Alt: (Stats) → ANOVA (variations)

Data Analyst

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:15
185: Solution with step by step
1st and Future - Player Contact
leetcode.com/problems/department-top-three-salaries/solutions/3214925/185-solution-with-step-by-step-explanation/
Update

Code

The given problem is to find the high earners in each department. In other words, we need to find employees in each department with the top three unique salaries.

To solve this problem, we can use the following query:

```
SELECT d.Name as Department, e.Name as Employee, e.Salary as Salary
FROM Department d, Employee e
WHERE (
    SELECT COUNT(DISTINCT Salary)
    FROM Employee
    WHERE Salary > e.Salary AND DepartmentId = d.Id
) < 3 AND e.DepartmentId = d.Id
ORDER BY d.Id, e.Salary DESC;
```

This query first performs a cross join between the Department and Employee tables to get all possible combinations of departments and employees.

Then, the WHERE clause filters out the employees who do not have a top three unique salary in their department. The subquery in the WHERE clause returns the count of distinct salaries greater than the employee's salary for the employee's department. If this count is less than 3, the employee's salary is in the top three unique salaries for the department.

Finally, the query selects the department name, employee name, and salary for the high earners and orders the result by department ID and employee salary in descending order.

Overall, this query efficiently finds the high earners in each department by utilizing a subquery and avoiding the use of expensive self-joins.

Related Solutions
SQL || DEPARTMENT TO...
MySQL
Beats 98% || EASY & SI...
MySQL
Using basic utilities of S...
MySQL 2+
185. Department Top T...
MySQL
Clean Code with Dense ...
MySQL
16 / 16

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:19

(Q2)

ML deploy

↳ SVM →

SV ; reduce # SV ...

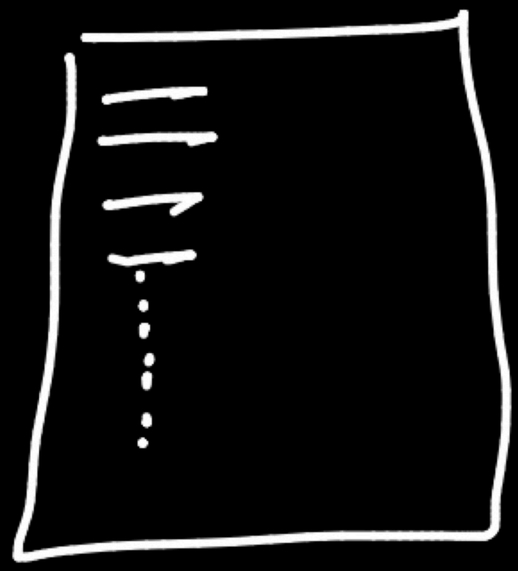
how do we make SVM run fast @ evaluation time

x_q → SVM → y_q (fast)

{ Random Forest → fewer & shallower trees
↳ GBDT preferred over RF ? (Math..)
↳ loss fn.

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:21

GBDT



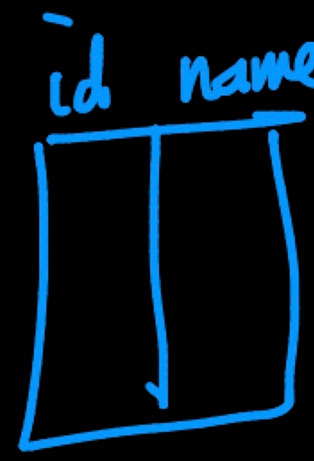
(1hr)

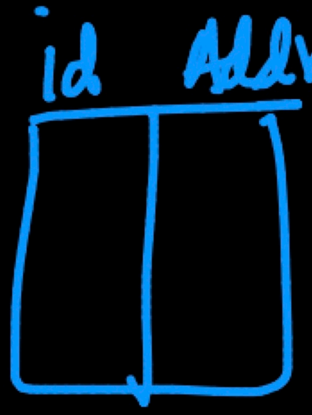
pseudo-code ..

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:23

(Q3) Python

DS | MLE





n

m

JOIN from "scratch"

implement library fn (Pandas- merge)

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:24

any library → pow(x,n) from Scratch

x^n ✓

↳ n is integer

→ n is real

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:27

(9) [Transformers]

↳ most time taking aspect of training

↓

how to Speed up this ?

GPT vs BERT

→ Speed up BERT →

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:33

Project → Data Analysis; ML; DL

Non-Tech

Showcase

2-3 solid

↳ Github (Code, ...) → end to end

→ Blog (2000-word)

→ 5 min video

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:57

Problems - LeetCode | Search | Kaggle

Search Google or type a URL

Update

23 / 23