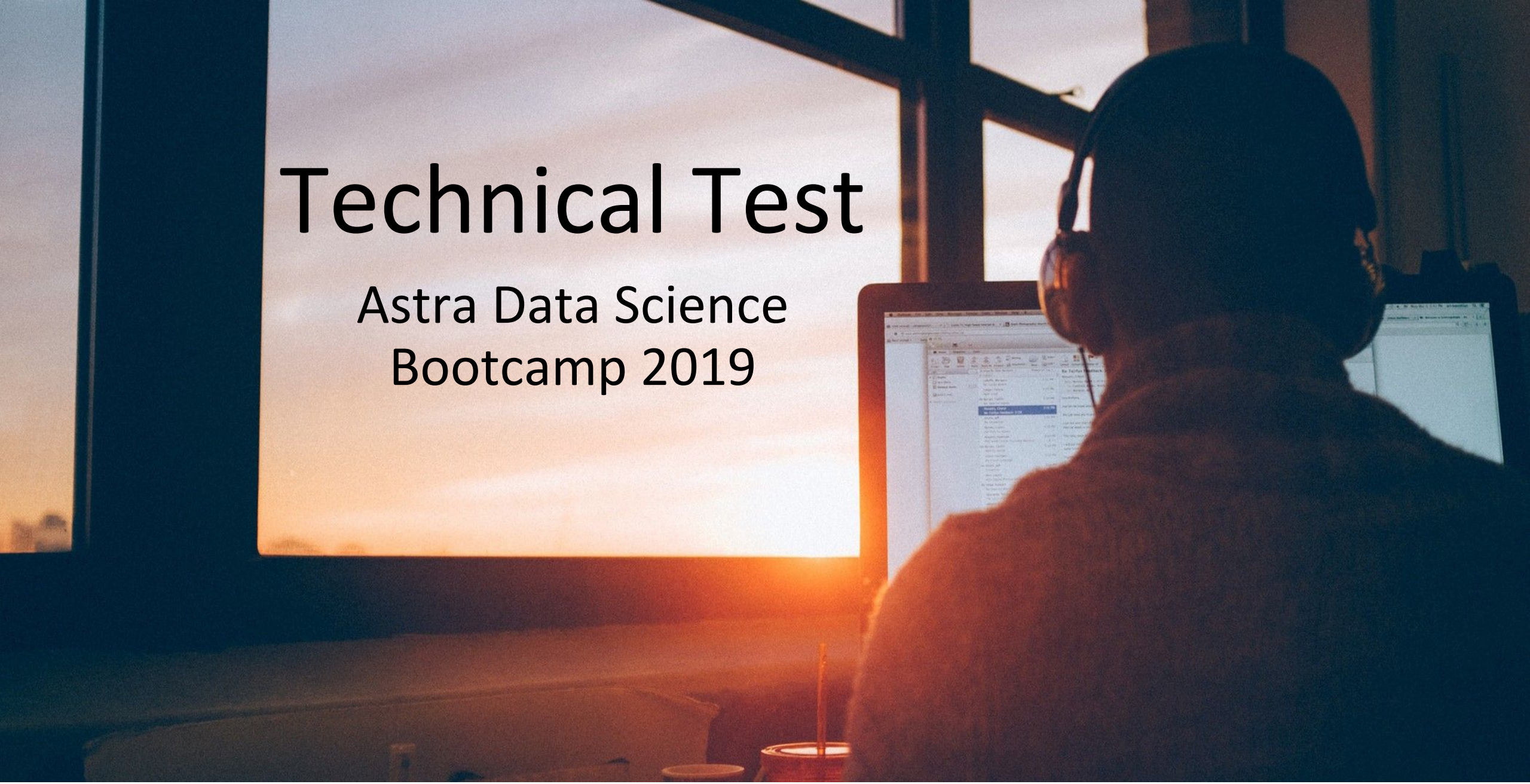


Technical Test

Astra Data Science Bootcamp 2019

AstraDigital

DATA·AI

Test Instruction

Test Overview

Test consists of 6 programming problems:

- 2 Logic problems
- 1 SQL problem
- 1 Statistical test problem
- 2 Basic machine learning problems

Test duration is 48 Hours

The expected output of each problem is a **function in R/Python** (SQL Script on SQL Problem) or at least **pseudocode** in english.

Each problem has different score, it will be appeared in problem title.

Test Problem

1. Valid Anagram (score: 15)

Given two strings *s* and *t* , write a function to determine if *t* is an anagram of *s*.

Example 1:
Input: *s* = "anagram", *t* = "nagaram"
Output: true

Example 2:
Input: *s* = "cat", *t* = "car"
Output: false

Note:
You may assume the string contains only lowercase alphabets.

Follow up:
What if the inputs contain unicode characters? How would you adapt your solution to such case?

2. Confused Driver (score: 25)

Truck driver carries secret package through desert.

He was instructed to go to **a checkpoint**
then **turn left** as long as **half** of the first journey
(from start to the checkpoint)

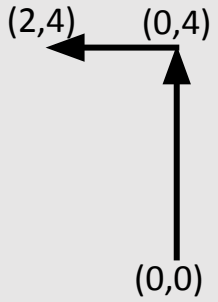
Help the driver to get the coordinate of the endpoint directly

Write a function
input : [(x1,y1), (x2,y,2)]
output : (x3,y3)

(x1,y1) = start coordinate
(x2,y2) = checkpoint coordinate
(x3,y3) = endpoint coordinate

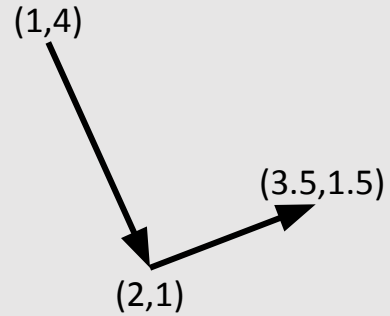
Example 1:

input : [(0,0), (0,4)]
output : (-2,4)



Example 2:

input : [(1,4), (2,1)]
output : (3.5,1.5)



AstraDigital

DATA·AI

3. Top Three Salary on each Dept (score: 20)

Id	Name	Salary	DepartmentId
1	Joe	85000	1
2	Henry	80000	2
3	Sam	60000	2
4	Max	90000	1
5	Janet	69000	1
6	Randy	85000	1
7	Will	70000	1

Employee Table

Id	Name
1	IT
2	Sales

Department Table

The Employee table holds all employees. Every employee has an Id, and there is also a column for the department Id.

The Department table holds all departments of the company.

Write a SQL query to find employees who earn the top three salaries in each of the department. For the above tables, your SQL query should return the following rows (order of rows does not matter).

Department	Employee	Salary
IT	Max	90000
IT	Randy	85000
IT	Joe	85000
Sales	Henry	80000
Sales	Sam	60000

4. Two sample t-test for Equal Means (score: 15)

Given two array of numbers which represent sample of two sample. Write a function to determine if two population means are equal.

input : Two array of numbers

output: Test Statistic & decision whether two population means are equal

5. Entropy in decision tree (Score : 10)

A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogeneous). One of algorithms for partitioning data is using **ID3**. **ID3** using entropy for calculate homogeneity of sample.

Create a function to calculate **entropy** of weather and play football from the following table.

Notes:
Input: [[‘rainy’,’no’], [‘rainy’,’no’], [‘overcast,’yes’], …]
Calculate: E(Weather,Play Football)
Output: Entropy score

Weather	Play Football
Rainy	No
Rainy	No
Overcast	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Overcast	No
Overcast	Yes
Sunny	No

6. Gradient Descent (Score : 15)

Create a function to compute local minimum and number of gradient descents steps from function $f(x) = X^5 - 3X^3 - 2X^2 + 10$ with initial X=0.1;