Candidate Screening Test

Sr. Algorithm Engineer

Screening content

- Linux/Bash
- C/C++
- Python
- ML Questionnaire

Linux/Bash

1. Build Bash script to extract system info and output into text file,
   - CPU number
   - Available hard drive space
   - Current memory usage
2. Build Bash script to do following processes and output summary into text file
   - Count how many lines in your resume file
   - Extract 2nd word of each line of your resume file
   - Extract first and last line of your resume file

C/C++

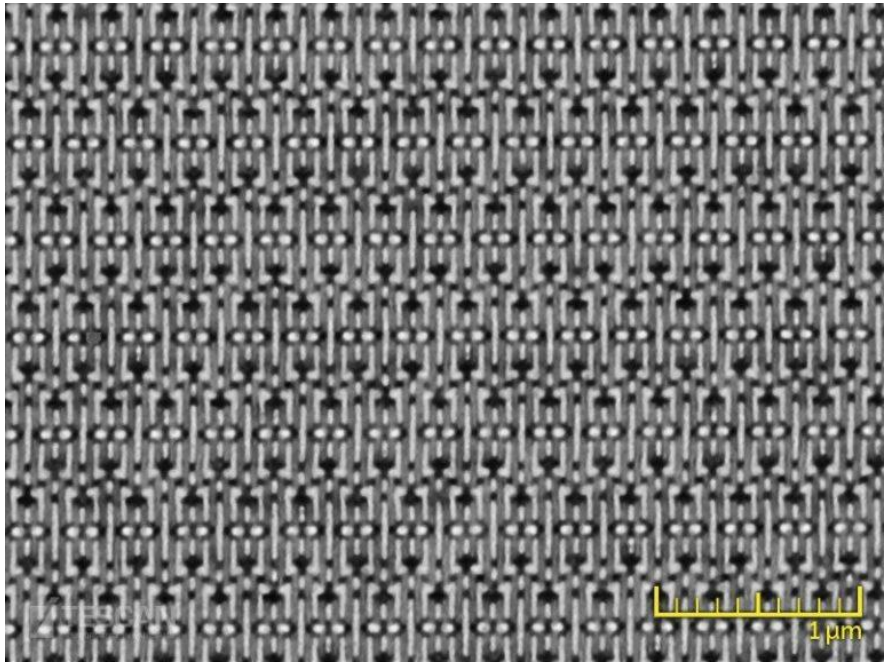
1. Write C/C++ program to,
   - Generate numerical array of 0-9, and print it on screen
   - Then print it in the format of

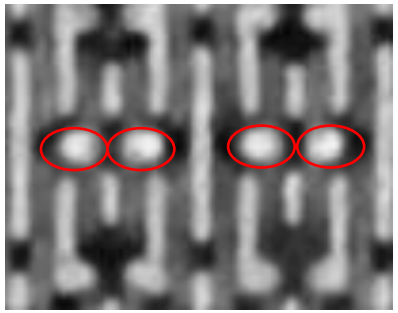
```
   0
  1  2
 3  4  5
6  7  8  9
```

   - Log CPU and Elapse time
   - Log Memory usage
2. Profile your C/C++ program and output summary file to show
   - Which function is called the most, for how many times
   - Which function costs the most, for how much time

Python

1. Build Python script to process image
   - Extract X, Y direction pixel number of the image
   - Extract pixel size (in nm) of the image
   - Plot grayscale histogram
   - Plot X direction average grayscale
   - Plot Y direction average grayscale
   - Extract all structures in the image like red circled on and create a new image
   - Generate one new image which only contains the structure like red circled and whose grayscale is significantly lower





ML Questionnaire

1. Please share your understanding of supervised learning, semi-supervised learning, and unsupervised learning, as well as their application scenarios.
2. Discuss your understanding of deep learning and how it differs from traditional machine learning methods.
3. Please explain the problem of "overfitting" in machine learning, and how you avoid or address it. Please explain from the perspectives of both traditional machine learning and deep learning.
4. How do you evaluate the performance of a machine learning model? Please explain some commonly used performance metrics.
5. How do you handle imbalanced datasets? Would the approach differ when the imbalance ratio is 1:10 versus 1:10000?