

## Technical Interview Questions:

Please try to answer some of the following questions to the best of your ability while explaining all of your logic. It is okay if you do not know the answer to these questions, feel free to email Joshua for hints if you are stuck. If you are unable to answer the question, please explain how you would go about solving the question. Programming using any developer environment (e.g., text editor) and/or Microsoft Word is sufficient to answer these questions.

### 1. Programming/Statistics:

#### a. Functional programming:

Write a python/R function that returns the mean, standard deviation and number of elements of a python/R list/vector. As an example, you can utilize this python list: [5.99342831, 4.7234714, 6.29537708, 8.04605971, 4.53169325, 4.53172609, 8.15842563, 6.53486946, 4.06105123, 6.08512009].

#### b. Statistics:

Let's imagine that these numbers represent the expression of a specific gene and we want to see whether this gene is highly expressed. If the genes are highly expressed, then the patient may be at risk for developing cancer. Given the list above, calculate the proportion of values which surpass 4, which is our hypothetical threshold of what constitutes "normal" expression. This will give us an indication of how many patients to follow-up with additional screening.

#### c. Object Oriented Programming:

What is a python class? How is this different from a function? Can you list a few examples of classes? What are attributes and methods and how do they differ? Can you give an example of a class? Optional: As an added challenge, see if you can turn the above calculation of proportion of genes that are highly expressed into a Python class that accepts as input such a list of a gene's expression across individuals.

#### d. Plotting:

Can you create a histogram of the following expression values? Comment on the shape of the histogram.

[https://matplotlib.org/3.5.0/api/as\\_gen/matplotlib.pyplot.hist.html](https://matplotlib.org/3.5.0/api/as_gen/matplotlib.pyplot.hist.html)

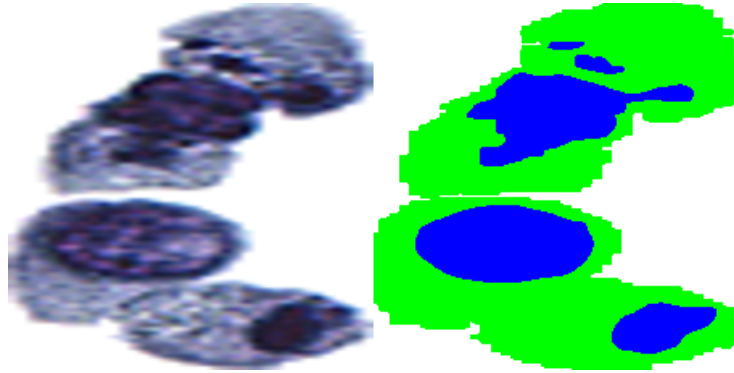
```
[19.47,13.76,20.83,28.71,12.89,12.89,29.21,21.91,10.77,19.88,10.83,10.81,17.18,-2.22,-0.52,9.94,5.88,17.83,6.83,2.29,28.19,12.97,15.61,2.18,10.1,16.0,4.64,18.38,9.59,12.37,9.58,31.67,14.88,5.48,22.4,4.01,16.88,-2.64,3.05,16.77,21.65,16.54,13.96,12.29,1.69,8.52,10.85,24.51,18.09,-0.87,17.92,11.53,8.91,20.51,24.28,23.38,7.45,12.22,17.98,23.78,10.69,13.33,5.04,4.23,22.31,27.21,14.35,24.03,18.25,9.19,18.25,28.84,14.68,29.08,-8.58,22.4,15.78,12.31,15.83,-2.89,13.02,18.21,28.3,10.34,7.72,10.48,23.24,17.96,10.23,19.62,15.87,23.72,8.68,12.05,11.47,1.83,17.67,17.35,15.05,12.89]
```

### 2. Bonus Questions (Going above and beyond, no need to do these; maybe select one if you are interested):

#### a. Machine learning:

What is machine learning? How does machine learning differ from other algorithms? What is the difference between supervised and unsupervised learning? Can you give examples of these approaches?

b. Image Analysis:



On the left hand side of the above image are urine cells. On the right the cells have been divided into their nuclei (blue) and cytoplasm (green). The nucleus to cytoplasm ratio, defined as  $\text{blue}/(\text{blue}+\text{green})$ , is a measure of potential malignancy (is the cell cancerous?). Normally, a pathologist would outline where the blue and green colors should be added to the diagram. Pathologists usually give an “eyeball” estimate of this ratio that can be unreliable. We want to develop a computer program that could emulate the pathologist’s markup of the image. How would you go about doing this?

c. Deep learning / Tensorflow for Image Classification:

On this website, <https://playground.tensorflow.org/> , you can fit machine learning model that learns how to classify datapoints in shape like patterns (e.g., orange is separate from blue, algorithm should also learn to separate them). Neural networks can learn to separate these two colors by mimicking processes of the central nervous system. Try playing around with the system to see if you can classify data in different shapes. What worked for you? What did not work? Try building the smallest neural network possible to classify the shapes.

d. Natural Language Processing: Check out this autocomplete text tool to see how neural networks can auto-complete your text!

<https://transformer.huggingface.co/doc/distil-gpt2> Did you find anything interesting? What are your thoughts on how to represent language and text using computer programs/algorithms?

Reference to Some of Our Previous Works:

1. <https://www.nature.com/articles/s41379-020-00718-1>
2. <https://www.nature.com/articles/s41379-020-0526-z>
3. [https://psb.stanford.edu/psb-online/proceedings/psb21/levy\\_j.pdf](https://psb.stanford.edu/psb-online/proceedings/psb21/levy_j.pdf)
4. <https://acsjournals.onlinelibrary.wiley.com/doi/full/10.1002/cncy.22099>
5. <https://pubmed.ncbi.nlm.nih.gov/31797614/>
6. <https://www.biorxiv.org/content/10.1101/2020.01.07.897801v2>
7. <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-3443-8>
8. <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-01046-3>