Mastercard AI Garage Interview Experiences

# **Slot: 1**

# **Procedure**

1. Test: Yes. (Online)
2. Interview Mode: (CV Raman Building[white building near Hoysala guesthouse])

----------------------------------- Physical ----------------------------------------------------------------------------

# Sofia Sunam

**Personal View**: [will update test experience later]. 2 rounds of Technial Interview and 1 HR round. Asked from CV/resume. Only important conclusions from lectures of MLDS and INLP from last sem will help to clear round 1. Round 1 and 2 were of the same difficulty [moderate]. Round 2 had coding for others but I said that I know coding so as to do assignments and also have SDE2 experience. Thus, i guess, they did not ask me for DSA coding. HR round was too generic where they confirmed on location and compensation.

1. Keep CV/resume print out ready ; ample number of them

2. Take water bottle umbrella and food

3. Prepare your resume well

**Status**: Attended till final round (HR) and did not get selected.

**Interview Description**

## **Round 1:**

Educational Background of the Interviewer (only one member): idk

Project Specific Discussions: Almost none.

1. What makes transformer stochastic (temperature var)
2. Explain rag
3. Projects on classifier
   * Input n model working n objective
   * Why random forest over knn : Bias variance trade off
   * Is knn parametric
4. Is the graph acyclic
   * Dfs algo time complexity

**Round 2:**

* Different interviewer (onemember), with a long history in this domain.
  + Why did we chng from rnn to transformer
  + N customers, rsD discount and k% available the discount. So total discount amt =nkd

**Round 3: HR**

* HR round was too generic where they confirmed on location and compensation.
* How do i destress myself
* Short intro and story to IISc
* Previous employer details and experience
* Bangalore experience and what challenges on first time coming out of state

---------------------------------------------------------------------------------------------------------------

# Tejas Tonde

**Personal View**: [will update in detail later] Interviews were moderate-level difficult. Round 2 interviewer focused more on the loss functions used in the projects from the resume. Keep at least 2 copies of the resume.

**Status**: Attended till Round 2. Not selected.

**Interview Description**

## **Round 1:**

Educational Background of the Interviewer: No idea

Project Specific Discussions: CNN,RNN, Transformers , Adam optimizer

* Introduce yourself
* **Adam**
* Started with one of my project related to Adam
* Asked me to write the formula for Adam. Why was Adam introduced in first place. Compare SGD(Stochastic Gradient Descent), SGD with momentum and Adam. What is bias correction term?
* Explain the entire process of how the project was done, experiments and conclusions of the project.
* **CNN and RNN**
* Explain how CNNs work. Given input image, stride,padding, filter size what is the output feature map.
* If you use a 3x3 kernel instead of a 7x7 kernel, what changes will occur? Pros and cons of each case.
* Do we encounter the problem of vanishing gradient in case of CNNs? Why is vanishing gradient more relevant in RNNs than CNNs? Explain vanishing gradient and ways to overcome it?
* Which layers in CNN will get a drastic update in the parameters? (Nothing about the layer was specified so I tried to answer with respect to the position of the layer from the output layer and size of kernel)
* Why is it preferred to use a ReLU activation instead of sigmoid or tanh? (Think in terms of the output range of these functions. Sigmoid and tanh are bounded, ReLU isn’t. Can there be any effect like vanishing gradient?)
* **Transformer**
* Have you worked on transformers? Explain different types of positional encoding in the transformer.
* Explain sinusoidal encoding in detail. Why specifically sin and cos, why not any other function? How does it ensure unique encoding for each token? In a m X n matrix of input embeddings (m = sequence length and n = embedding dimension), is it embedding applied across m or n? Explain with formula.
* **Ungrouped questions**
* Explain the difference between and auto-encoder and a variational auto encoder. What is the difference in their loss functions?
* Explain hierarchal softmax.
* **Puzzle**
* Puzzle: You have a jar of 4 litres and a jar of 9 litres. Having unlimited supply of water, how will you measure 1,2,3,4,5,6,7,8,9 litres?

**Round 2:**

1. Introduce yourself
2. Write binary cross entropy, categorical cross entropy. Write loss function for multi-class and multi-label classification.
3. Picked one project and asked to write the loss functions (dice loss and IoUloss(Intersection over Union). Which one is better and why? Are these losses differentiable?
4. How will you use VAE or an Auto-Encoder for binary classification task? Difference between VAE and Auto-encoder? Explain their loss functions.
5. How is the distance between the distributions measured?
6. If the distance is measured using KL divergence, why is it called KL divergence rather than KL distance? Write the formula for KL divergence.

---------------------------------------------------------------------------------------------------------------