Amazon's Data Scientist Interview Practice Problems

A walkthrough of some of Amazon's interview questions!

Terence S
Feb 26 · 6 min read ★

Photo by Christian Wiediger on Unsplash

Given the popularity of my articles, Google's Data Science Interview Brain Teasers, 40 Statistics Interview Problems and Answers for Data Scientists, Microsoft Data Science Interview Questions and Answers, and 5 Common SQL Interview Problems for Data

Scientists, this time I collected a number of Amazon’s data science interview questions on the web and answered them to the best of my ability. Enjoy!

If this is the kind of stuff that you like, be one of the FIRST to subscribe to my new YouTube channel here! While there aren’t any videos yet, I’ll be sharing lots of amazing content like this but in video form. Thanks for your support :)

. . .

Q: If there are 8 marbles of equal weight and 1 marble that weighs a little bit more (for a total of 9 marbles), how many weighings are required to determine which marble is the heaviest?

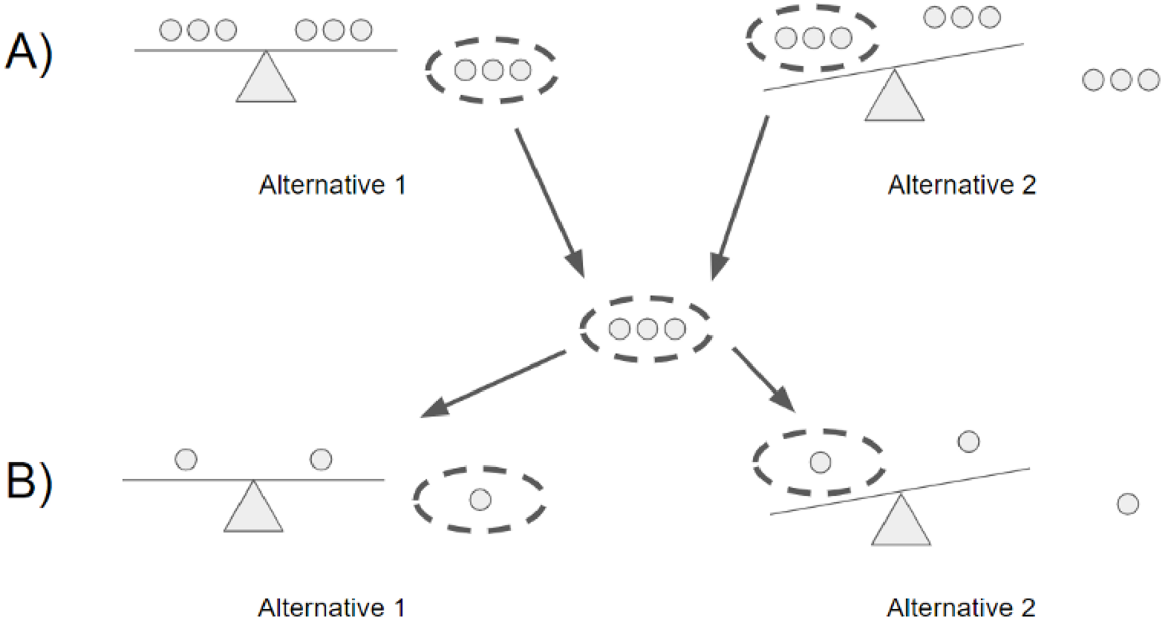


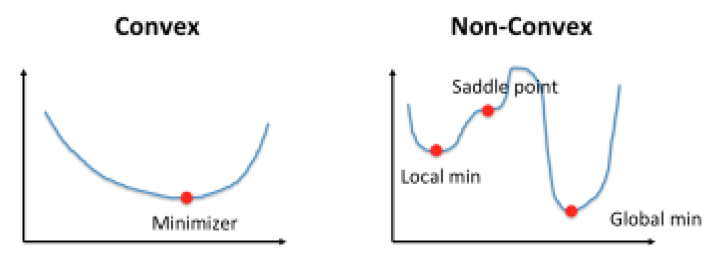
Image created by author

Two weighings would be required (see part A and B above):

1. You would split the nine marbles into three groups of three and weigh two of the groups. If the scale balances (alternative 1), you know that the heavy marble is in the third group of marbles. Otherwise, you’ll take the group that is weighed more heavily (alternative 2).

2. Then you would exercise the same step, but you’d have three groups of one marble instead of three groups of three.

Q: Difference between convex and non-convex cost function; what does it mean when a cost function is non-convex?



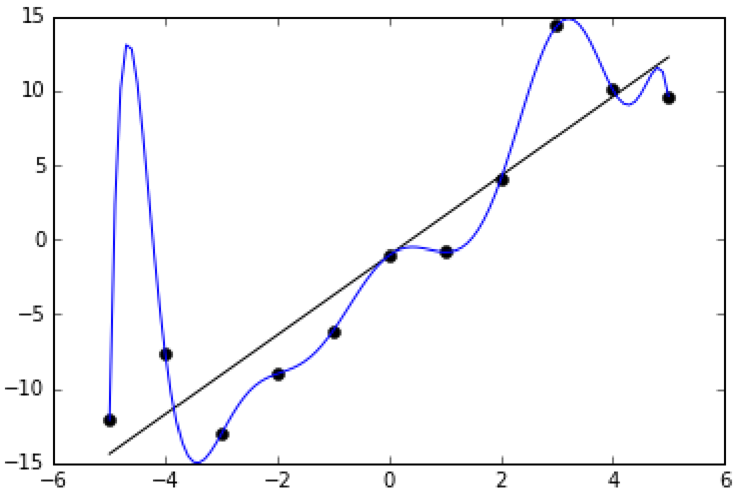
Taken from Cho-Jui Hsieh, UCLA

A **convex function** is one where a line drawn between any two points on the graph lies on or above the graph. It has one minimum.

A **non-convex function** is one where a line drawn between any two points on the graph may intersect other points on the graph. It characterized as "wavy".

When a cost function is non-convex, it means that there's a likelihood that the function may find local minima instead of the global minimum, which is typically undesired in machine learning models from an optimization perspective.

Q: What is overfitting?



Taken from Wikipedia

Overfitting is an error where the model 'fits' the data too well, resulting in a model with high variance and low bias. As a consequence, an overfit model will inaccurately predict new data points even though it has a high accuracy on the training data.

Q: How would the change of prime membership fee affect the market?

I'm not 100% sure about the answer to this question but will give my best shot!

Let's take the instance where there's an increase in the prime membership fee — there are two parties involved, the buyers and the sellers.

For the buyers, the impact of an increase in a prime membership fee ultimately depends on the price elasticity of demand for the buyers. If the price elasticity is high, then a given increase in price will result in a large drop in demand and vice versa. Buyers that continue to purchase a membership fee are likely Amazon's most loyal and active customers — they are also likely to place a higher emphasis on products with prime.

Sellers will take a hit, as there is now a higher cost of purchasing Amazon's basket of products. That being said, some products will take a harder hit while others may not be impacted. It is likely that premium products that Amazon's most loyal customers purchase would not be affected as much, like electronics.

Q: Describe Tree, SVM and Random forest. Talk about their advantage and disadvantages.

Decision Trees: a tree-like model used to model decisions based on one or more conditions.

- Pros: easy to implement, intuitive, handles missing values
- Cons: high variance, inaccurate

Support Vector Machines: a classification technique that finds a **hyperplane** or a boundary between the two classes of data that maximizes the margin between the two classes. There are many planes that can separate the two classes, but only one plane can maximize the margin or distance between the classes.

- Pros: accurate in high dimensionality
- Cons: prone to over-fitting, does not directly provide probability estimates

Random Forests: an ensemble learning technique that builds off of decision trees. Random forests involve creating multiple decision trees using bootstrapped datasets of the original data and randomly selecting a subset of variables at each step of the decision tree. The model then selects the mode of all of the predictions of each decision tree.

- Pros: can achieve higher accuracy, handle missing values, feature scaling not required, can determine feature importance.
- Cons: black box, computationally intensive

Q: Why is dimension reduction important?

Dimensionality reduction is the process of reducing the number of features in a dataset. This is important mainly in the case when you want to reduce variance in your model (overfitting).

Wikipedia states four advantages of dimensionality reduction (see here):

1. *It reduces the time and storage space required*
2. *Removal of multi-collinearity improves the interpretation of the parameters of the machine learning model*
3. *It becomes easier to visualize the data when reduced to very low dimensions such as 2D or 3D*
4. *It avoids the curse of dimensionality*

Q: The probability that item an item at location A is 0.6, and 0.8 at location B. What is the probability that item would be found on Amazon website?

We need to make some assumptions about this question before we can answer it. **Let’s assume that there are two possible places to purchase a particular item on Amazon and the probability of finding it at location A is 0.6 and B is 0.8. The probability of finding the item on Amazon can be explained as so**:

We can reword the above as P(A) = 0.6 and P(B) = 0.8. Furthermore, let’s assume that these are independent events, meaning that the probability of one event is not impacted by the other. We can then use the formula…

P(A or B) = P(A) + P(B) — P(A and B)
P(A or B) = 0.6 + 0.8 - (0.6*0.8)
P(A or B) = 0.92

Q: Describe SVM.

Answer already provided in previous question

Q: What is boosting?

Boosting is an ensemble method to improve a model by reducing its bias and variance, ultimately converting weak learners to strong learners. The general idea is to train a weak learner and sequentially iterate and improve the model by learning from the previous learner. *You can learn more about it here*.

Thanks for Reading!

If you like my work and want to support me…

1. The BEST way to support me is by following me on **Medium** here.
2. Be one of the FIRST to follow me on **Twitter** here. *I’ll be posting lots of updates and interesting stuff here!*
3. Also, be one of the FIRST to subscribe to my new **YouTube channel** here!
4. Follow me on **LinkedIn** here.
5. Sign up on my **email list** here.
6. Check out my website, **terenceshin.com**.

Resources

Amazon Data Scientist Interview Questions

62 Amazon Data Scientist interview questions and 61 interview reviews. Free interview details posted anonymously by…

www.glassdoor.ca

Amazon Data Science Interview

Amazon is hiring more developers for Alexa than Google for everything.

medium.com

Overfitting

In statistics, overfitting is "the production of an analysis that corresponds too closely or exactly to a particular…

en.wikipedia.org

http://web.cs.ucla.edu/~chohsieh/teaching/CS260_Winter2019/lecture3.pdf

Data Science Machine Learning Artificial Intelligence Amazon Work