Machine Learning Design

Duration: 20 Minutes
Difficulty: Medium
Domains: Product

Problem

How would you calculate the customer lifetime value (CLV) of Amazon customers?

Solution

[Candidate] Before talking about the solution, I'd like to first frame the problem. Based on my experience, I believe customer lifetime value is defined by the following:

1. Lifetime history (i.e. 6 months, 9 months, 3 years and e.t.c.)
2. Prediction point (i.e. CLV prediction based on the first week of sale or first month of sale)
3. Customer segment (i.e. all Amazon customers or subset?)

Are my assumptions valid?

[Interviewer] They are all reasonable assumptions about CLV. Let's discuss them more in details one by one. What do you believe would be a reasonable lifetime history?

[Candidate] Well, I believe this depends on the data history available. For instance, if I have three months worth of sales data per customer, I won't be able to build a CLV model with greater than 3 months of forecast horizon. Also, it's not possible to build a "lifetime" value of a customer given churn and limited data history. So, it needs to be confined to some months or years. To start, perhaps I can design a system that predicts the CLV of 1 year.

[Interviewer] Sounds good. Now, let's talk about how you would predict 1 year.

[Candidate] To calculate the 1-year CLV of a customer, I need their transaction history. But, it needs to be designed based on two different approaches.

Design 1 - Predict 1 year based on the first X weeks/months/years of the customer lifetime value. An approach could be: predict 1 year given the customers' first month of purchasing goods on Amazon.

Design 2 - Moving window approach. Regardless of whatever stage a customer is on, continue to predict the next 12 months of sales given the past X weeks/months/years sales patterns. For instance, predict sales of the next 12 months given the past 3 months of spending patterns. Update the model on a monthly basis.

[Interviewer] Now, let's talk about variables. What do you believe are relevant signals for predicting the CLV of customers? Let's say you have access to the customer's purchase pattern data.

[Candidate] I think I could come up with a couple. I could apply apply a window-based functions aggregated at the customer level to derive the following features:

1. Average sales the past 1 days

2. Average sales the past 7 days
3. Average sales the past 30 days
4. Average sales the past 60 days
5. Average sales the past 120 days
6. Count sales the past 1 days
7. Count sales the past 7 days
8. Count sales the past 30 days
9. Count sales the past 60 days
10. Count sales the past 120 days
11. # of days since first purchase
12. # of logins this week

I could also apply the same strategy aggregated at the customer x department level to extract additional features. For instance, the average sales in grocery the past 7 days. Using this strategy, I could literally create hundreds of signals.

[Interviewer] Great. Now, one quick question on that. It seems that the signals could be highly correlated, how would you address that?

[Candidate] There are various statistical techniques to handle multicollinearity. One of them is using the Lasso model to remove collinearity in the feature set.

[Interviewer] Sounds good. Now, what's your next step toward designing the solution?

[Candidate] With the features select, I can start with a simple benchmark model such as random forest with hyper-parameter tuning on N trees, depths, # of samples per terminal node and e.t.c.

[Interviewer] If you were to perform hyper-parameter tuning, how much should it improve your model performance?

[Candidate] I'd say it's reasonable to gain 1 to 5% improvement using hyper-paramter tuning. It's quite rare for the improvement to happen beyond that unless the default parameter setting was overfitting the data to begin with.

[Interviewer] Great, one quick question on the granularity. Would you want to forecast the CLV of 1 year in a single data-point of prediction or 12 months?

[Candidate] I suppose that depends on what the business stakeholders want. Sometimes, single data-point suffice. However, in many cases, the business wants more granularity. So, forecasting the 12 month in a month-to-month granularity could work as well.

[Interviewer] Great. Now, let's talk about productionization. How would you deploy your model?

[Candidate] If the offline model performance is decent (let's say that MAPE is <10%), then I can deploy a model that predicts real-time. Meaning for every customer that has just completed the 1 month as a customer customer, predict their CLV for the next 12 months. I would deploy monitoring such as data validations, unit-testing and performance tracking to ensure that the model is performing quite well.

Interviewer Solution

A naive solution would say, "build a regression model that trains on customer behavior data and predicts customer lifetime value. You can use a model like XGBoost, perform hyper-parameter tuning and optimize on mean squared error."

Although the explanation above describes the model, many vital details are neglected. First you need two elements - time horizon and prediction point.

(1) Define Time Horizon

The CLV is the total revenue a business generates from a customer over time horizon. CLV is not always a measure of revenue across a customer's lifetime, from the moment of sign-up until cancellation. CLV is also measured over a period. For instance, a business could be interested in the lifetime value of the first month, six months or 12 months from a customer since the sign-up.

(2) Define Prediction Point

At what point do you predict the lifetime value of a customer? Are you predicting on a new customer? Or, are you predicting on a current customer based on the purchase history?

Depending on your design decisions, the data size, feature availability and algorithm would differ.

You could explain the following to interviewer:

You: "To measure customer lifetime value, I need to define the time horizon and prediction point. For instance, I could predict the total lifetime value of months three through six given a customer's purchase history on the first two months. Or, I could predict lifetime value of the first six months since sign-up."

Interviewer: "Very good. How would you decide?"

This follow-up question is an opportunity to showcase your business sense. You could explain the following:

You: “In either one of the cases, you could project the average or total revenue earned across customers. The forecasting could be useful for reporting purpose to address questions like, ‘is the business alluring customers that will yield high revenue?’

One additional note about the second approach, prediction at sign-up, is that this model could be useful for A/B testing. Suppose an experimentation is conducted comparing sign-up rates on two variations. One question to address is the revenue earned from each of the two groups. You could wait six months to gather total revenues from the customers or, forecast the expected revenue earned at conversion.”

Interviewer: “Your application on A/B testing certainly makes sense. Now, how would you predict customer lifetime value given that you are predicting at sign-up?”

You’ve established the prediction point and lifetime value. In addition, you selected and justified an approach. Now, you can discuss modeling part.

You: “Given that I do not have user purchase pattern. I could use other data source such as device information, location data, time period, ip address, and cookie history.”

Interviwer: “Can you elaborate?”

Prediction without user history is a cold-start problem in machine learning. Many creative approaches are proposed. One approach is to gather as much information about the user’s device through a backend or third-party service that collects the information every time a user enters the site. This is called device intelligence and, oftentimes, companies use this to resolve cold-start problems in recommender system or fraud problems.

For instance, device intelligence can collect information on the user’s browser type, operating system, latitude and longitude of location, device type and such. There are start-ups that use this technology to collect hundreds of signals about a new user even before he clicks anything on the site.

You: “I presume you can collect device information from browser type (i.e. Chrome, Safari) device type (i.e. mobile, desktop or tablet), country location (i.e. U.S., Japan) and such. Additionally, I assume you would know the source of how the new customer stumbled upon the page as in, ‘did the new customer enter the site through search engine like Google or Facebook Ads?’ Lastly, I would also presume that you would know some information about the user based on their profile information populated during the sign-up.”

> Interviewer: “That is correct. Now, I’d like to see you design the model.”
>
> Once you laid the grouandwork of the problem and variables, the approach is the problem is straightforward. You cn explain how you would use the raw variable to derive model features through feature engineering and selection. Lastly, you could explain what model you would use. As a starting point, do not be concerned about the sophistication of your algorithm. Interviwers generally do not care whether you would decide to use neural networks or logistic regression. The focus point is on your design. Lastly, you can evaluate your model on regression evaluation metrics like mean squared error, mean absolute error, mean absolute percentage error and such.
>
> In each step explained, expect one or two follow-up questions. As long as you explain and justify your approach, you will ace this machine learning design interview.

Interviewer Assessment

In the statistics section, a candidate is assessed based on correctness and soundness of statistical methodology, product sense and communication. For each dimension the candidate is rated in the following scale: (5) superior, (4) good, (3) adequate, (2) marginal, (1) not competent.

Assessments	Rating	Comments
ML Methodology	5	The candidate's approach to the CLV problem was exceptional. She seems to have experience in solving such problem in the past experience. She understands the parameters that define the problem. She also has a strong grasp in ML fundamentals as she provided sound approaches to feature engineering and selection. As a bonus, she also understands how to productionize a model.
Product Sense	5	The candidate demonstrated a strong understanding of Amazon customer patterns and data. She leveraged her understanding of the Amazon e-commerce platform to address the ML problem.
Communication	5	The candidate asked great questions, clarifying assumptions required to solve her problem. She also explained her answers clearly and comprehensively. As a plus, she understands that some of the business problem scope is based on the business stakeholder needs. This is a key quality assessed, and she has shown that she understands how to engage business stakeholders.