Assignment 4: Evaluation Quiz

5/6 points earned (83%)

Quiz passed!

Continue Course

Back to Week 5

0 / 1 points

1.

You are responsible for a recommender system for an e-commerce site that has three slots in which to recommend products at check-out time. In general, you are most interested in whether people buy additional products based on the recommendations, and you will be measuring sales and lift when your system is live, but you want an offline measure to help determine which potential recommenders are worth trying online. A few other details: this is a domain where people do occasionally re-purchase items, but not frequently. Most regular customers visit between once a week and once every three months. The site has complete details on all previous purchases, and has customer ratings for about 5% of the purchased items. Also, the site does a pretty good job recommending to new customers based on demographics and overall popularity, so you are focused on finding recommenders that take advantage of information learned from customers who have already purchased several items.
Which of the following evaluation plans/metrics seems best?

✔ 1 / 1 points

2.
Which of the following is a situation in which Mean Absolute Error is a reasonable choice as a metric for evaluating recommender performance?

✔ 1 / 1 points

3.
Which of the following is a situation in which it would be most useful to tune the recommender using a metric such as the receiver operating characteristic -- specifically, tuning the algorithm to find the right trade-off between true positive and false positive rates?

✔ 1 / 1 points

4.

All of these situations are ones where it would make sense to test different recommenders empirically through A/B tests or other field tests. In which situation is it LEAST LIKELY that you could get useful data by asking users which set of outputs they prefer? In other words, in which situation are users least likely to know whether the recommender is actually achieving its goals?

1 / 1 points

5.

We have argued that the real proof of a recommender system is in the usage, and that offline evaluation has serious problems. At the same time, there are many situations where offline evaluation does make sense. Which of these is a valid reason for carrying out off-line metric-based evaluation rather than a live user study of a recommender?

1 / 1 points

6.

Which of the following is a valid objection to the validity of offline evaluation using metrics such as MAE, top-n Precision, or nDCG?