

Google Discussions



Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 56 DISCUSSIO..

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 56

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You work for a credit card company and have been asked to create a custom fraud detection model based on historical data using AutoML Tables. You need to prioritize detection of fraudulent transactions while minimizing false positives. Which optimization objective should you use when training the model?

- A. An optimization objective that minimizes Log loss
- B. An optimization objective that maximizes the Precision at a Recall value of 0.50
- C. An optimization objective that maximizes the area under the precision-recall curve (AUC PR) value
- D. An optimization objective that maximizes the area under the receiver operating characteristic curve (AUC ROC) value

Show Suggested Answer

by [inder0007](#) at July 6, 2021, 11:36 a.m.

Comments

Type your comment...

Submit

Paul_Dirac Highly Voted 3 years, 2 months ago

This is a case of imbalanced data.

Ans: C

<https://stats.stackexchange.com/questions/262616/roc-vs-precision-recall-curves-on-imbalanced-dataset>

<https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc>

   upvoted 20 times

  **GogoG** 3 years ago

C is wrong - correct answer is D. ROC basically compares True Positives against False Negative, exactly what we are trying to optimise for.

   upvoted 2 times

  **ralf_cc** Highly Voted 3 years, 3 months ago

D - https://en.wikipedia.org/wiki/Receiver_operating_characteristic

   upvoted 8 times

  **omar_bh** 3 years, 3 months ago

True. The true positive is presented by Y axis. The bigger the area the graph take, the higher TP ratio

   upvoted 2 times

  **tavva_prudhvi** 1 year, 3 months ago

A larger area under the ROC curve does indicate a better model performance in terms of correctly identifying true positives. However, it does not take into account the imbalance in the class distribution or the costs associated with false positives and false negatives.

In contrast, the AUC PR curve focuses on the trade-off between precision (Y-axis) and recall (X-axis), making it more suitable for imbalanced datasets and applications with different costs for false positives and false negatives, like credit card fraud detection.

   upvoted 2 times

  **tavva_prudhvi** 1 year, 3 months ago

AUC ROC is more suitable when the class distribution is balanced and false positives and false negatives have similar costs.

In the case of credit card fraud detection, the class distribution is typically imbalanced (fewer fraudulent transactions compared to non-fraudulent ones), and the cost of false positives (incorrectly identifying a transaction as fraudulent) and false negatives (failing to detect a fraudulent transaction) are not the same.

By maximizing the AUC PR (area under the precision-recall curve), the model focuses on the trade-off between precision (proportion of true positives among predicted positives) and recall (proportion of true positives among actual positives), which is more relevant in imbalanced datasets and for applications where the costs of false positives and false negatives are not equal. This makes option C a better choice for credit card fraud detection.

   upvoted 2 times

  **jkkim_jt** Most Recent 2 days, 13 hours ago


Selected Answer: C

o AUC-PR focuses on how well the classifier performs for the positive class (precision and recall are both concerned with positives) --> more suitable when the focus is on indentifying the positive class in imbalanced data

o AUC-ROC looks at the trade-off between the true positive rate (sensitivity) and the false positive rate, considering both classes. --> general purpose meric that works well when both classes are of similiar size

(ChatGPT)

   upvoted 1 times

  **PhilipKoku** 4 months, 2 weeks ago

Selected Answer: C

C) PR (Precision Recall)

   upvoted 1 times

  **PhilipKoku** 4 months, 2 weeks ago

Selected Answer: C

C) PR ROC

   upvoted 1 times

  **tavva_prudhvi** 1 year, 3 months ago


Selected Answer: C

In fraud detection, it's crucial to minimize false positives (transactions flagged as fraudulent but are actually legitimate) while still detecting as many fraudulent transactions as possible. AUC PR is a suitable optimization objective for this scenario because it provides a balanced trade-off between precision and recall, which are both important metrics in fraud detection. A high AUC PR value indicates that the model has high precision and recall, which means it can detect a large number of fraudulent transactions while minimizing false positives.

minimizing false positives.

Log loss (A) and AUC ROC (D) are also commonly used optimization objectives in machine learning, but they may not be as effective in this particular scenario. Precision at a Recall value of 0.50 (B) is a specific metric and not an optimization objective.

   upvoted 4 times

  **M25** 1 year, 5 months ago

Selected Answer: C

Went with C

   upvoted 1 times

  **John_Pongthorn** 1 year, 8 months ago

Selected Answer: C

Hi Everyone

I discover, there are some clues that this question is likely to refer to the last section of <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>

This is what it tries to tell us especially with the last sentence

Classification-threshold invariance is not always desirable. In cases where there are wide disparities in the cost of false negatives vs. false positives, it may be critical to minimize one type of classification error. For example, when doing email spam detection, you likely want to prioritize minimizing false positives (even if that results in a significant increase of false negatives). AUC isn't a useful metric for this type of optimization.

Additionally, it tells me which of the following choices is the answer to this question as below.

<https://cloud.google.com/automl-tables/docs/train#opt-obj>.

   upvoted 1 times

  **enghabeth** 1 year, 8 months ago

Selected Answer: D

What is different however is that ROC AUC looks at a true positive rate TPR and false positive rate FPR while PR AUC looks at positive predictive value PPV and true positive rate TPR.

Detect Fraudulent transactions = Max TP

Minimizing false positives -> min FP

<https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc#:~:text=ROC%20AUC%20vs%20PR%20AUC&text=What%20is%20different%20however%20is,and%20true%20positive%20rate%20TPR>

   upvoted 1 times

  **John_Pongthorn** 1 year, 9 months ago

Selected Answer: C

Detection of fraudulent transactions seems to be imbalanced data.

<https://cloud.google.com/automl-tables/docs/train#opt-obj>

AUC ROC : Distinguish between classes. Default value for binary classification.

AUC PR Optimize results for predictions for the less common class.

it is straightforward to answer, you just have to capture key word to get the right way. (Almost banlanced Or Imbalanced)

<https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/>

When to Use ROC vs. Precision-Recall Curves?

Generally, the use of ROC curves and precision-recall curves are as follows:

ROC curves should be used when there are roughly equal numbers of observations for each class.

Precision-Recall curves should be used when there is a moderate to large class imbalance.


   upvoted 3 times

  **ares81** 1 year, 9 months ago

Selected Answer: C

Fraud Detection --> Imbalanced Dataset ---> AUC PR --> C, for me

   upvoted 1 times

  **wish0035** 1 year, 10 months ago

Selected Answer: C

ans: C

Paul_Dirac and giaZ are correct.

   upvoted 1 times



  **hiromi** 1 year, 10 months ago

Selected Answer: C

-

C
<https://towardsdatascience.com/on-roc-and-precision-recall-curves-c23e9b63820c>

   upvoted 2 times

  **itallix** 2 years, 1 month ago

"You need to prioritize detection of fraudulent transactions while minimizing false positives."

Seems that answer B fits this well. If we want to focus exactly on minimizing false positives we can do that by maximising Precision at a specific Recall value. C is about balance between these two, and D doesn't care about false positive/negatives.

   upvoted 2 times

  **suresh_vn** 2 years, 2 months ago

Selected Answer: D

D

https://en.wikipedia.org/wiki/Receiver_operating_characteristic




C optimize precision only



   upvoted 1 times

  **suresh_vn** 2 years, 2 months ago

Sorry, C is my final decision

<https://cloud.google.com/automl-tables/docs/train#opt-obj>



   upvoted 1 times

  **rtnk22** 2 years, 2 months ago

Selected Answer: C

Answer is c.

   upvoted 1 times

  **giaZ** 2 years, 7 months ago

https://icaiit.org/proceedings/6th_ICAIIT/1_3Fayzrakhmanov.pdf

The problem of fraudulent transactions detection, which is an imbalanced classification problem (most transactions are not fraudulent), you want to maximize both precision and recall; so the area under the PR curve. As a matter of fact, the question asks you to focus on detecting fraudulent transactions (maximize true positive rate, a.k.a. Recall) while minimizing false positives (a.k.a. maximizing Precision). Another way to see it is this: for imbalanced problems like this one you'll get a lot of true negatives even from a bad model (it's easy to guess a transaction as "non-fraudulent" because most of them are!), and with high TN the ROC curve goes high fast, which would be misleading. So you wanna avoid dealing with true negatives in your evaluation, which is precisely what the PR curve allows you to do.

   upvoted 6 times

[Load full discussion...](#)

Start Learning for free



Social Media

[Facebook](#) , [Twitter](#)

[YouTube](#) , [Reddit](#)



We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.



© 2024 ExamTopics

ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.