Stanford University

CS 329P : Practical Machine Learning (2021 Fall)

4.1 Evaluation Metrics

Qingqing Huang, Mu Li, Alex Smola

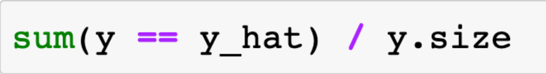
https://c.d2l.ai/stanford-cs329p

Model Metrics

- Loss measures how good the model in predicting the outcome in supervised learning
- Other metrics to evaluate the model performance
  - Model specific: e.g. accuracy for classification, mAP for object detection
  - Business specific: e.g. revenue, inference latency
- We select models by multiple metrics
  - Just like how you choose cars

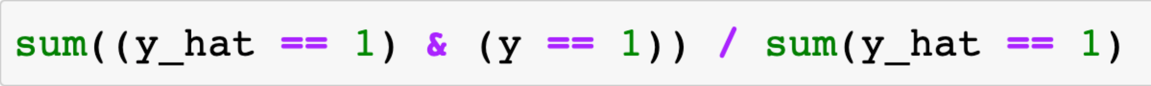
Metrics for Binary Classification

- Accuracy: # correct predictions / # examples



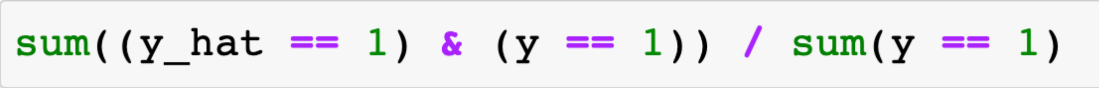
```
sum(y == y_hat) / y.size
```

- Precision: # True positive / # (True positive + False positive)



```
sum((y_hat == 1) & (y == 1)) / sum(y_hat == 1)
```

- Recall: # True positive / # Positive examples

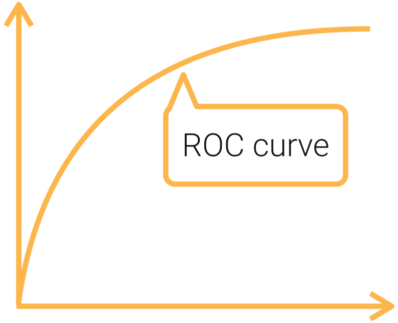


```
sum((y_hat == 1) & (y == 1)) / sum(y == 1)
```

- Be careful of division by 0
- One metric that balances precision and recall
  - F1: the harmonic mean of precision and recall: $2pr/(p+r)$

AUC-ROC

- Measures how well the model can separate the two classes
- Choose decision threshold θ, predict positive if $o \geq \theta$ else neg
- In the range [0.5, 1]



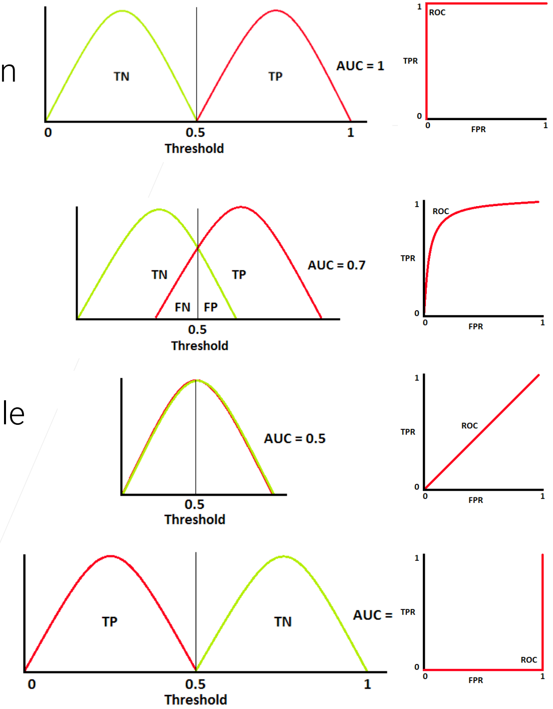
$$\text{TPR} = \frac{\text{\# True Pos predictions}}{\text{\# pos examples}}$$

FPR= # False Pos predictions / # neg examples

Distribution of output o

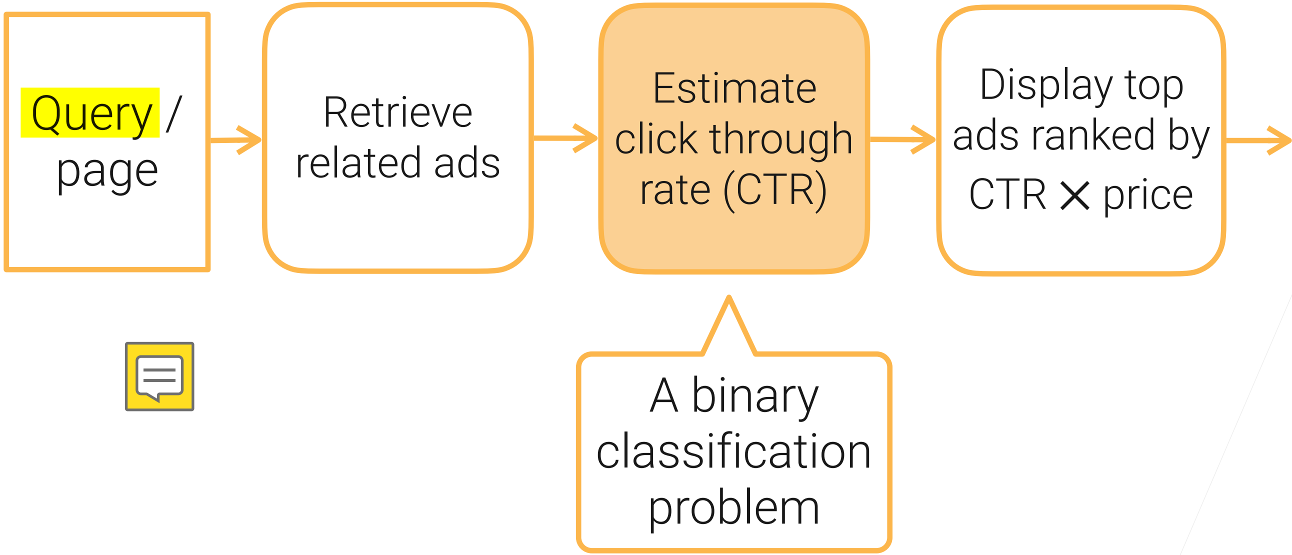
Ideal separation

Non separable



Case Study: Displaying Ads

- Ads is one major revenue source for Internet companies

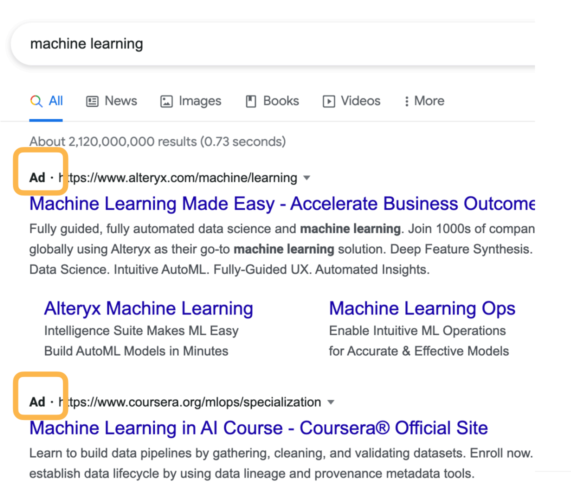

Query / page

Retrieve related ads

Estimate click through rate (CTR)

Display top ads ranked by CTR × price

A binary classification problem


machine learning

About 2,120,000,000 results (0.73 seconds)

Ad · https://www.alteryx.com/machine/learning

Machine Learning Made Easy - Accelerate Business Outcome

Fully guided, fully automated data science and **machine learning**. Join 1000s of compan globally using Alteryx as their go-to **machine learning** solution. Deep Feature Synthesis. Data Science. Intuitive AutoML. Fully-Guided UX. Automated Insights.

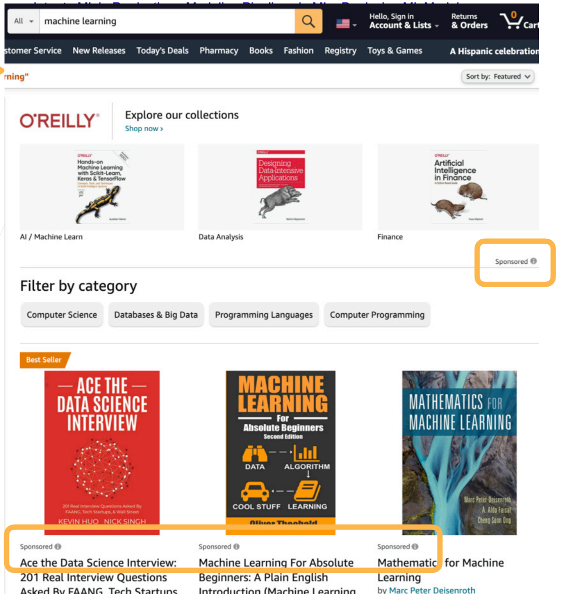
Alteryx Machine Learning
Intelligence Suite Makes ML Easy
Build AutoML Models in Minutes

Machine Learning Ops
Enable Intuitive ML Operations
for Accurate & Effective Models

Ad · https://www.coursera.org/mlops/specialization

Machine Learning in AI Course - Coursera® Official Site

Learn to build data pipelines by gathering, cleaning, and validating datasets. Enroll now. establish data lifecycle by using data lineage and provenance metadata tools.



Explore our collections
Shop now ›

AI / Machine Learn

Data Analysis

Finance

Sponsored

Filter by category

Computer Science | Databases & Big Data | Programming Languages | Computer Programming

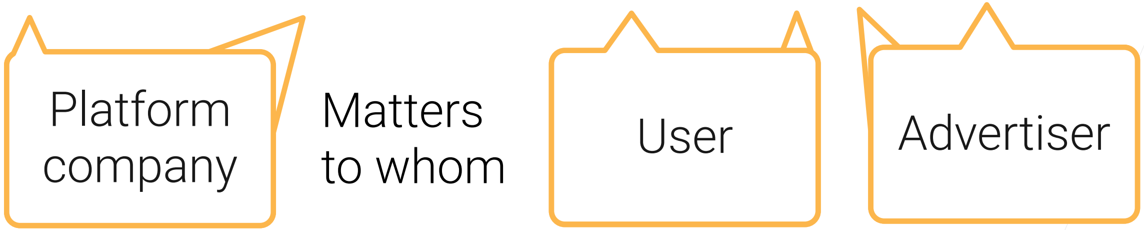
Sponsored
Ace the Data Science Interview: 201 Real Interview Questions Asked By FAANG, Tech Startups

Sponsored
Machine Learning For Absolute Beginners: A Plain English Introduction (Machine Learning

Sponsored
Mathematic for Machine Learning
by Marc Peter Deisenroth

Business Metrics for Displaying Ads

- Optimize both revenue and customer experience
  - Latency: ads should be shown to users at the same time as others
  - ASN: average #ads shown in a page
  - CTR: actual user click through rate
  - ACP: average price advertiser pays per click
- revenue = #pageviews × ASN × CTR × ACP



Platform company

Matters to whom

User

Advertiser

Displaying Ads: Model → Business Metrics

- The key model metric is AUC
- A new model with increased AUC may harm business metrics, possible reasons:
  - Lower estimated CTR → less ads displayed
  - Lower real CTR because we trained and evaluated on past data
  - Lower prices
- Online experiment: deploy models to evaluate on real traffic data

Summary

- We evaluate models with multiple metrics
- Model metrics evaluate model performance on examples
  - E.g. accuracy, precision, recall, F1, AUC for classification models
- Business metrics measure how models impact the product