

Log In

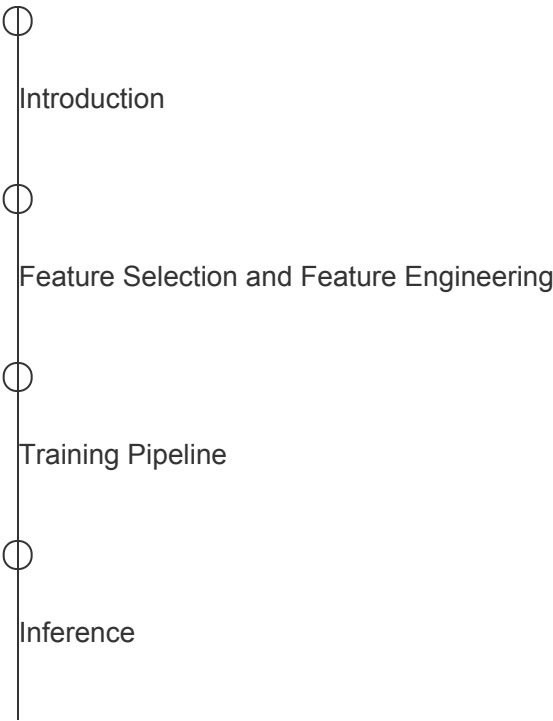
Join

Back To Module Home

Machine Learning System Design

0% completed

Machine Learning Primer



Metrics Evaluation

Video Recommendation

Feed Ranking

Ad Click Prediction

Rental Search Ranking

Estimate Food Delivery Time

Machine Learning Knowledge

Machine Learning Model Diagnosis

Conclusion

Mark Module as Completed

Metrics Evaluation

Learn different metrics for model evaluations.

We'll cover the following

- Metrics evaluation
 - Offline metrics
 - Online metrics

Metrics evaluation#

In practice, it's common that the model performs well during offline evaluation but does not perform well when in production. Therefore, it is important to measure model

performance in both on and offline environments.

Offline metrics#

- During offline training and evaluating, we use metrics like logloss, MAE, and R2 to measure the goodness of fit. Once the model shows improvement, the next step would be to move to the staging/sandbox environment to test for a small percentage of real traffic.

Online metrics#

- During the staging phase, we measure certain metrics, such as Lift in revenue or click through rate, to evaluate how well the model recommends relevant content to users. Consequently, we evaluate the impact on business metrics. If the observed revenue-related metrics show consistent improvement, then it is safe to gradually expose the model to a larger percentage of real traffic. Finally, when we have enough evidence that new models have improved revenue metrics, we can replace the current production models with new models. For further reading, explore how Sage Maker enables A/B testing or LinkedIn A/B testing.
- This diagram shows one way to allocate traffic to different models in production. In reality, there will be few a dozen models, each getting real traffic to serve online requests. This is one way to verify whether or not a model actually generates lift in the production environment.

Allocate traffic for multiple models in production

- AB testing is an extensive subject and is use case specific. Read more about A/B

testing here.

Back

Inference

Next

Problem Statement and Metrics

Mark as Completed

Report an Issue