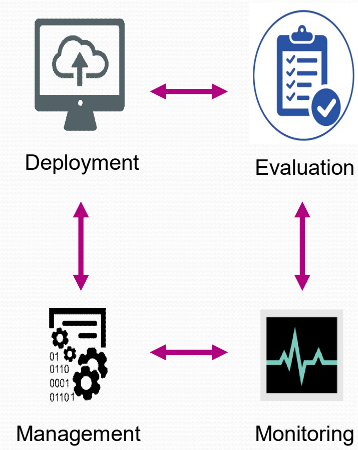
ML_After deployment

Manisha V Jadhav

What happens after (initial) deployment

ML production life cycle

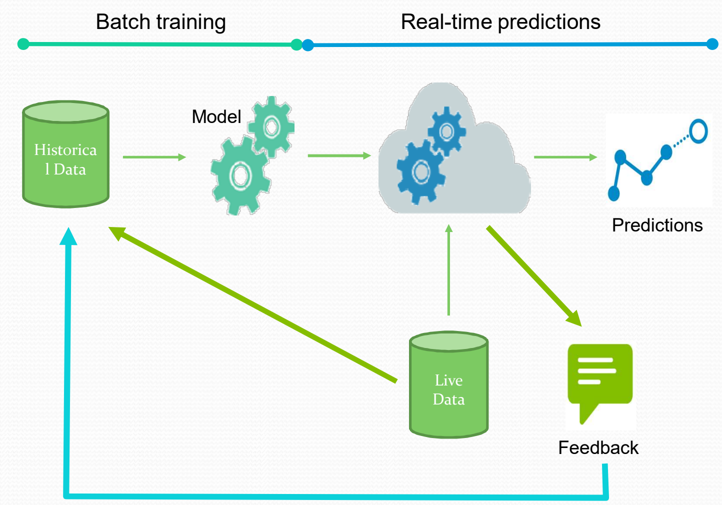


After deployment

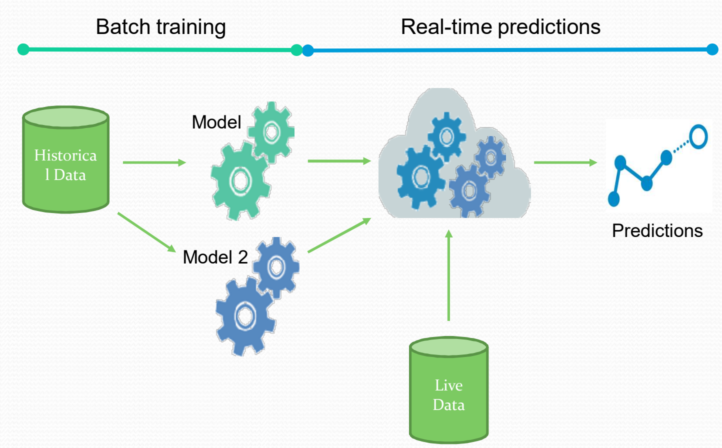
Evaluate and track metrics over time.

React to feedback from deployed models.

ML in production



ML in production -



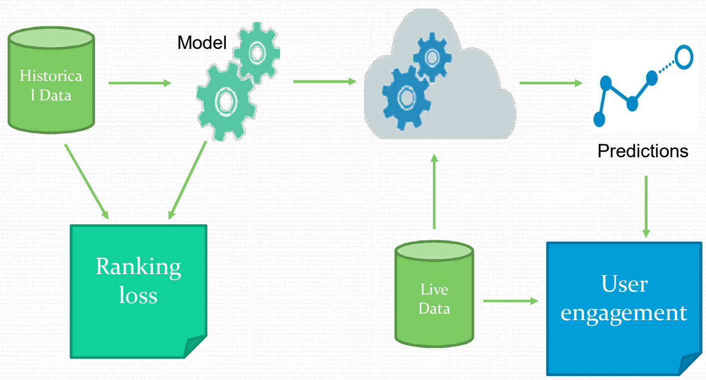
Key questions

- When to update a model?
- How to choose between existing models?
- Answer: continuous evaluation and testing

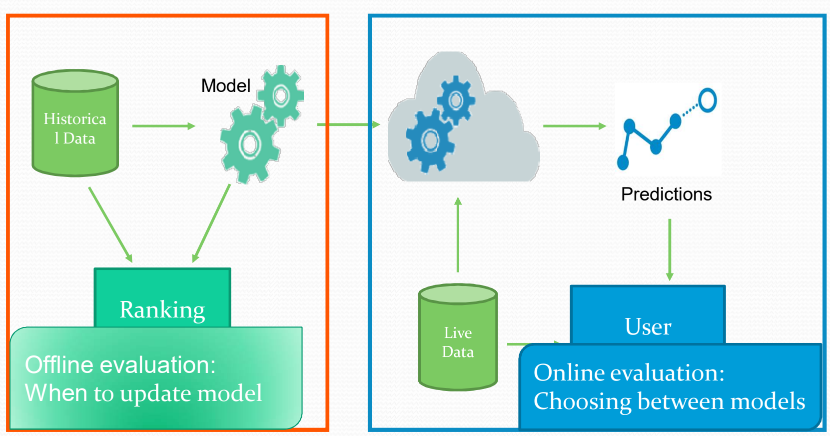
What is evaluation?

What data?
Which metric?

Evaluating a recommender



Evaluating a recommender



Updating ML models

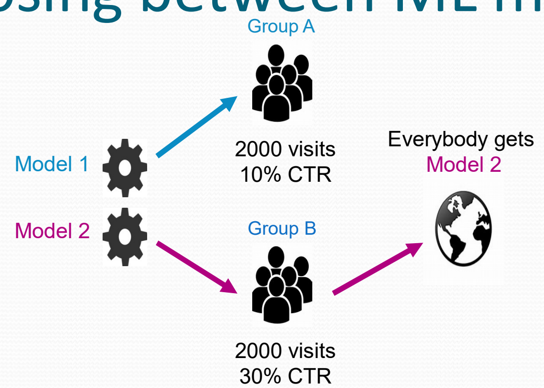
Why update?

- Trends and user tastes change over time
- Model performance drops

When to update?

- Track statistics of data over time
- Monitor both offline & online metrics on live data
- Update when offline metric diverges from online metrics

Choosing between ML models



Strategy 1: A/B testing—select the best model and use it all the time

Choosing between ML models

A statistician walks into a casino…

Multi-armed bandits

Pay-off \$1:\$1000

Play this 5% of the time

Pay-off \$1:\$200

Play this 85% of the time

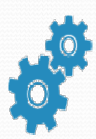
Pay-off \$1:\$500

Play this 10% of the time

Choosing between ML models

A statistician walks into an ML production environment

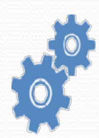
Model 1



Pay-off \$1:\$1000

Use this 5% of the time (Exploration)

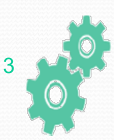
Model 2



Pay-off \$1:\$200

Use this 85% of the time (Exploitation)

Model 3



Pay-off \$1:\$500

Use this 10% of the time (Exploration)

MAB vs. A/B testing

Why MAB?

- Continuous optimization, “set and forget”
- Maximize overall reward

Why A/B test?

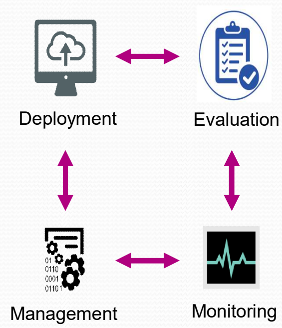
- Simple to understand
- Single winner
- *Tricky* to do right

Other production considerations

- Versioning
- Logging
- Provenance
- Dashboards
- Reports

Conclusions

Distributed
&
Predictive Services



A/B testing,
multi-armed bandits
& much more