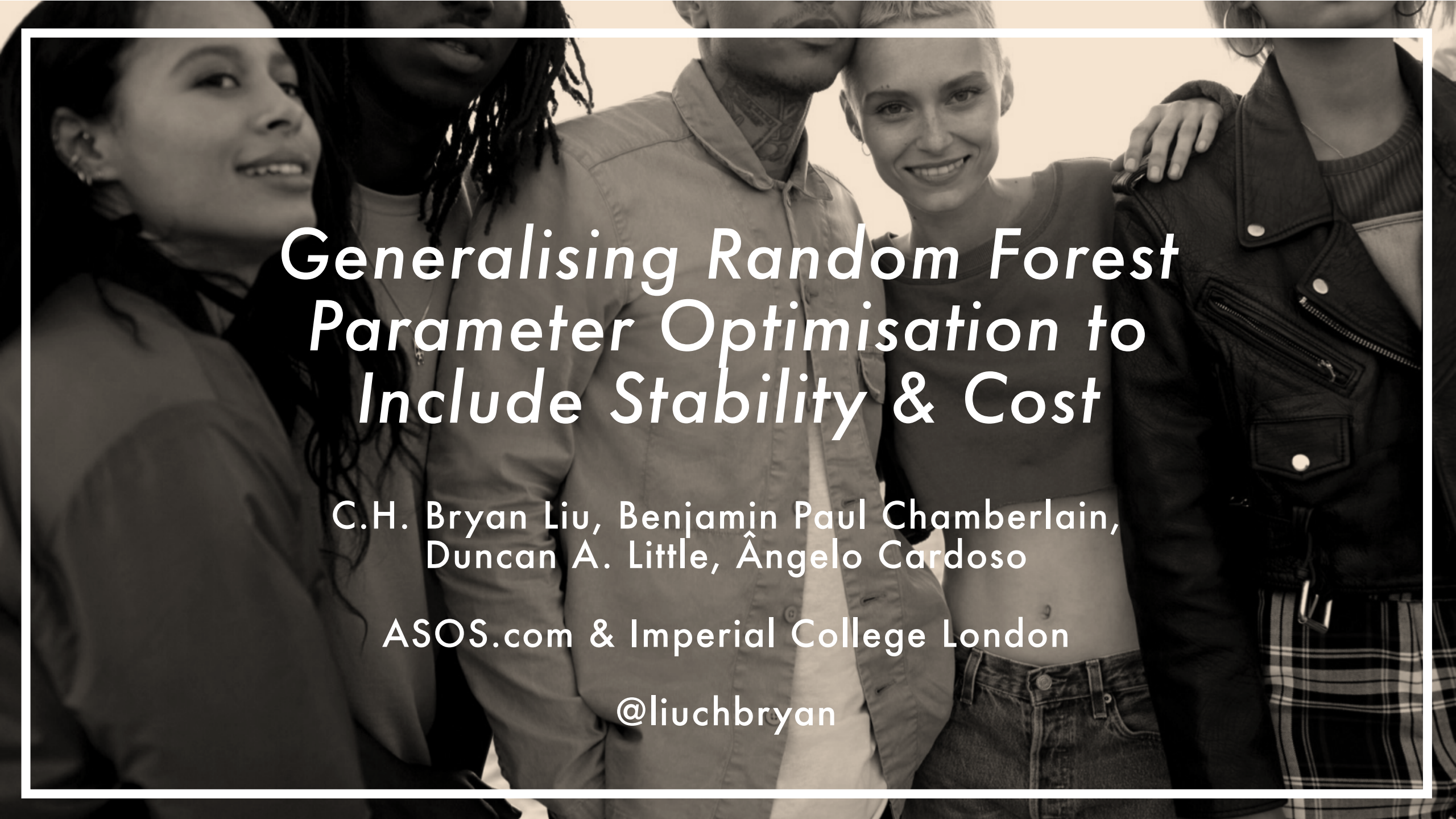


Generalising Random Forest Parameter Optimisation to Include Stability & Cost

C.H. Bryan Liu, Benjamin Paul Chamberlain,
Duncan A. Little, Ângelo Cardoso

ASOS.com & Imperial College London

@liuchbryan

About ASOS.com

We aspire… to be the online shopping destination for fashion-loving 20-somethings

We have… 15m+ active customers
85k products on at anytime (3k new / week)
Billions of customer-product interaction

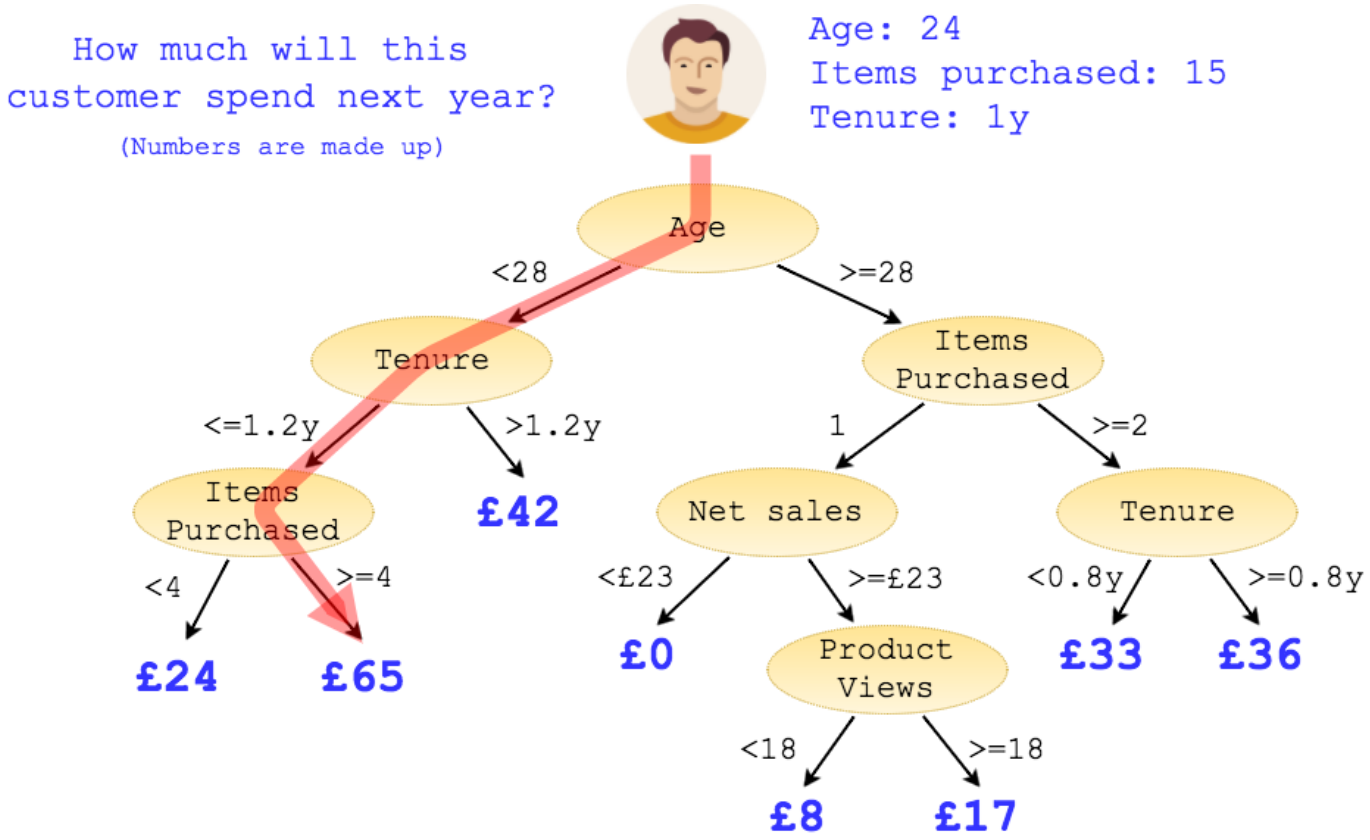
We do R&D in… Customer lifetime value prediction
Recommender systems
Visual properties in products
Demand forecasting

Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos discover fashion online

Imperial College London

Let's talk about random forest



A hand-wavy definition:

- Collection of *decision trees* that learn the set of rules leading to the output
- Randomly sample a subset of training rows & columns when building each tree
- Returns the mode/mean from the trees' outputs

Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos discover fashion online

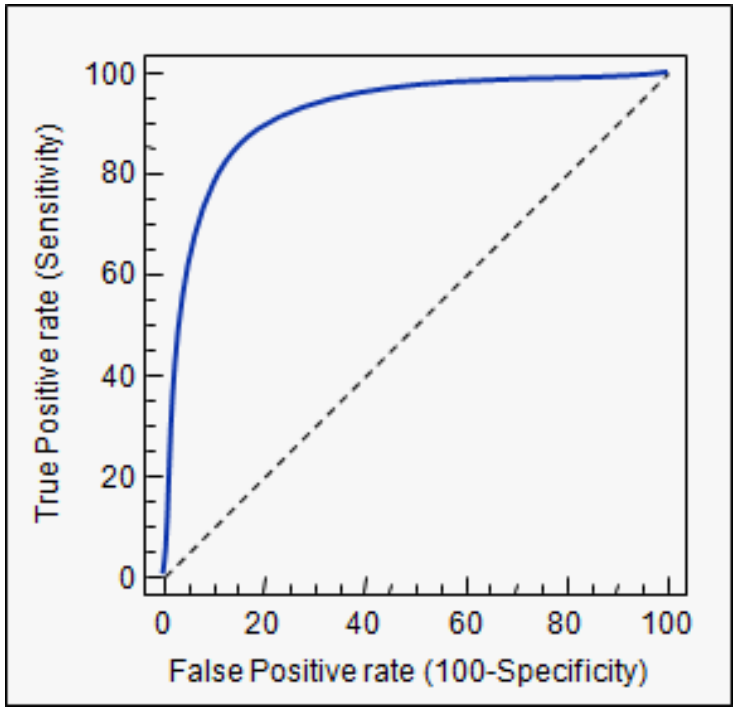
Imperial College London

It requires parameter tuning

E.g.: No. of trees, tree depth, no. of features sampled, …

Goal – achieve better performance (measured by *error reduction* metrics):

- Area Under Receiver Operating Characteristic Curve (AUC)
  (True positive rate vs false positive rate under different decision threshold)
- Root mean squared error (RMSE)
- Classification accuracy
- …

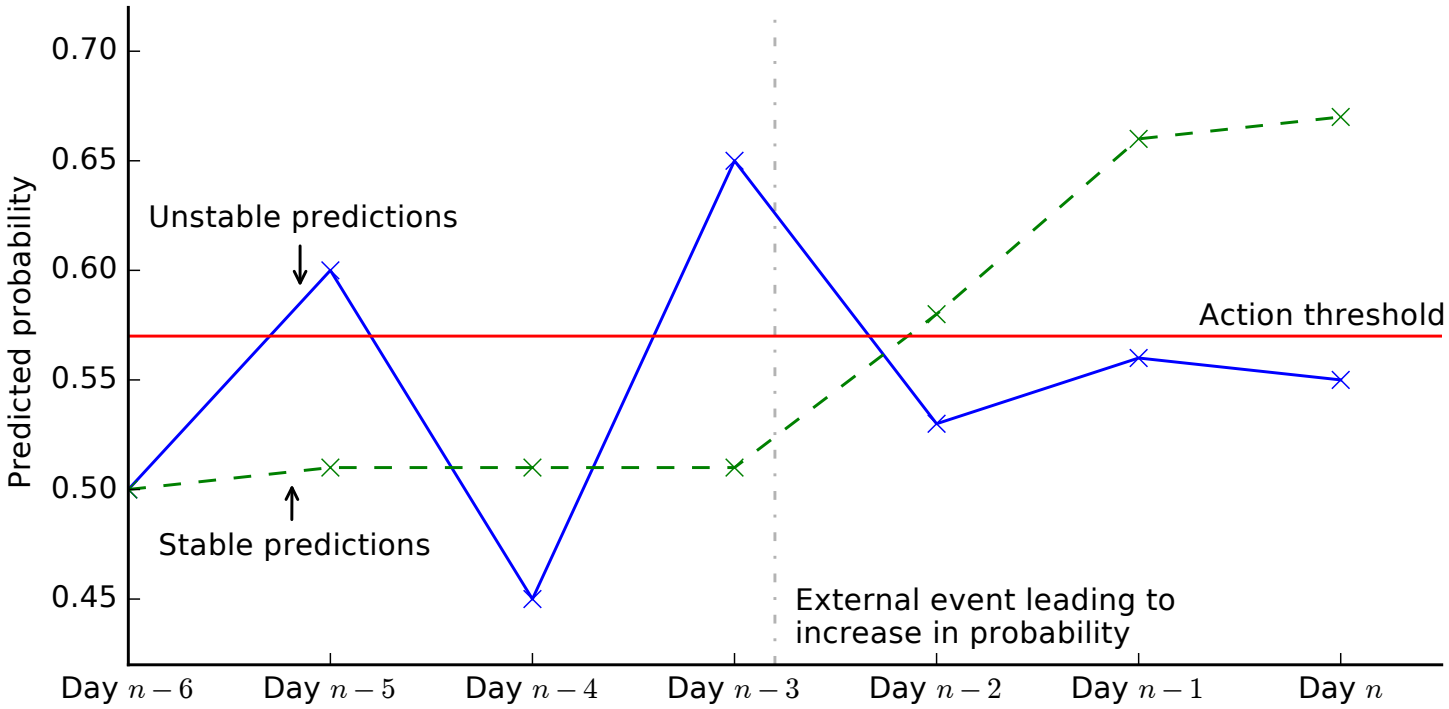


Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos discover fashion online

Imperial College London

Just optimising AUC is not enough!



<u>Are your predictions *stable*?</u>

E.g.: Monitoring the predicted probability and taking action when it changes due to *external* event.

Unstable predictions (arising from model fluctuation, training data sampling) will give you a hard time if the variability is large!

Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos discover fashion online

Imperial College London

Just optimising AUC is not enough!

Do your improved predictions justify the extra *training cost*?

Computing as a service (e.g. Amazon Elastic Cloud, Microsoft Azure) means computation is paid for by the hour.

(Training) Time is money!

Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos discover fashion online

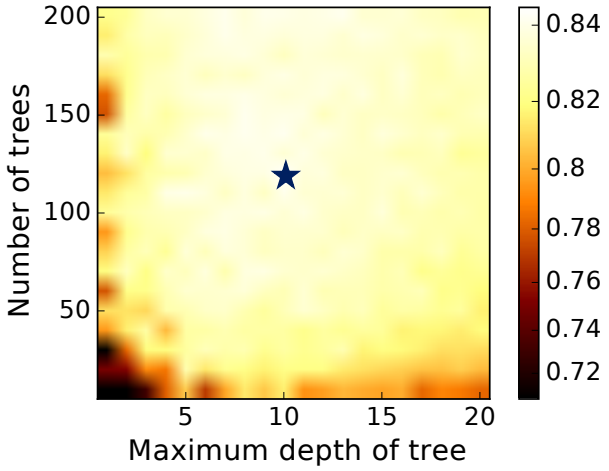
Imperial College London

The cost-stability-error reduction trilemma

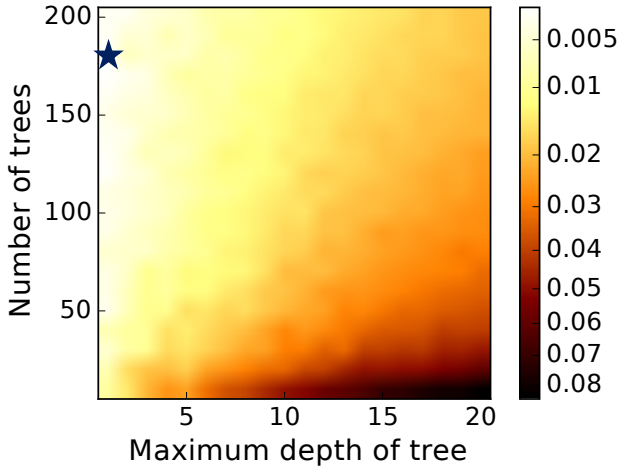
Result of grid search: No. of trees (y-axis) & max. depth of tree (x-axis) against different criteria (lighter = better, star = optimal setting)

Hard to optimise all three criteria at the same time!

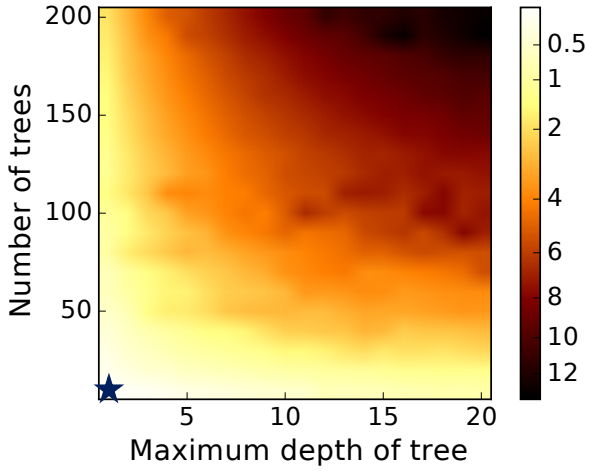
Moderate no. of deep trees
→ high AUC



Large no. of shallow trees
→ high stability



Small no. of shallow trees
→ low cost



Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos
discover fashion online

Imperial College London

Outline

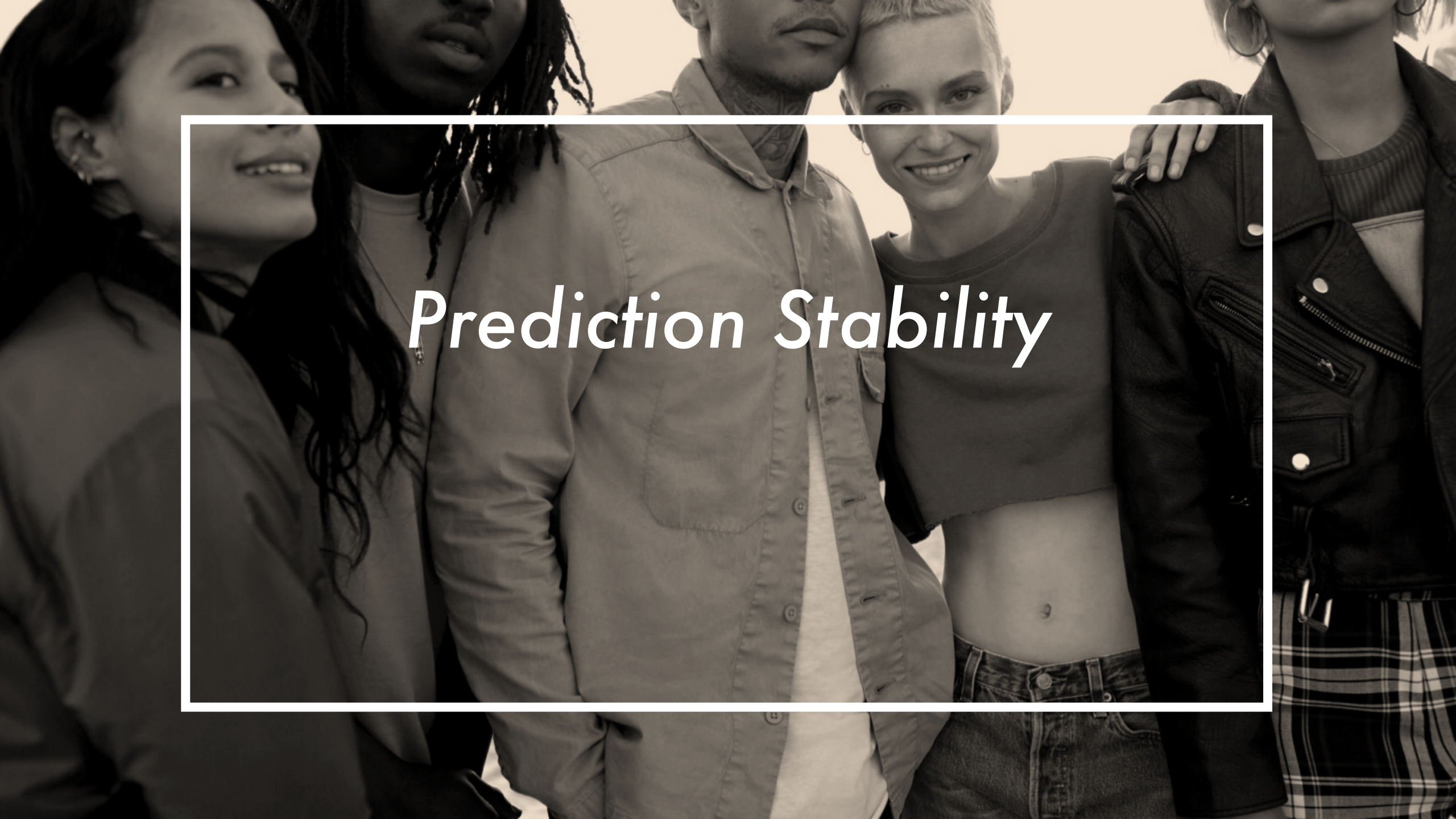
Prediction stability – how do we measure it?

The cost-stability-error reduction trilemma – how do we address it?

Application on two public datasets:

- Orange small (KDD Cup 2009)
- Criteo display advertising challenge (Kaggle 2014)

asos discover fashion online
Imperial College London


Prediction Stability

Prediction Stability – Prior Work

Studied extensively in Computational Learning Theory:

Breiman (1996) relates predictive loss to instability
(A procedure is unstable if a small change in training data causes large changes in predictions)

Elisseeff et al. (2005) extends the notion explicitly to randomised learning algorithms
(Constructed generalisation / leave-one-out error bounds based on model instability)

We want to explore the stability of the prediction themselves in their own right.

Setup

Start with a pool of training & validation data

- The training data may be further sampled in each run to keep training cost low

Do multiple train / validation run on the same datasets

- By doing so, we are measuring the endogenous variability (model fluctuation, training data sampling) only

Read the predicted probability in classification setting

asos discover fashion online
Imperial College London

Key Idea

If I retrain the model and make a fresh prediction on the same validation input,
how far away should I expect that prediction to be compared to the current prediction?

asos
discover fashion online
Imperial College London

Our new stability metric

If I retrain the model and make a fresh prediction on the same validation input,
how far away should I expect that prediction to be compared to the current prediction?

1. Take the squared prediction delta: (old prediction – new prediction)^2
2. Average over all validation data
3. Average over all pairs of runs
4. (For reporting purpose) Take the square root of the average

→ (Root) Mean Squared Prediction Delta, or (R)MSPD

asos discover fashion online
Imperial College London

More in our paper...

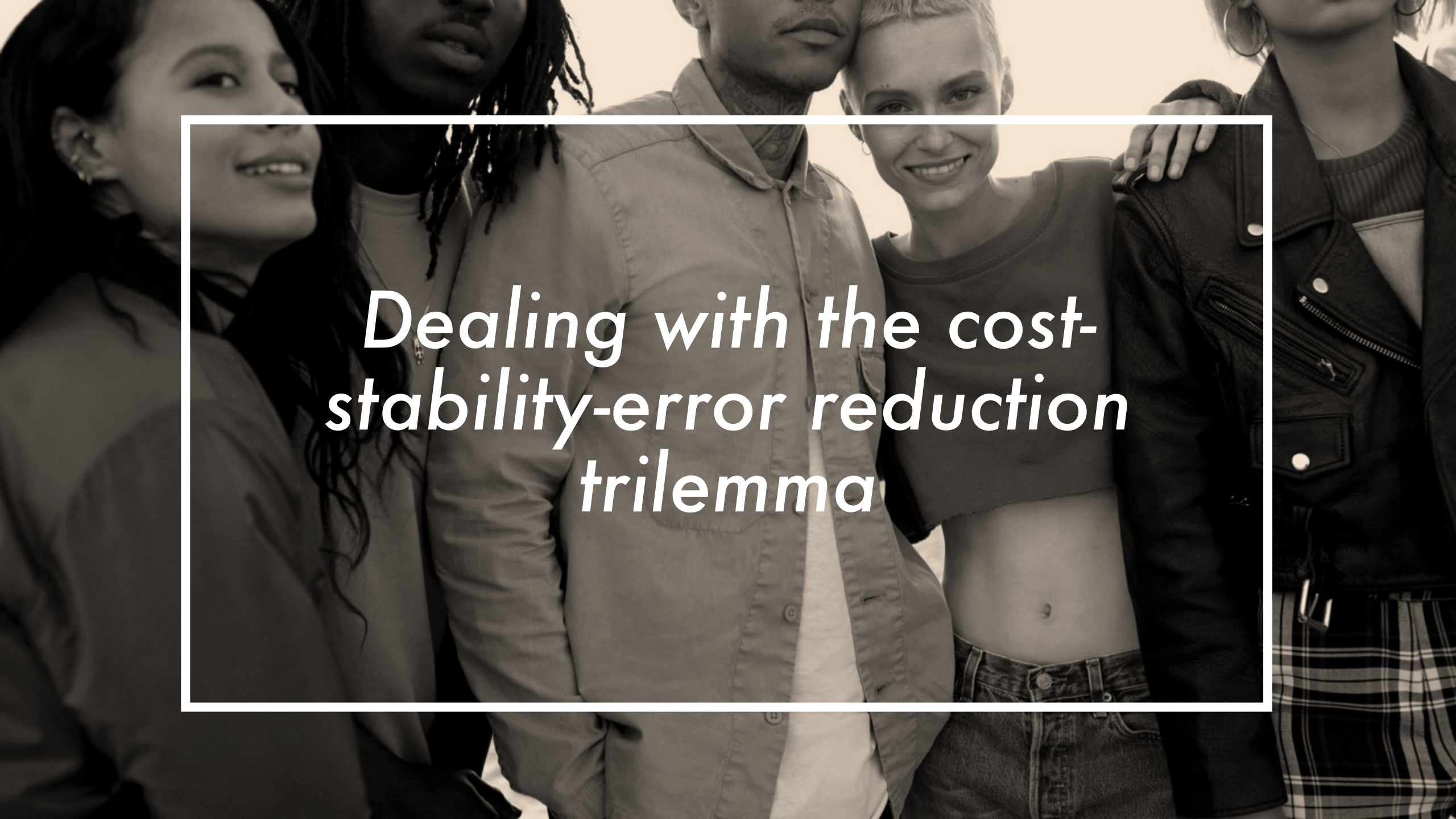
How MSPD captures the interaction between:

- Variance arising from the model itself
- Covariance arising from the further sampling of training data before training in each run

How prediction stability is affected by changing RF parameters

Visualising RMSPD in terms of the distribution of prediction deltas

asos
discover fashion online
Imperial College London



Dealing with the cost-stability-error reduction trilemma

Optimising all three criteria

We want to minimise the generalised loss function:

$$L = \beta\ RMSPD + \gamma\ Runtime\ - \alpha\ AUC$$

… with provision to prioritise any criteria with weight parameters.

asos discover fashion online
Imperial College London

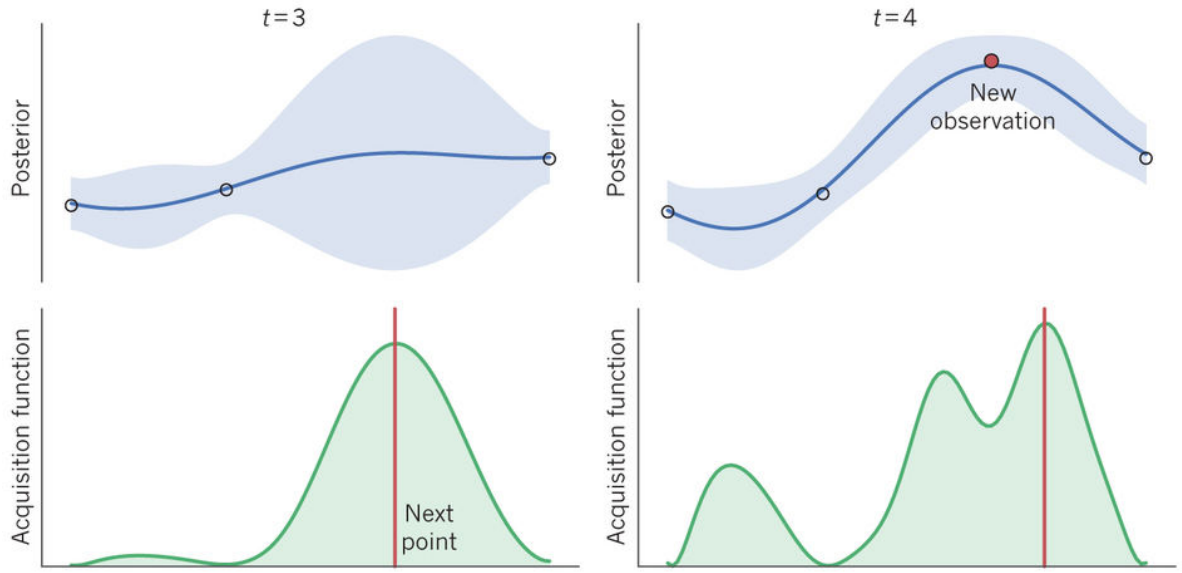
Optimising all three criteria

We want to minimise the generalised loss function:

$$L = \beta\ RMSPD + \gamma\ Runtime\ - \alpha\ AUC$$

- Expensive to evaluate (need to train/validate for each setting)
- Black-box function (no gradient information)
- Potentially noisy

→ Use Bayesian optimisation



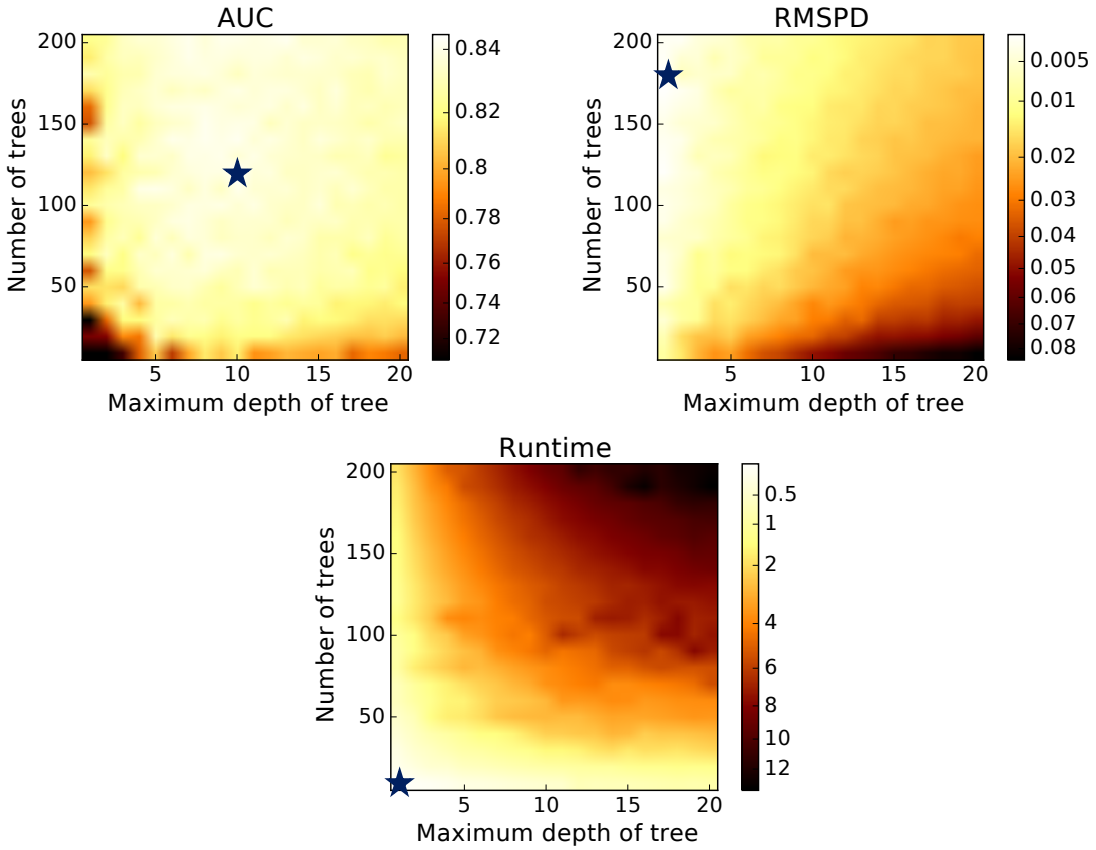
Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos discover fashion online

Imperial College London

Some results

This is what we had when we first touched on the trilemma…



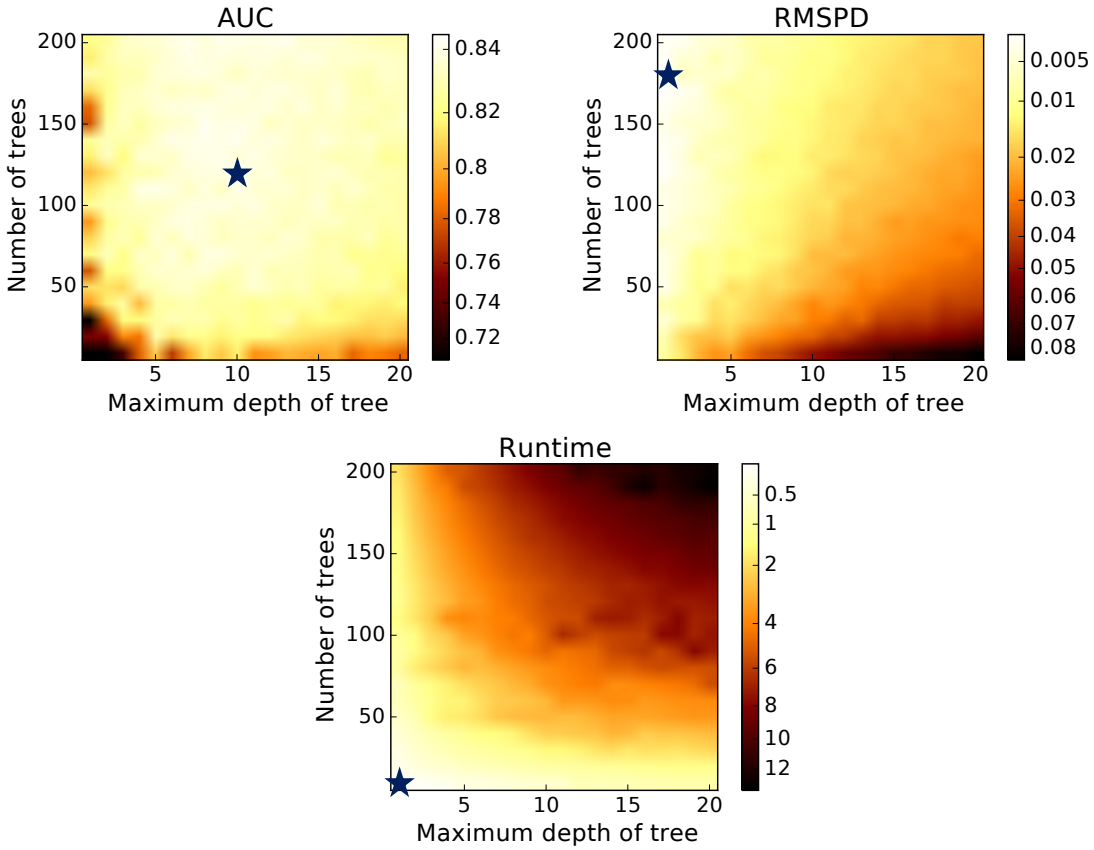
Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

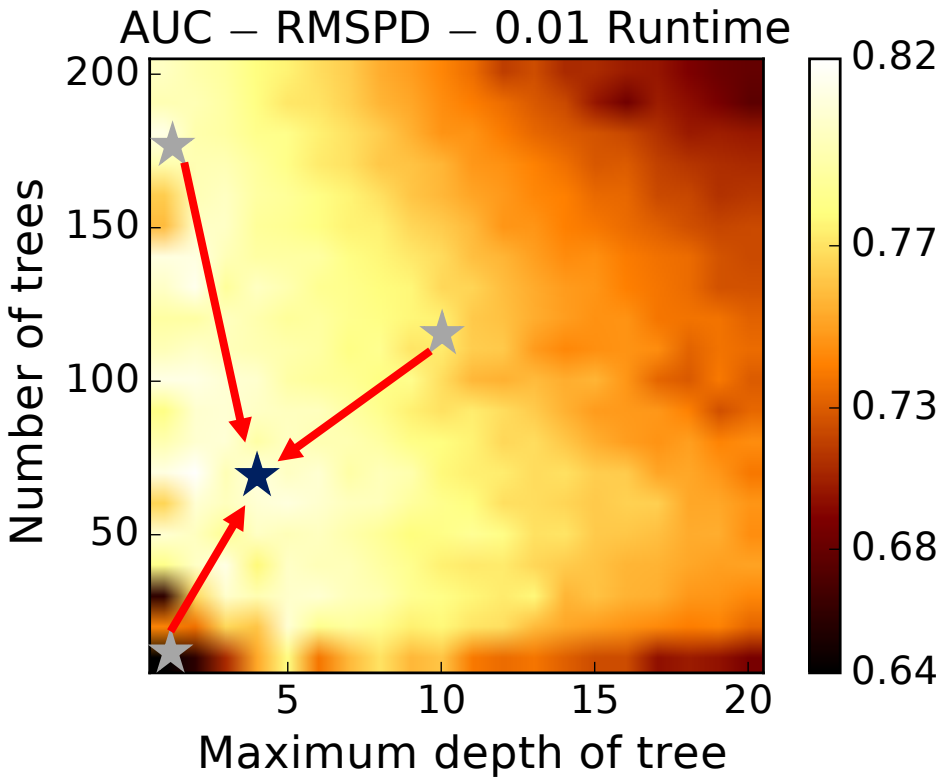
asos discover fashion online

Imperial College London

Some results

Solution found considering the joint effect is markedly different from only optimising AUC.





Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos discover fashion online

Imperial College London

Wrapping up

We developed a new measure for prediction stability – the MSPD.

We also created a new optimisation framework to address the cost-stability-error reduction trilemma via Bayesian optimisation.

The solution found by the framework is markedly different from optimising error reduction metric alone.

Generalising Random Forest Parameter Optimisation - Liu et al. (2017)

asos
discover fashion online

Imperial College London

Thank you!

Any questions?

Code & Docs on:
github.com/liuchbryan/generalised_forest_tuning