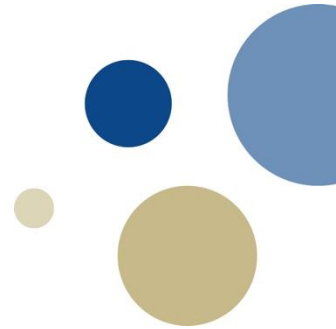


Ensemble methods – Bagging & Boosting

Andreas Gudahl Tufte, Saygin Ileri, Niclas Flehmig

Overview

- 🤓 Motivation
- 💰 Bagging
- 💉 Boosting
- 🤼 Showdown
- 🔍 Pros & Cons
- 👦 💻 Coding Example

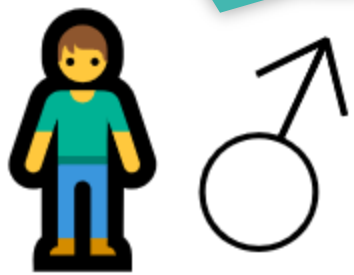
Motivation

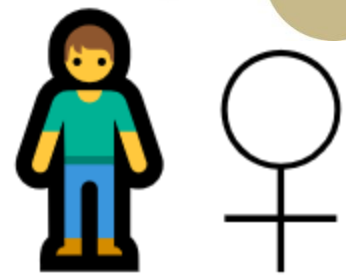
We are better together

It's huge. I guess **300.000**.

I meet the same people everyday. **50.000**.

I think it's **200.000**.





How many people live in Trondheim?

So, it is 183.333

"weak" estimator = better than random guess

Ensemble learning

Combine predictions of several base / ”weak” estimators to increase accuracy (better than a random guess)

Aggregation (Regression) and Voting (Classification)

Statistically

If we average over several good models we find a good approximation of the best

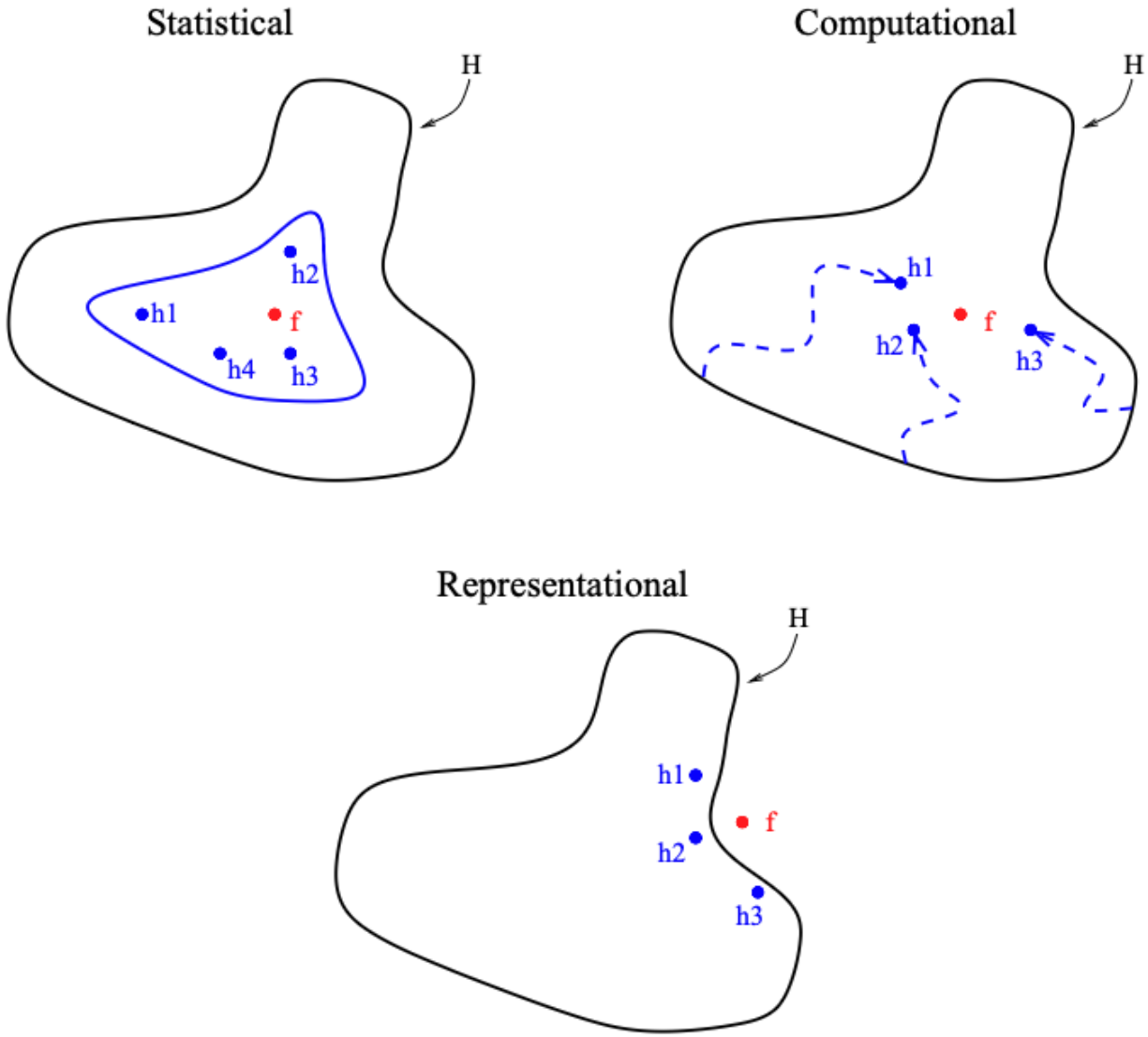
Computationally

- Optimization often gets stucked in local optima, often not solveable
- Starting from different points and getting different local optima may provide a better optimization

Representational

True model cannot be represented by one model but by combining several ones

A figure helps with the explanation



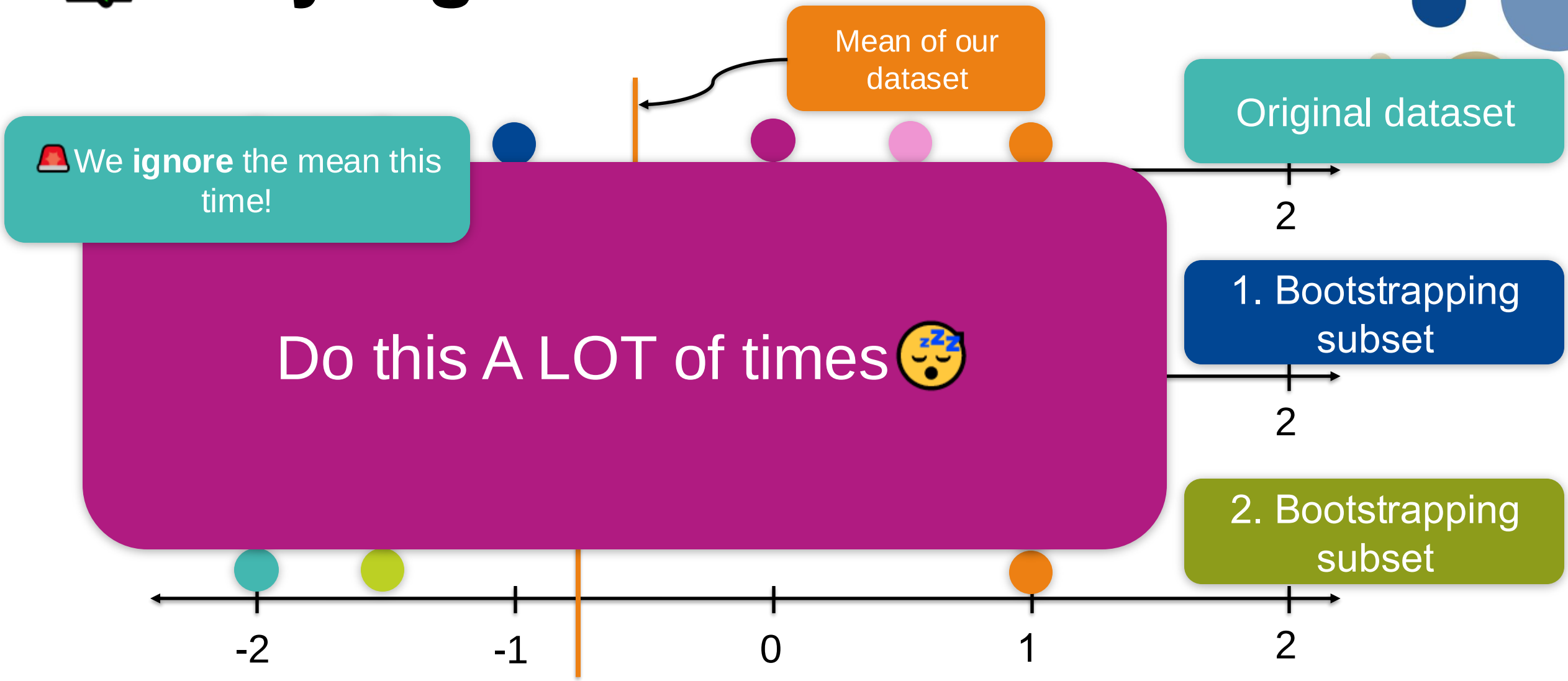
H = space of ”good” models

h_i = “good” models that we found

f = best model

Bagging

Recyling the old boot



Bagging

Bagging = Bootstrapping + Aggregating

Work great for **unstable procedures** (=small changes in data largely affect outcome of model)

Run in **parallel**

Outcomes of model are **equally weighted** in final prediction

Bagging is **NEVER destructive** either not effective or improves the estimation in terms of variance

👦🍳Recipe

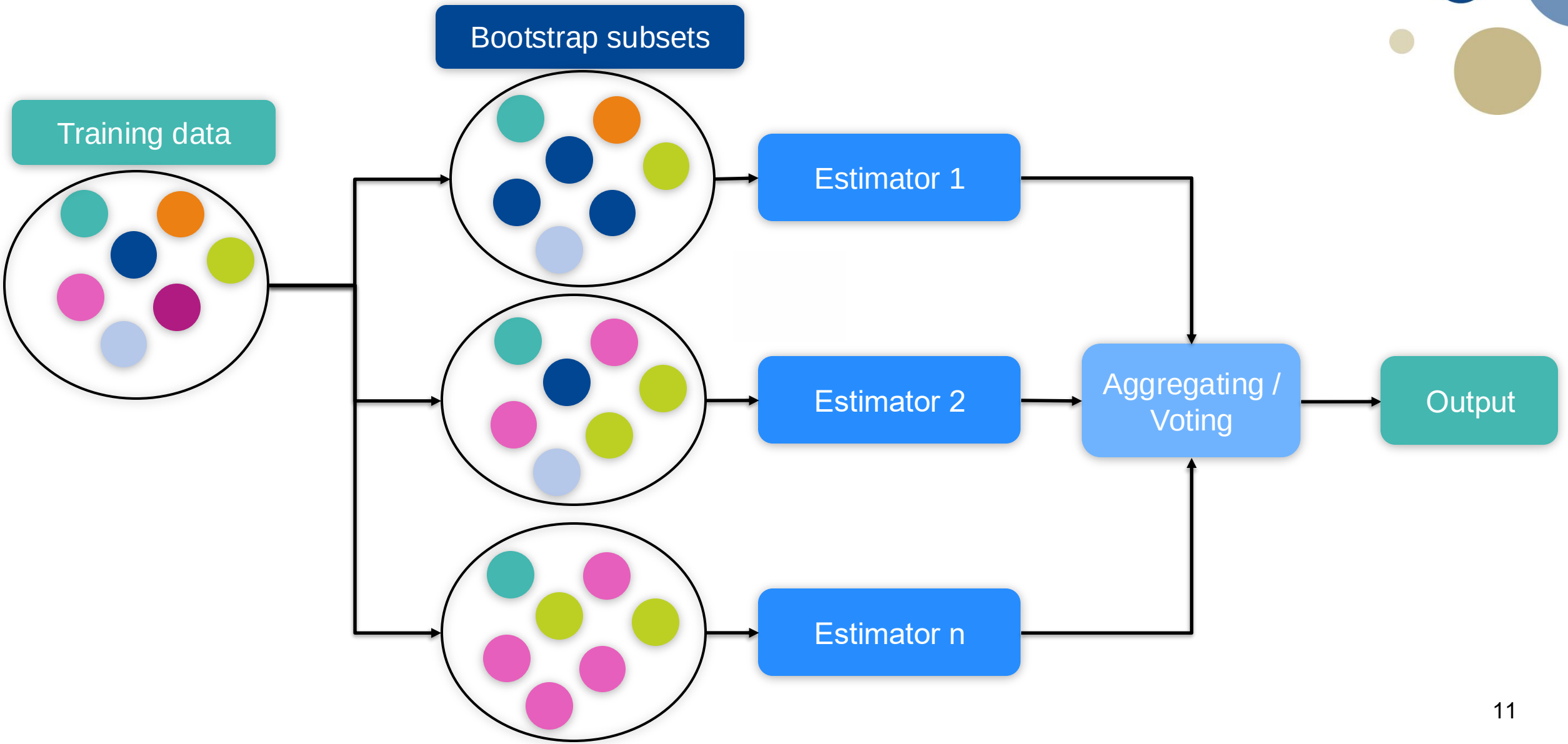
Perform **bootstrapping** to create subsets

Run all subsets in **parallel** with models

Aggregate all predictions / **vote** on final prediction

🥳Congratulations you got a good model! 🥳

How does it look like?



Boosting

Boosting

Take wrong predictions **and increase probability** to be picked in the next round

Next step model **focuses on misclassified** points

Sequential / **hierarchy** structure

Weighted average on final prediction (misclassification rate to determine strength of vote)

🧑🍳 Recipe

Run a model on a dataset

Re-configure the dataset based on the misclassification of the model

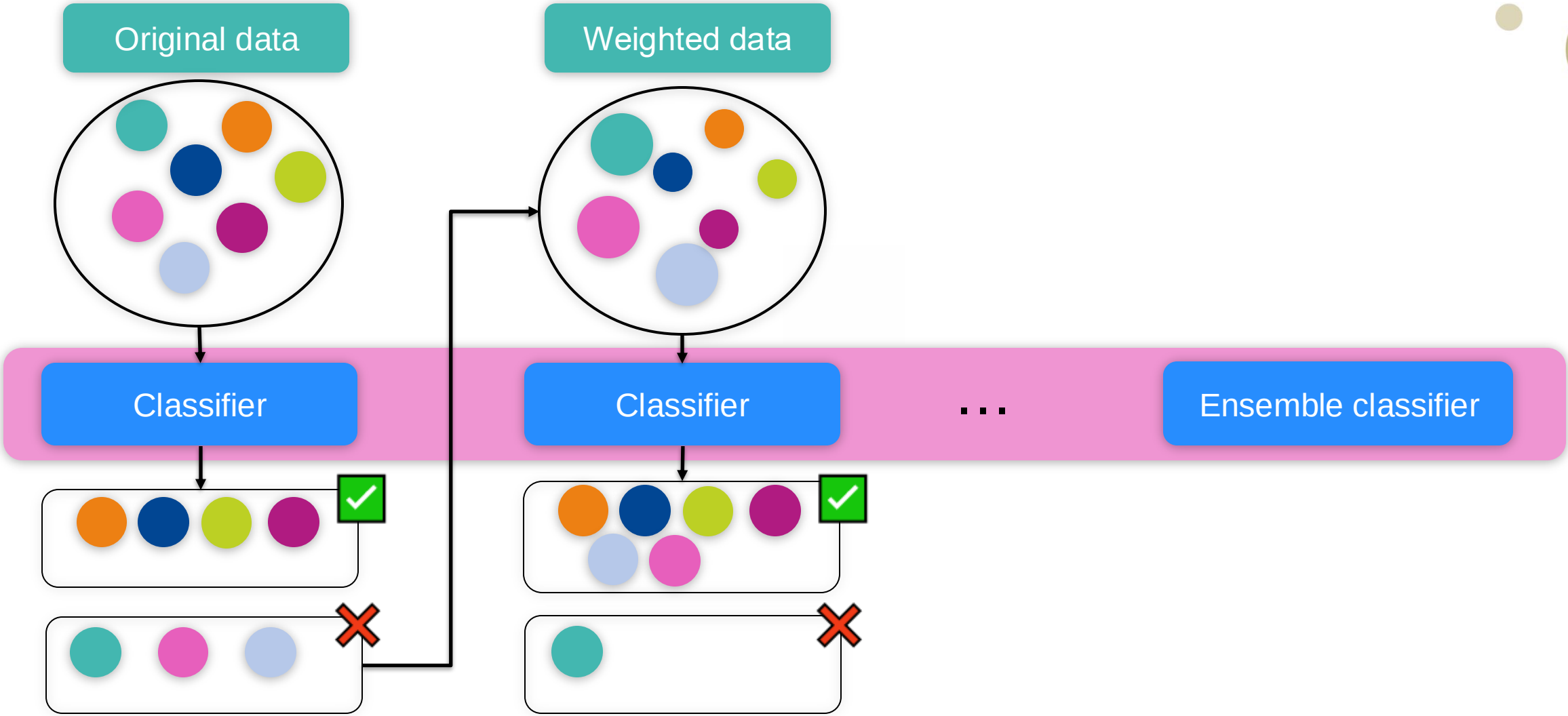
🔁

Run the model on the **new weighted dataset**

Stop the process after some time (e.g. iterations, error threshold) & **aggregate estimates**

🥳 Congratulations you got a good model! 🥳

Show me some example!



Showdown

 Bagging vs  Boosting

Boosting uses same dataset, **Bagging** uses bootstrapping subsets.

Boosting reduces bias of model, **Bagging** not.

Bagging works in parallel while **Boosting** is sequential.

Boosting outperforms **Bagging** on accuracy.

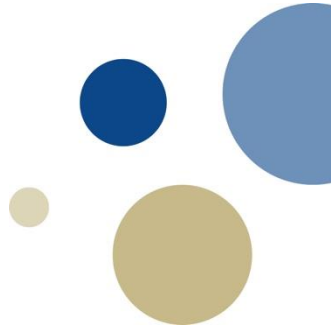
Boosting profits from an increasing number of models while **Bagging** stagnate relatively fast.

🔍Pros&Cons

Pros&Cons

- May increase accuracy
- Reduce variance / avoid overfitting
- Use simple models to create great results

- Does not work well on stable model
- May exclude data while **Boosting**



👨💻Coding example

References

Look it up

- Ensemble Learning:
  - https://scikit-learn.org/stable/modules/ensemble.html#b1996
  - https://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf
  - https://machinelearningmastery.com/ensemble-methods-for-deep-learning-neural-networks/
- Bagging:
  - https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/
  - https://link.springer.com/article/10.1023/A:1018054314350#article-info
- Boosting:
  - https://www.stat.berkeley.edu/~breiman/arcing-the-edge.pdf
  - https://cseweb.ucsd.edu/~yfreund/papers/boostingexperiments.pdf
- Both:
  - https://arxiv.org/pdf/1905.12787