

Boosting



Week 7 - Day 2



Linear Regression



Gradient Boosting

<https://www.youtube.com/watch?v=5CWwwtEM2TA>

Recap on Bagging vs. Boosting

Weak Classifiers + Aggregations

Weak Classifier

=

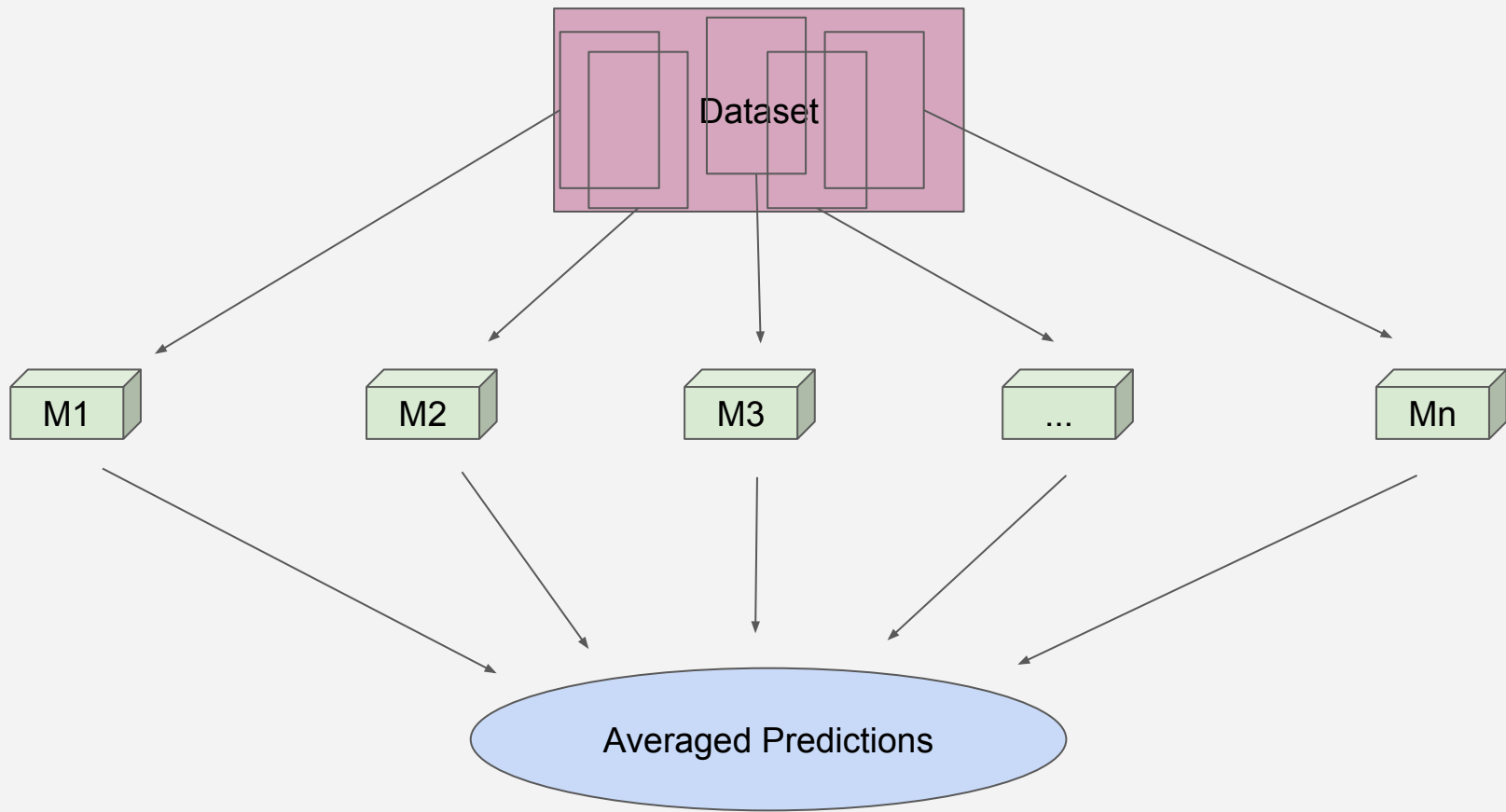
Slightly better than random

1 . Bagging

Bagging

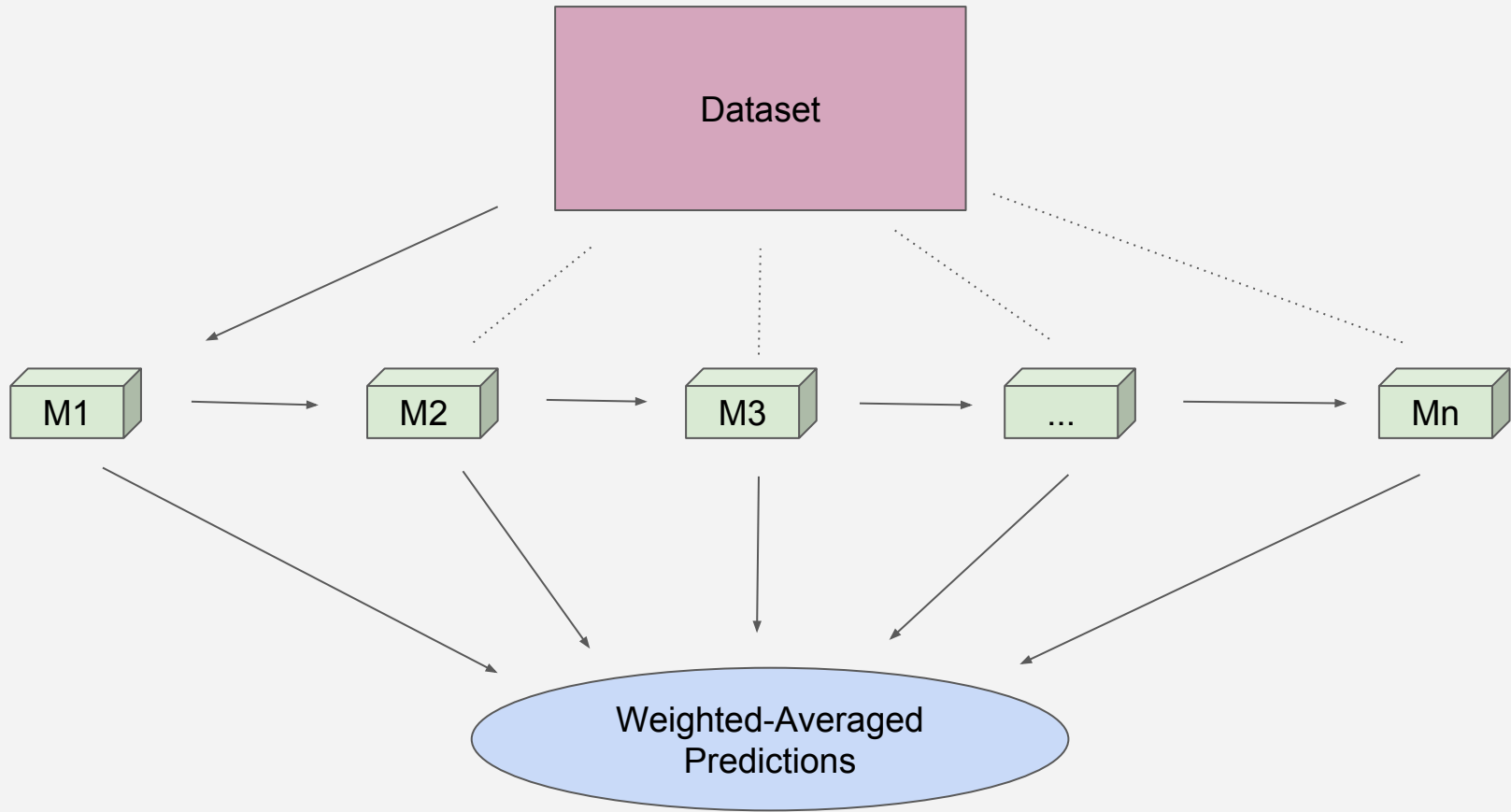
=

Bootstrap + Aggregating



E.g. Random Forest

2. Boosting



E.g. AdaBoost, Gradient Boost

AdaBoost

1. Model building phase

1. Create a simple model
2. Calculate predictions
3. Give weights to the points
(wrong predictions = large weights)
4. Repeat

2. Prediction phase

1. Get a prediction for every model
2. Calculate the weighted average of the predictions (bad models = small weights)

[https://www.youtube.com/watch?v=ix6l
vwbVpw0](https://www.youtube.com/watch?v=ix6lvwbVpw0)

Gradient Boost

Boosting + Gradient Descent

Grad. Boost >> AdaBoost

1. Create a simple model
2. Calculate predictions
3. Calculate the residuals
4. Build a model on the residuals
5. Repeat 2,3,4

<https://www.youtube.com/watch?v=sRktKszFmSk>

Pros/Cons

+ Performances

- Black box

- Complex+slow to tune

Libraries

1. XGBoost (sklearn API)
2. LightGBM (from Microsoft)
3. CatBoost (from Yandex)

Parameter tuning in XGBoost