

Стекинг

Гущин Александр
Осенний семестр 2015 года

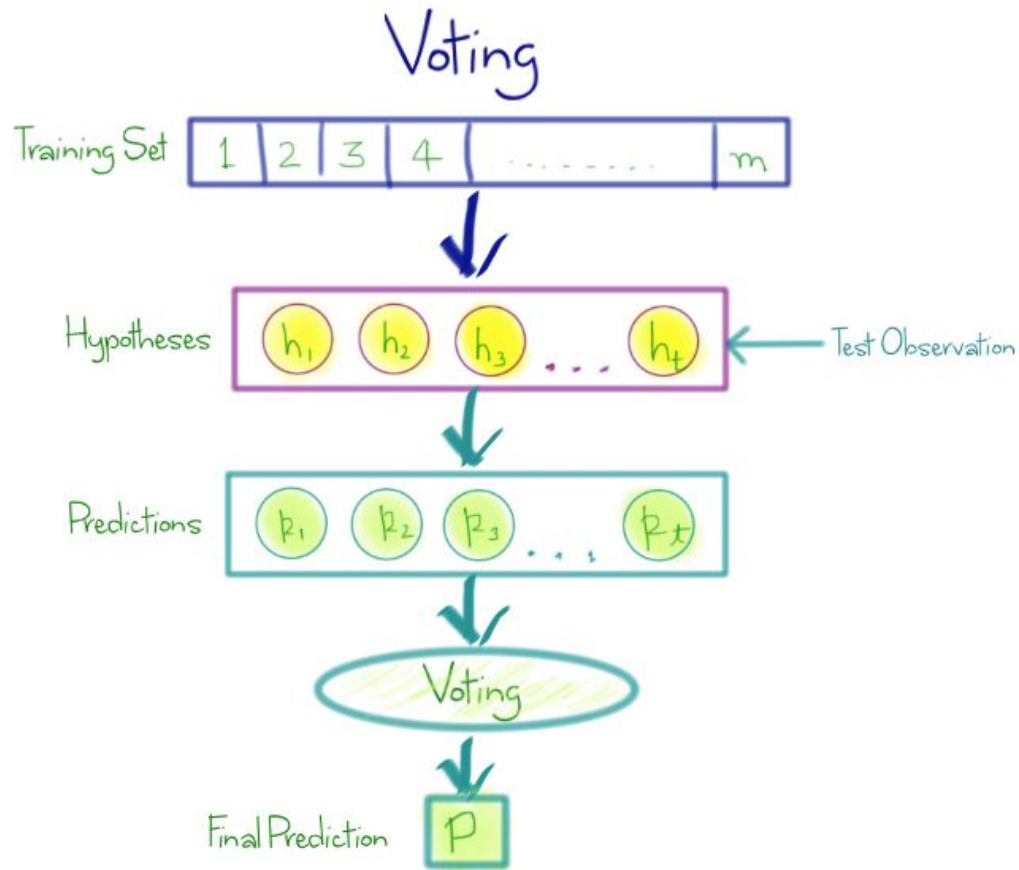
Голосование (Voting)

$$p = (p_1 + p_2 + \dots + p_n) / n$$

Взвешенное голосование

$$p = p_1 * w_1 + p_2 * w_2 + \dots + p_n * w_n$$

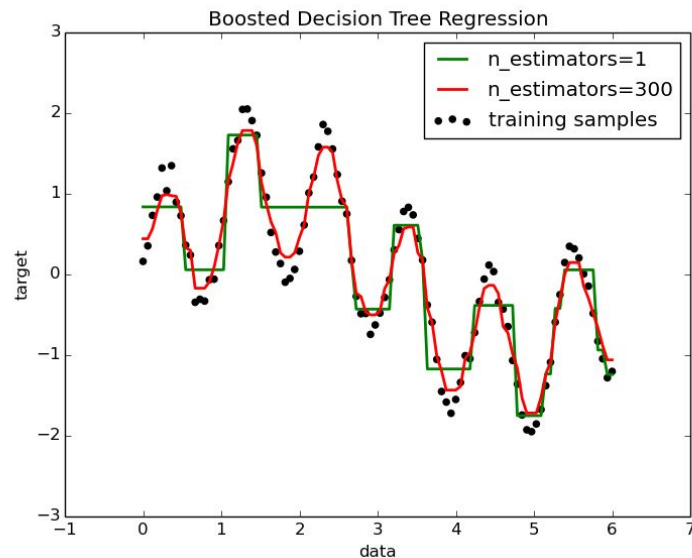
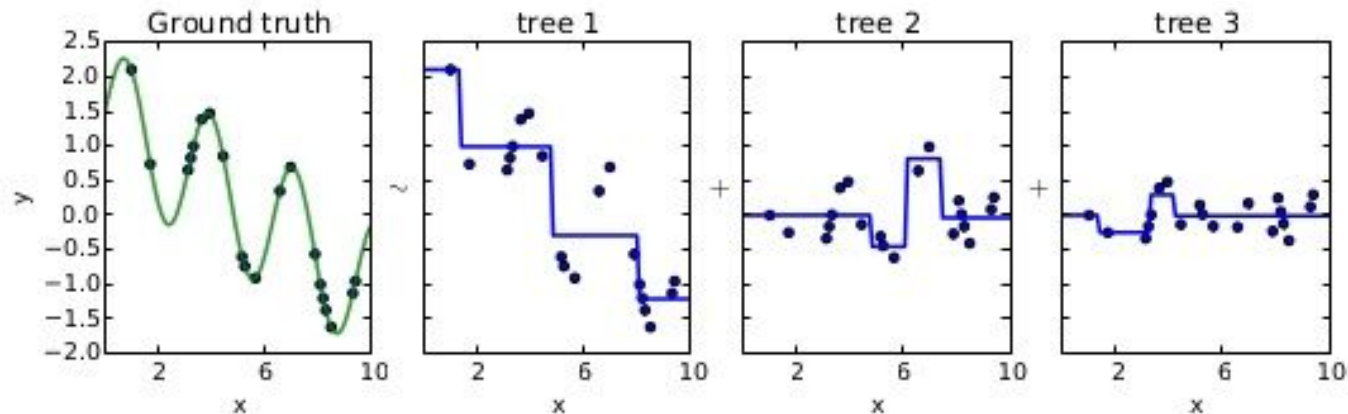
$$w_1 + w_2 + \dots + w_n = 1$$



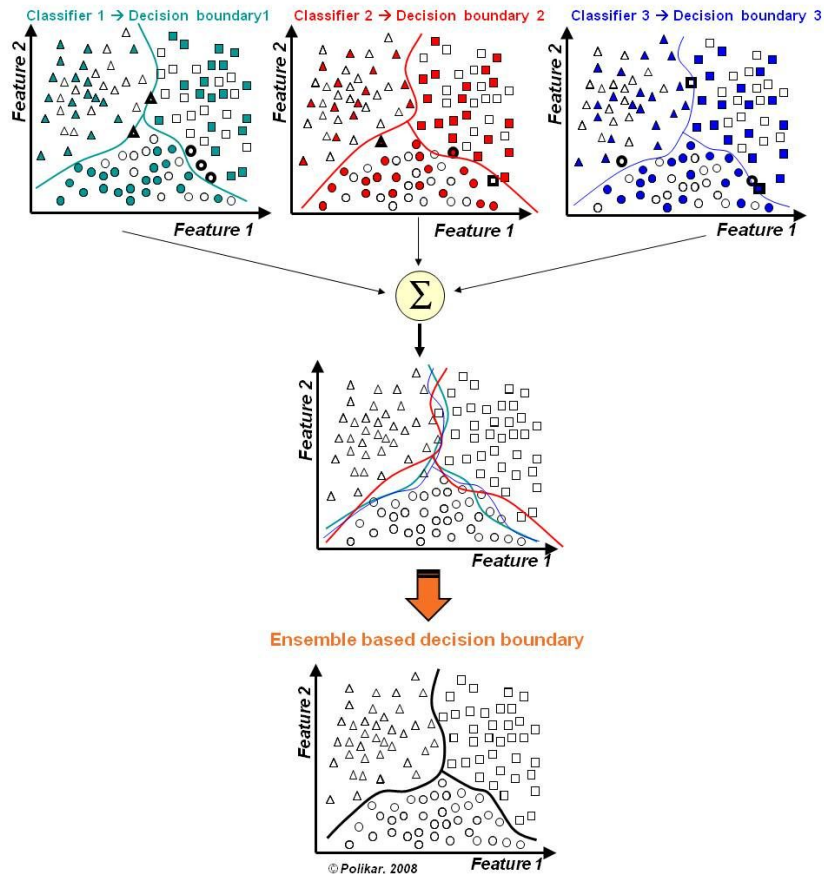
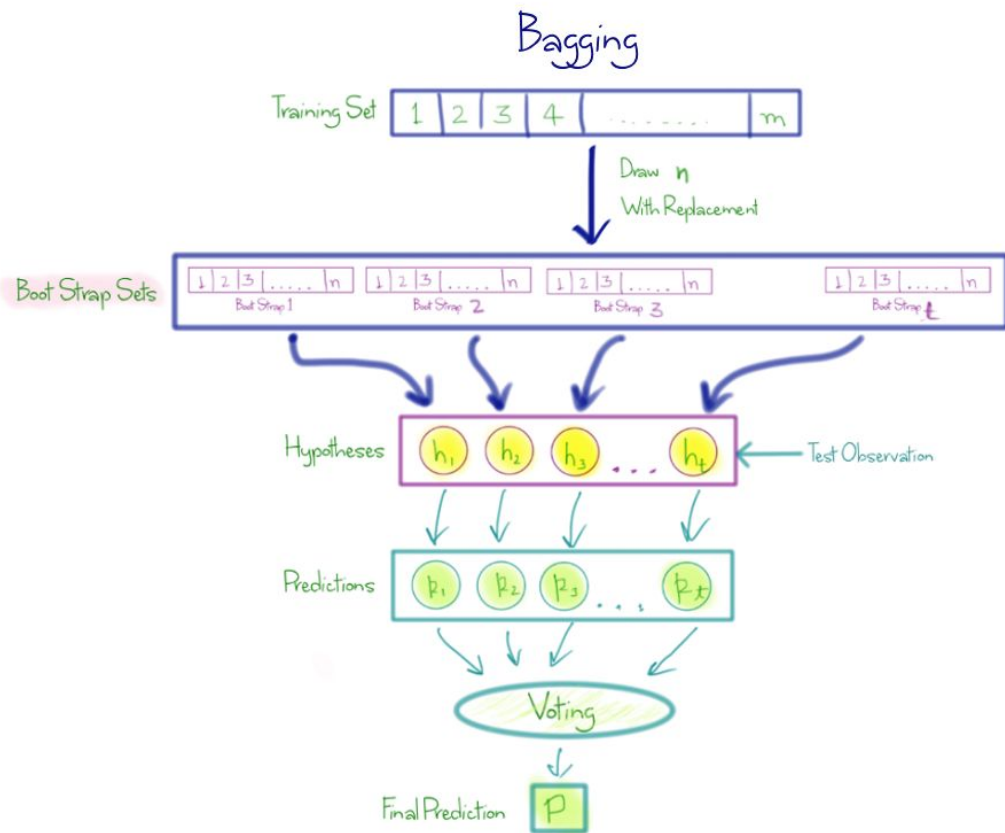
Boosting

Задача регрессии

Residual fitting



Bagging (Bootstrap aggregating)



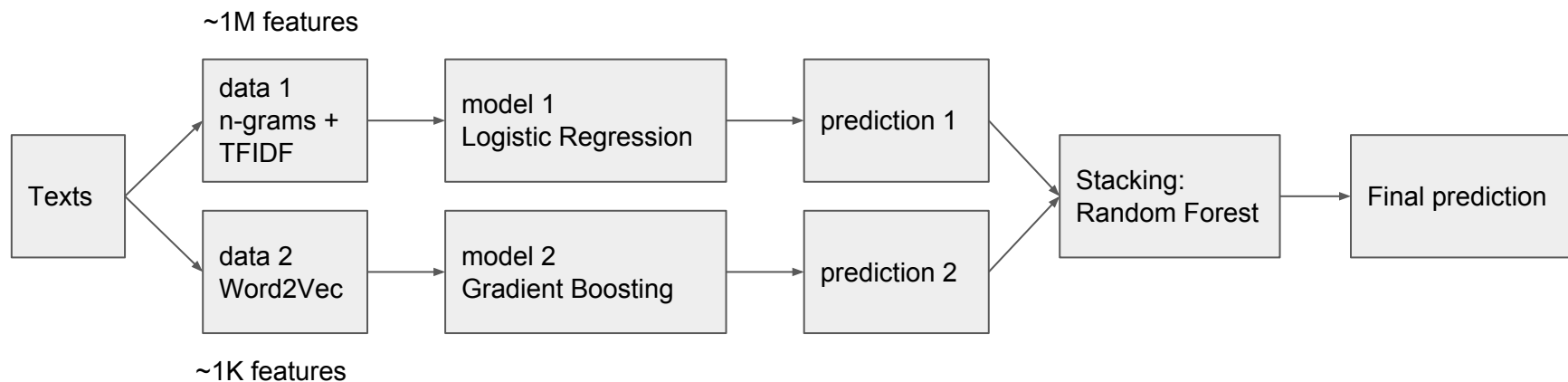
Вопросы

Почему это работает?

Как ансамблировать предсказания алгоритмов ещё эффективнее?

Классификация тестов

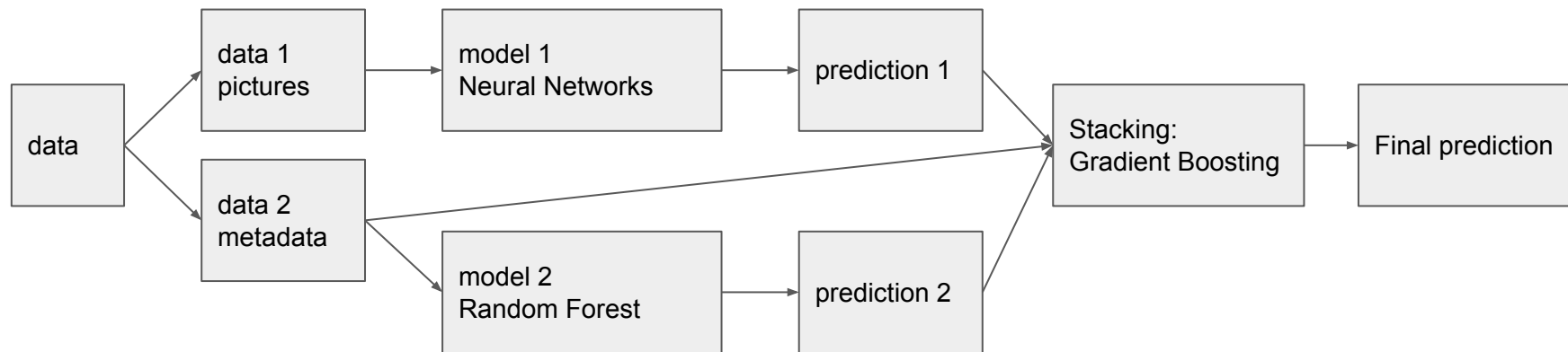
МНОГО КЛАССОВ



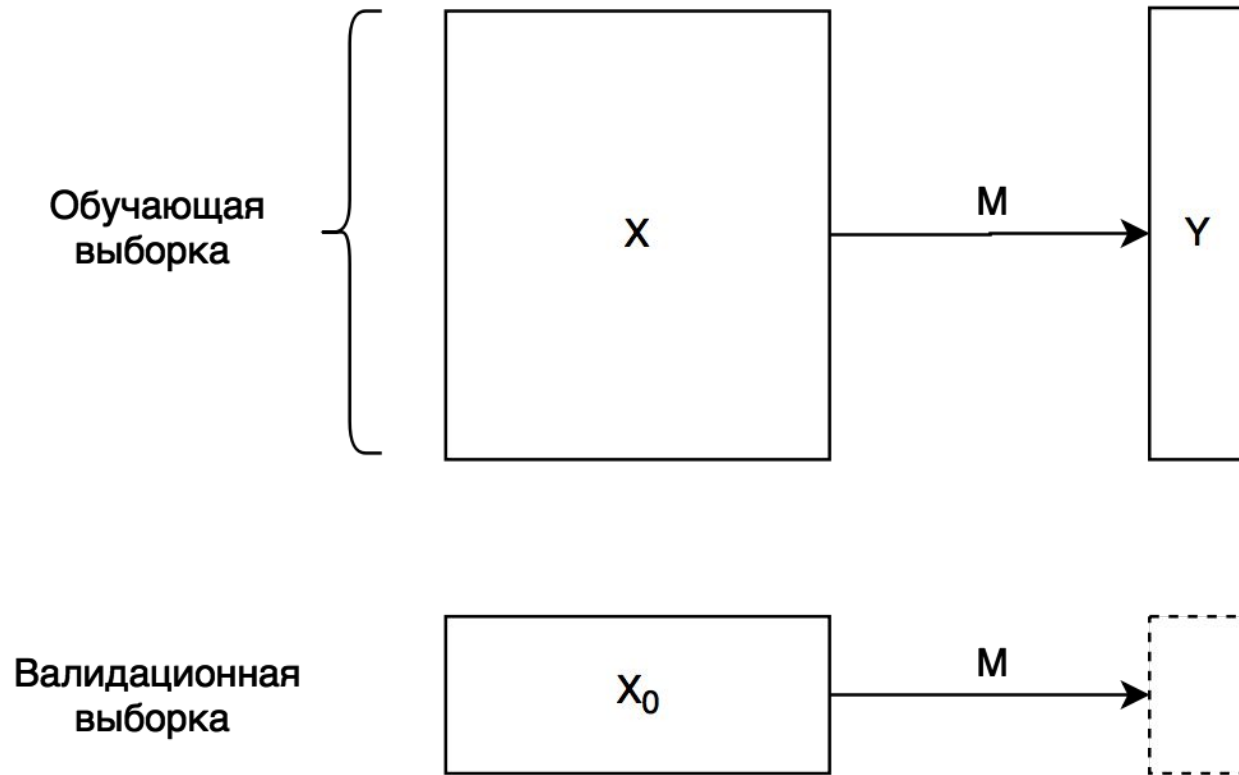
Картинки

Классификация: что изображено на картинке?

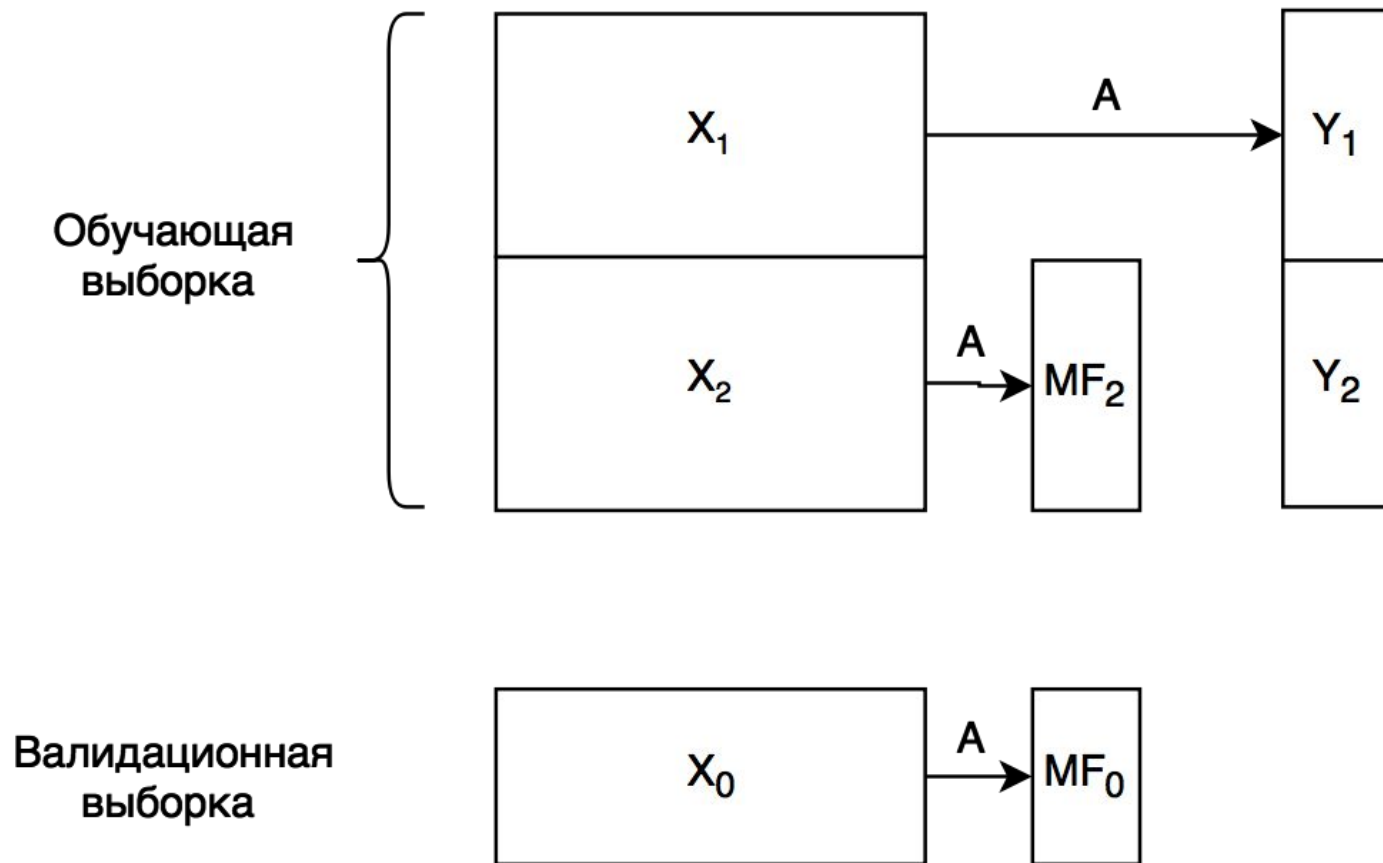
Регрессия: какую оценку получит картинка в социальных сетях?



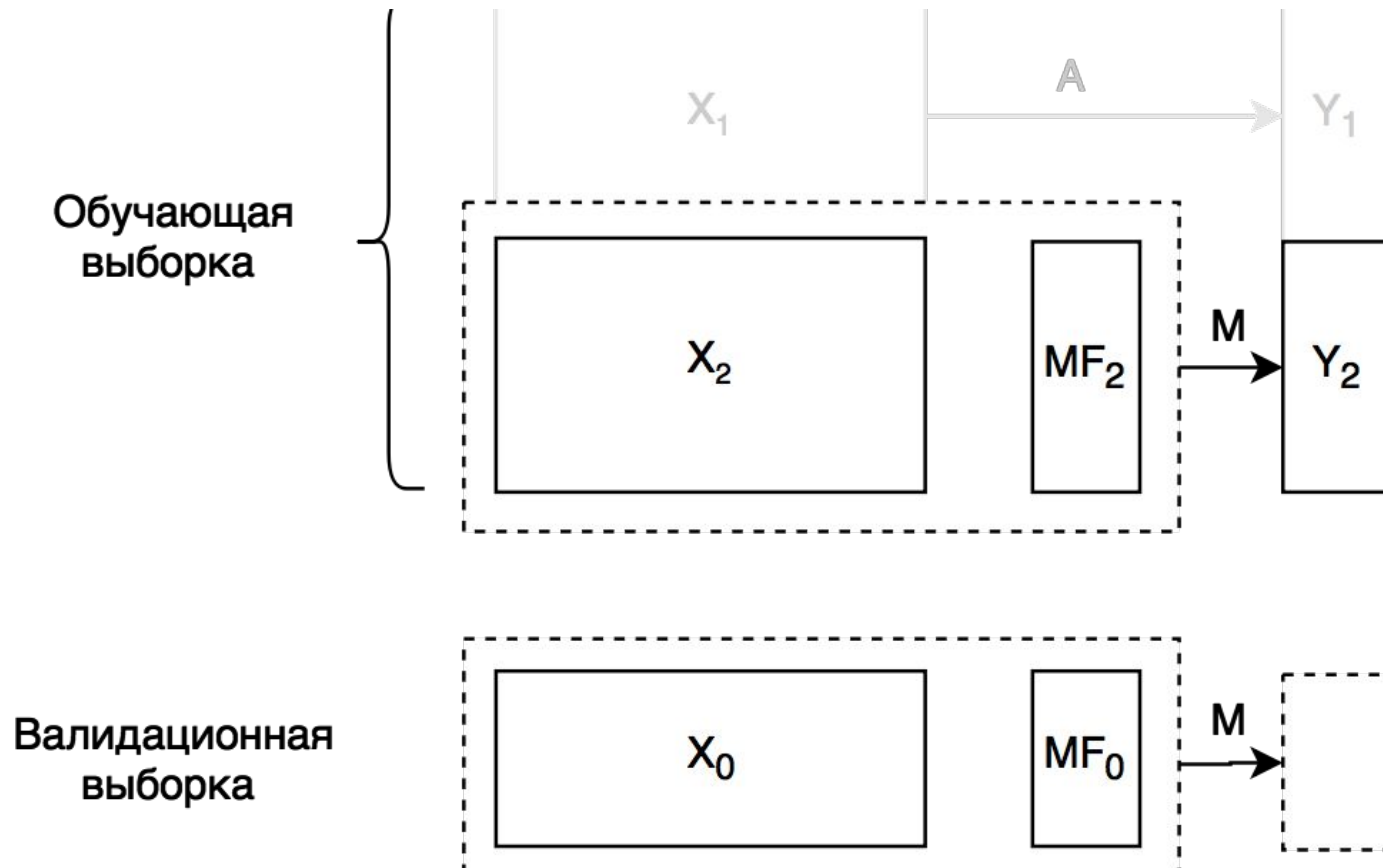
Обычная задача обучения с учителем



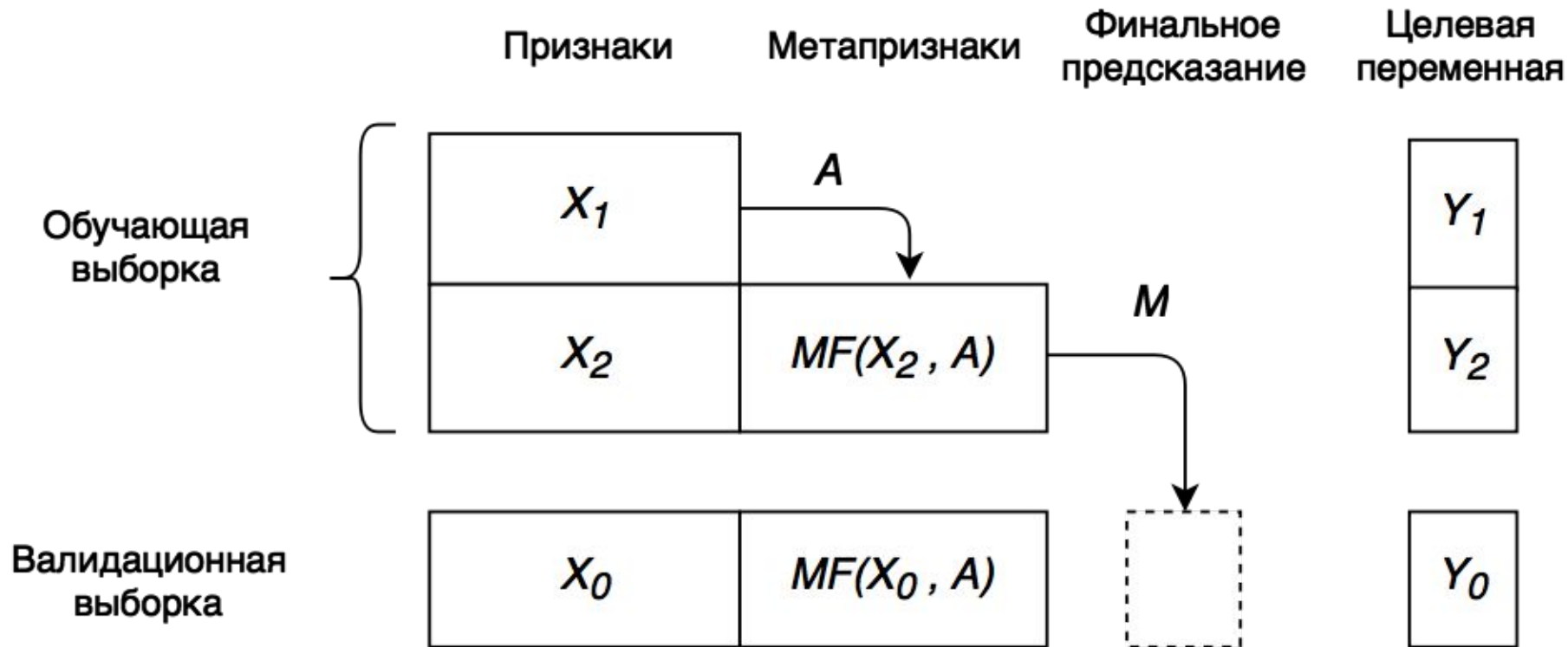
Стекинг - общий принцип (1)



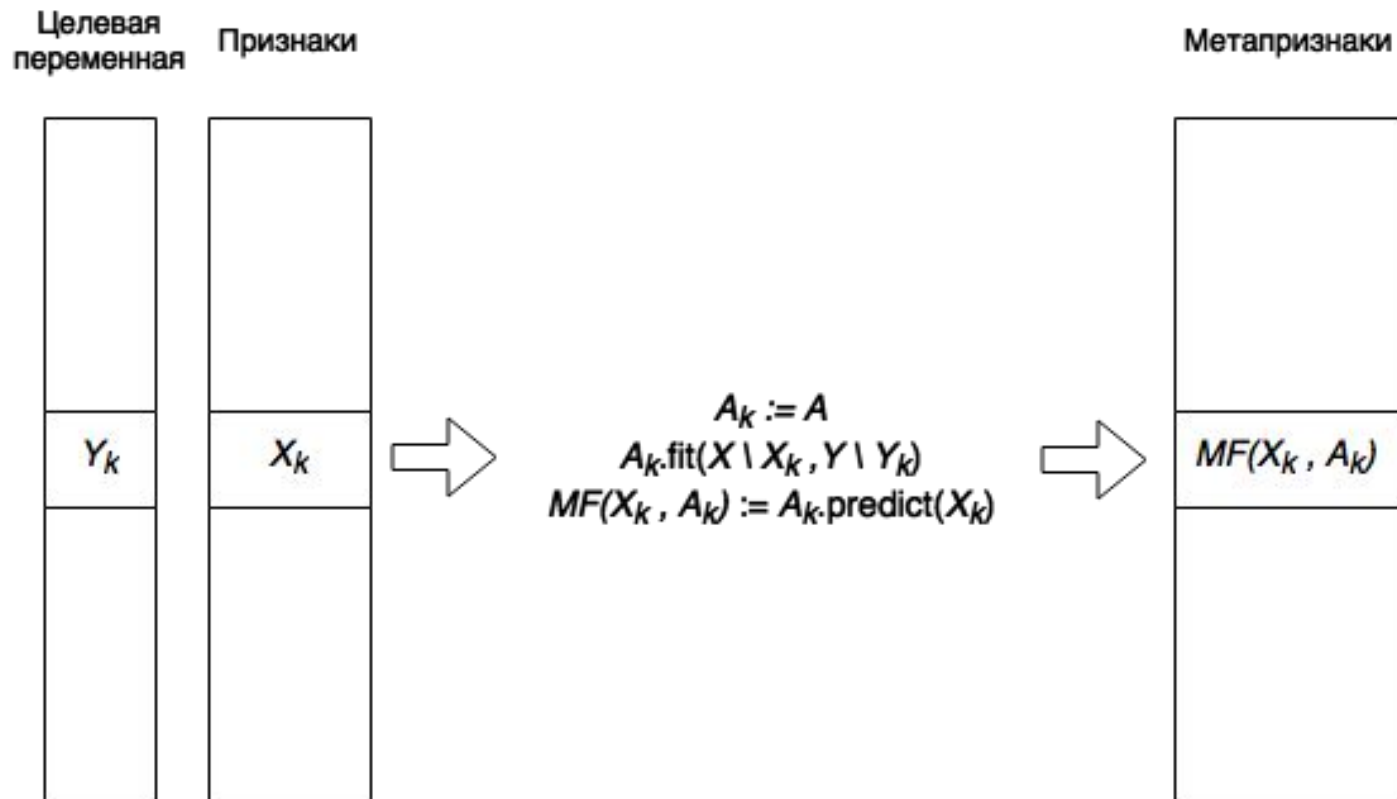
Стекинг - общий принцип (2)



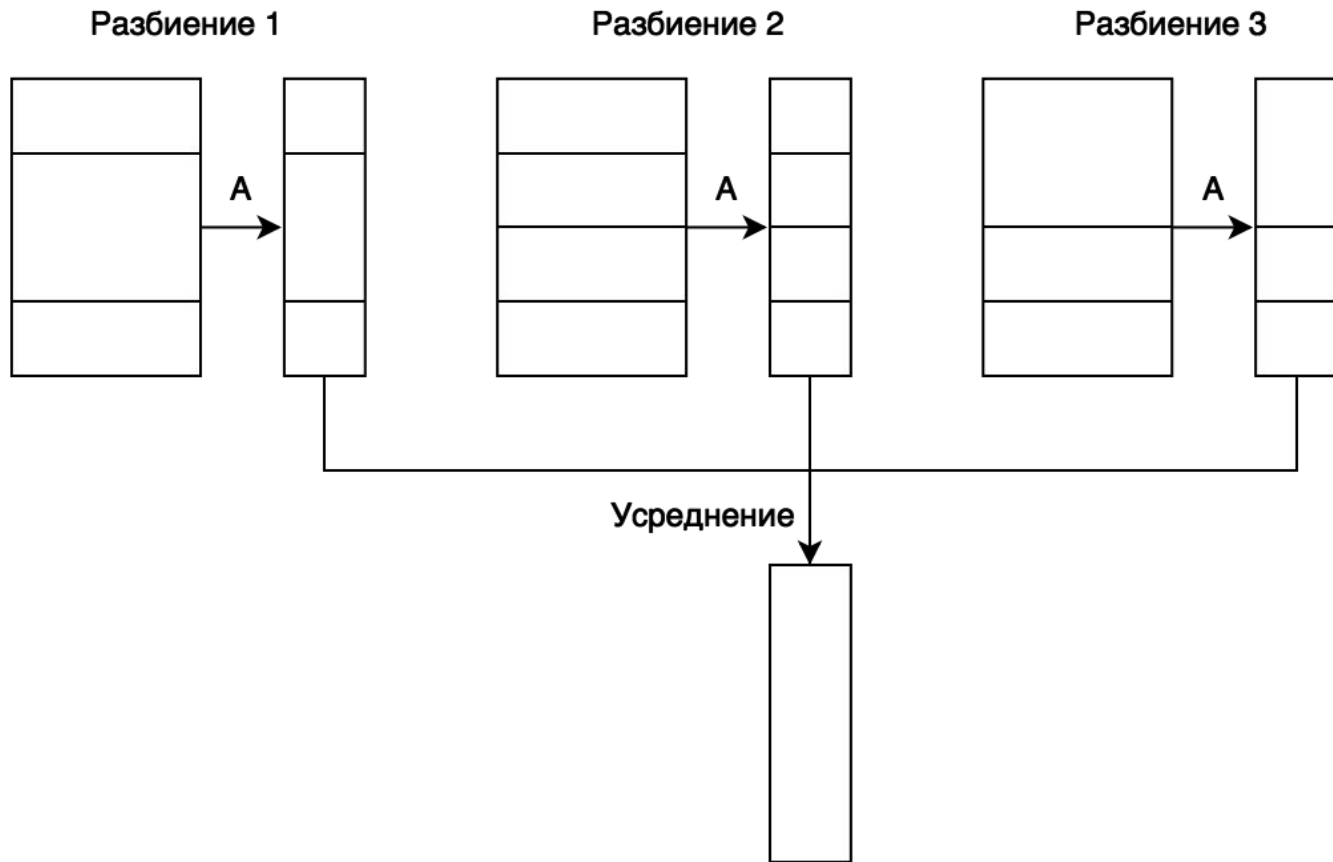
Стекинг по принципу Holdout (Blending)



Получение метапризнака

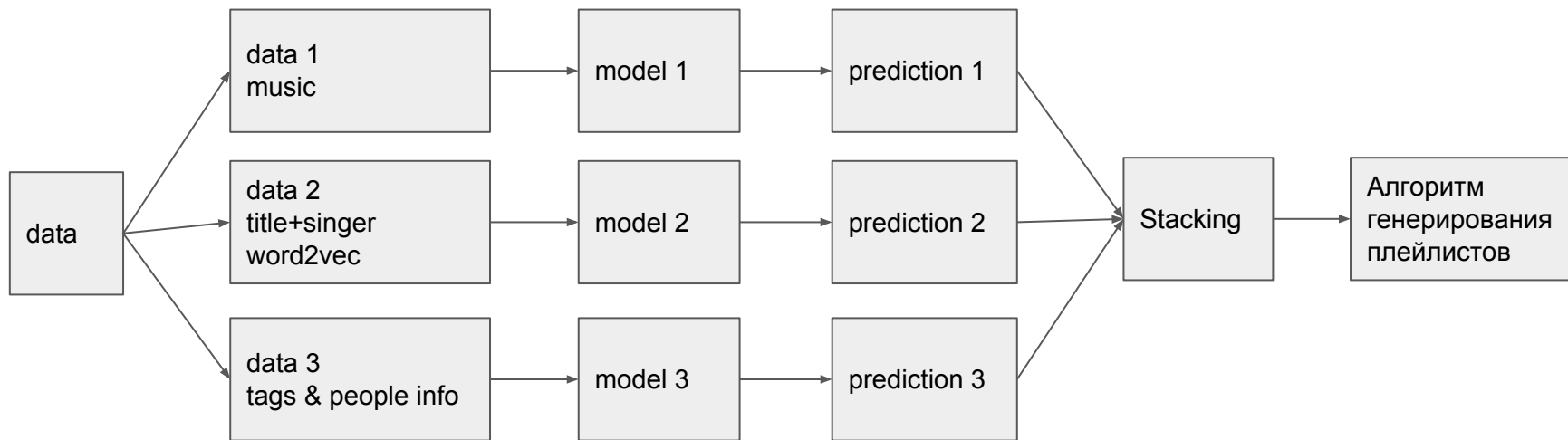


Получение метапризнака с усреднениями



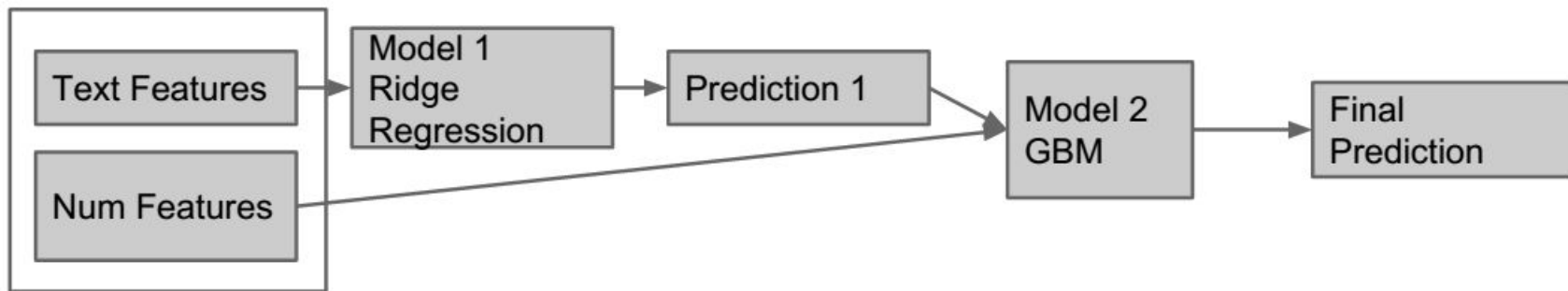
Составление плейлистов

Задача: генерировать плейлист по данной песне



Technical Tricks -- Stacking

- Basic idea -- use one model's output as the next model's input



- It is NOT a good idea to use in sample prediction for stacking
 - The problem is over-fitting
 - The more “over-fit” prediction1 is , the more weight it will get in Model 2

Материалы по ансамблям и стекингу

Tips Tricks and Use-cases of Ensembling in Practice by Misha Trofimov

<http://www.slideshare.net/MoscowDataFest/df1-dmc-trophimov-tips-tricks-and-usecases-of-ensembling-in-practice>

Kaggle ensembling guide by Triskellion

<http://mlwave.com/kaggle-ensembling-guide/>

Tips for data science competitions by [Owen Zhang](#)

<http://www.slideshare.net/OwenZhang2/tips-for-data-science-competitions>

Стекинг с усреднениями:

<https://github.com/aguschin/Guschin2015Stacking>

Использованные картинки

Классные картинки:

<http://manish-m.com/?p=794>

Boosting:

http://scikit-learn.org/stable/auto_examples/ensemble/plot_adaboost_regression.html

<http://www.slideshare.net/DataRobot/gradient-boosted-regression-trees-in-scikitlearn>

Bagging

http://www.scholarpedia.org/article/Ensemble_learning