

# Лекция 10

## Ансамбли моделей

Габдуллин Р.А., Макаренко В.А.

МГУ им. М.В. Ломоносова

16 марта 2021

# Ансамбли моделей

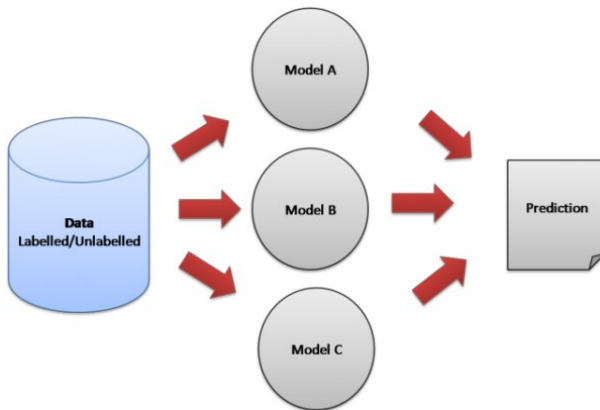


Рис.: Источник: [kdnuggets.com](https://www.kdnuggets.com)

Для построения прогноза можно использовать сразу несколько моделей.

# Бэггинг (Bootstrap aggregating)

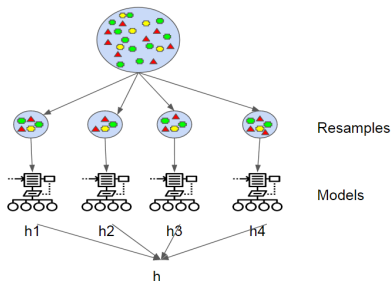


Рис.: Источник: [hackernoon.com](https://hackernoon.com)

Пусть  $(x_1, y_1), \dots, (x_\ell, y_\ell)$  – обучающая выборка.

- Строим  $N$  выборок длины  $\ell$ , сэмплируя с возвращением данные из тренировочной выборки.
- Обучаем модель на каждой из выборок.
- Усредняем прогнозы (для регрессии) / берем моду (для классификации).

# Обучение мета-алгоритмов

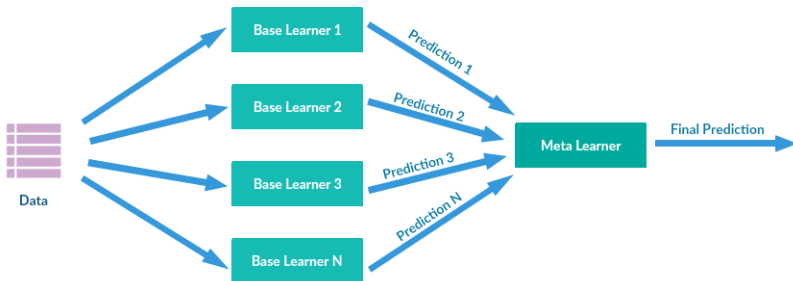


Рис.: Источник: [towardsdatascience.com](https://towardsdatascience.com)

# Блендинг (Blending)

Пусть  $(X, y)$  – тренировочная выборка,  $(X_{\text{test}}, y_{\text{test}})$  – тестовая.

- Разобьем тренировочную выборку на две части  $(X_{\text{train}}, y_{\text{train}})$  и  $(X_{\text{meta}}, y_{\text{meta}})$ .
- Обозначим  $(X_{\text{train}}, y_{\text{train}})$ ,  $(X_{\text{meta}}, y_{\text{meta}})$  и  $(X_{\text{test}}, y_{\text{test}})$  через  $A$ ,  $B$  и  $C$  соответственно.
- Обучим модели  $b_1(x), \dots, b_N(x)$  на  $A$ .
- Обучим модель  $a(x)$  на выборке  $B$ , используя в качестве признаков предсказания моделей  $b_1(x), \dots, b_n(x)$ .
- Оцениваем итоговое качество модели  $a(x)$  на выборке  $C$ .

# Блендинг (Blending)

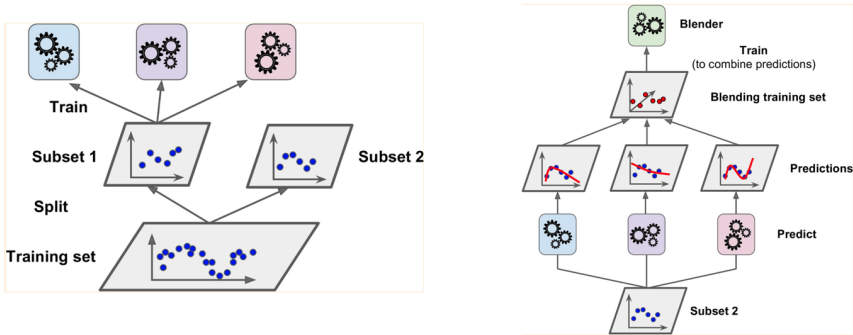


Рис.: Источник: [stats.stackexchange.com](https://stats.stackexchange.com)

# Стекинг (Stacking)

Пусть  $(X, y)$  – тренировочная выборка,  $(X_{\text{test}}, y_{\text{test}})$  – тестовая.

- Разобьем тренировочную выборку на  $K$  фолдов  $(X_{\text{train},i}, y_{\text{train},i})$ .
- Обозначим  $(X_{\text{train},i}, y_{\text{train},i})$  и  $(X_{\text{test}}, y_{\text{test}})$  через  $A_i$  и  $C$  соответственно.
- Для каждого фолда  $A_i$  обучаем модели  $b_1(x), \dots, b_N(x)$  на остальных фолдах, получаем предсказание на данном фолде.
- Обучаем модель  $a(x)$  на всей тренировочной выборке, используя в качестве признаков прогнозы, данные моделями  $b_1(x), \dots, b_N(x)$ .
- Обучаем модели  $b_1(x), \dots, b_N(x)$  на всей тренировочной выборке.
- Тестируем итоговый алгоритм на выборке  $C$ .

# Стекинг (Stacking)

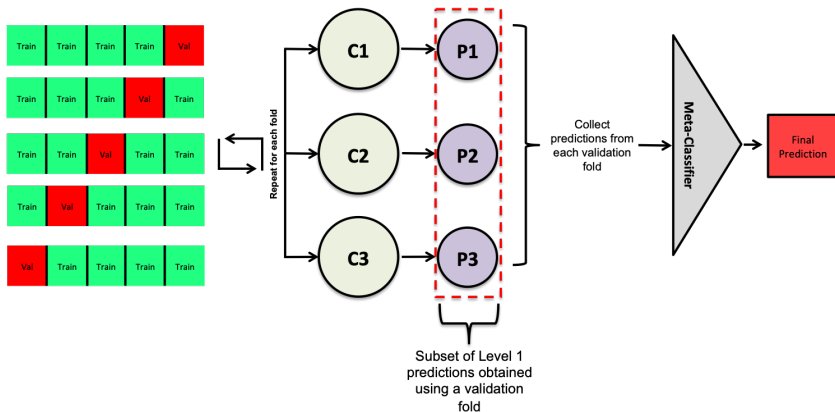


Рис.: Источник: [towardsdatascience.com](https://towardsdatascience.com)



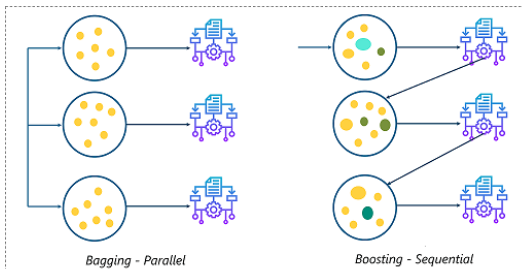
# Бустинг (Boosting)

Идея:

- Будем использовать  $N$  «слабых» моделей  $b_1(x), b_2(x), \dots, b_n(x)$  и с помощью них строить одну «сильную» модель в виде:

$$b(x) = \alpha_1(x)b_1(x) + \alpha_2(x)b_2(x) + \dots + \alpha_N(x)b_N(x).$$

- Процесс построения итеративный: каждая новая модель учится исправлять ошибки итогового алгоритма с прошлой итерации.



# Бустинг (Boosting)

Цель – построить алгоритм, минимизирующий эмпирический риск:

$$b(x) = \sum_{j=1}^N \alpha_j b_j(x),$$

$$\alpha_k, b_k = \operatorname{argmin}_{\alpha_k, b_k} \frac{1}{\ell} \sum_{i=1}^{\ell} L \left( y_i, \sum_{j=1}^N \alpha_j b_j(x) \right).$$

Итеративный процесс:

$$\alpha_m, b_m = \operatorname{argmin}_{\alpha_m, b_m} \frac{1}{\ell} \sum_{i=1}^{\ell} L \left( y_i, \sum_{j=1}^{m-1} \alpha_j b_j(x) + \alpha_m b_m(x) \right).$$

# Градиентный бустинг (Gradient boosting)

Итеративный процесс:

$$\alpha_m, b_m = \operatorname{argmin}_{\alpha_m, b_m} \frac{1}{\ell} \sum_{i=1}^{\ell} L \left( y_i, \hat{b}(x) + \alpha_m b_m(x) \right).$$

Приближим с помощью  $b_m(x)$  антиградиент функции потерь

$$b_m(x_i) \approx - \left. \frac{\partial L(y_i, z)}{\partial z} \right|_{z=\hat{b}(x_i)},$$

после чего решим задачу оптимизации

$$\alpha_m = \operatorname{argmin}_{\alpha_m} \frac{1}{\ell} \sum_{i=1}^{\ell} L \left( y_i, \hat{b}(x) + \alpha_m b_m(x_i) \right).$$

Таким образом, эмпирический риск будет убывать с каждой итерацией.