Boosting,
stacking and blending,
Bias-variance tradeoff

Вспоминаем

Likelihood

MSE

МНК

Linear regression

Logistic regression

Decision trees

Random forest

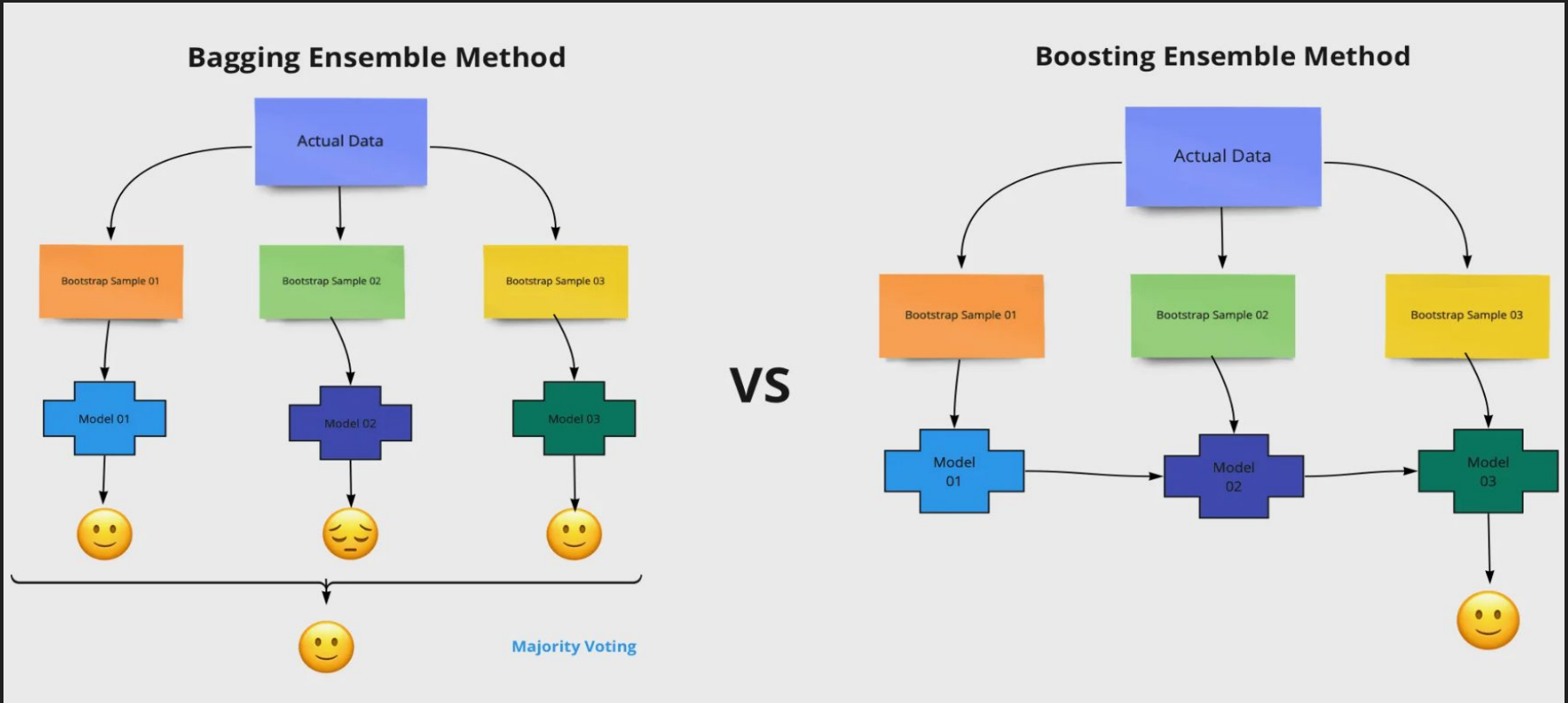
Bagging

Классификация, регрессия

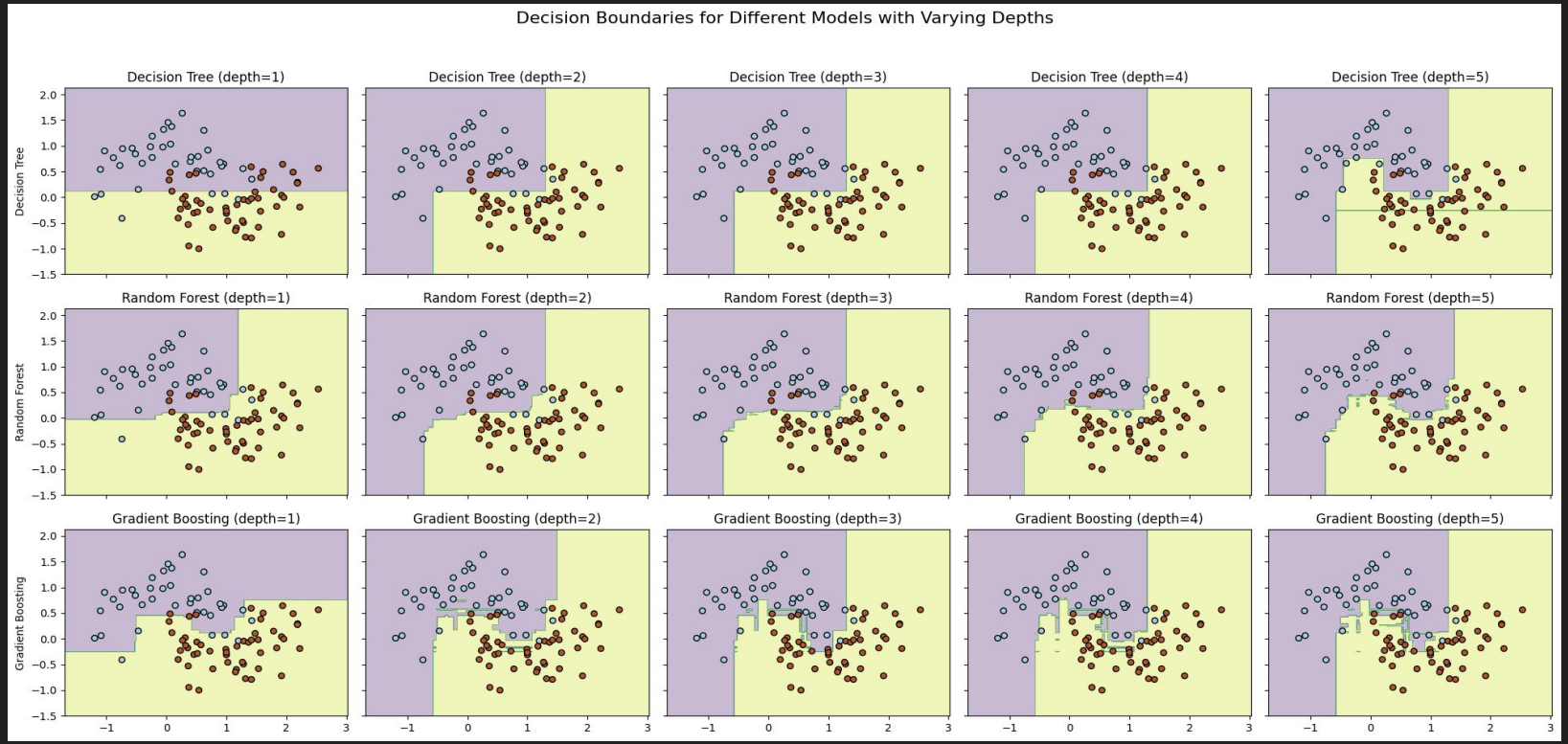
One-hot encoding, кодирование переменных

Out of bag

Boosting


Bagging Ensemble Method
Actual Data
Bootstrap Sample 01
Bootstrap Sample 02
Bootstrap Sample 03
Model 01
Model 02
Model 03
Majority Voting
VS
Boosting Ensemble Method
Actual Data
Bootstrap Sample 01
Bootstrap Sample 02
Bootstrap Sample 03
Model 01
Model 02
Model 03

Итак, бустинг - ансамбль, который улучшает точность модели за счет последовательного обучения нескольких слабых моделей, каждая из которых исправляет ошибки предыдущих



Какие бустинги бывают

Adaboost - первая версия бустинга

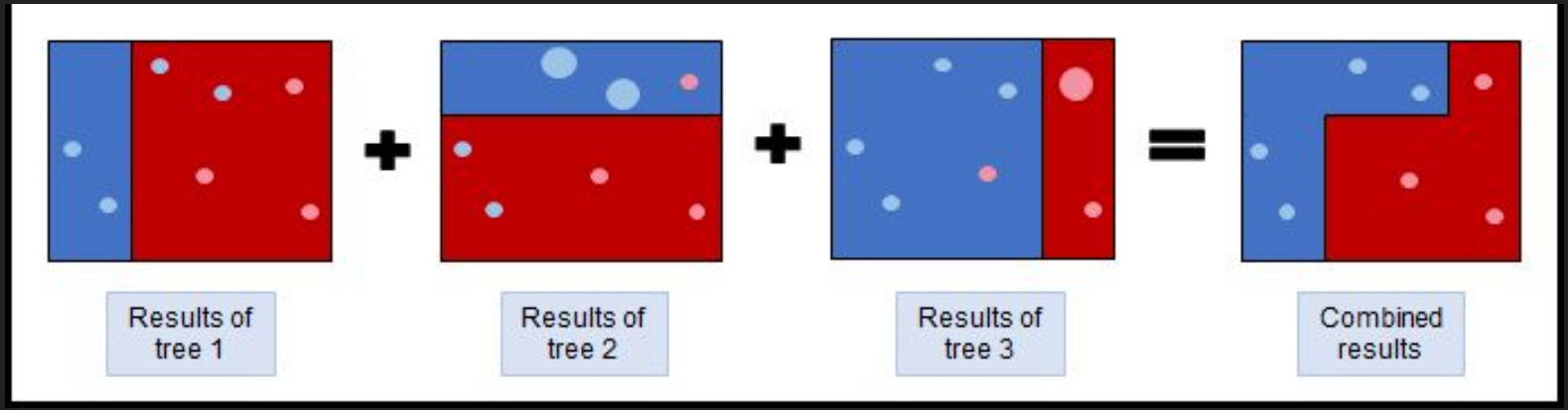
Gradient boosting - папа всех современных бустингов

Современные бустинги:

- Catboost
- LightGBM
- XGBoost

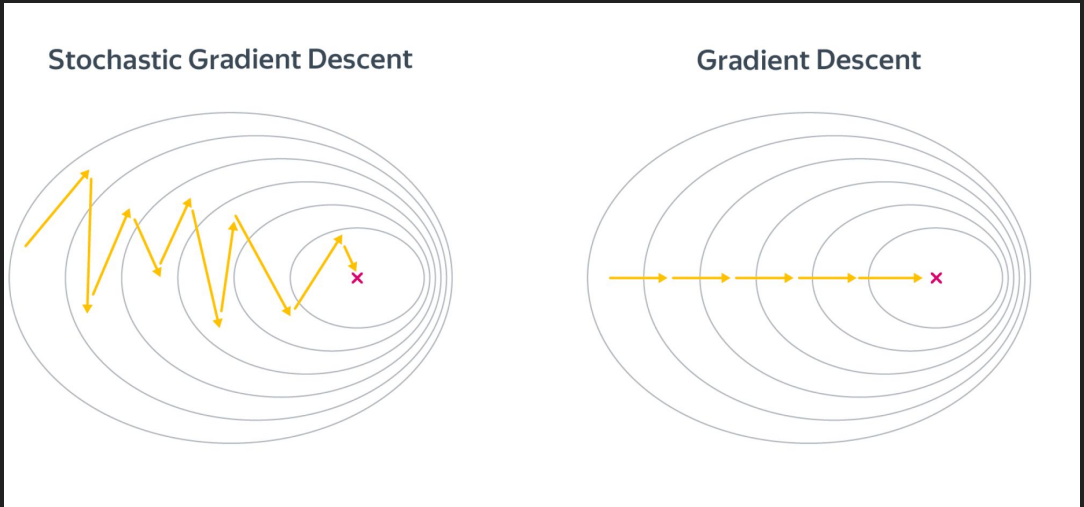
Adaboost (Adaptive Boosting)

В AdaBoost ошибки предыдущих моделей влияют на веса объектов, делая их более значимыми для следующей модели. На каждом шаге веса корректируются, чтобы уменьшить ошибку на более сложных для классификации примерах.



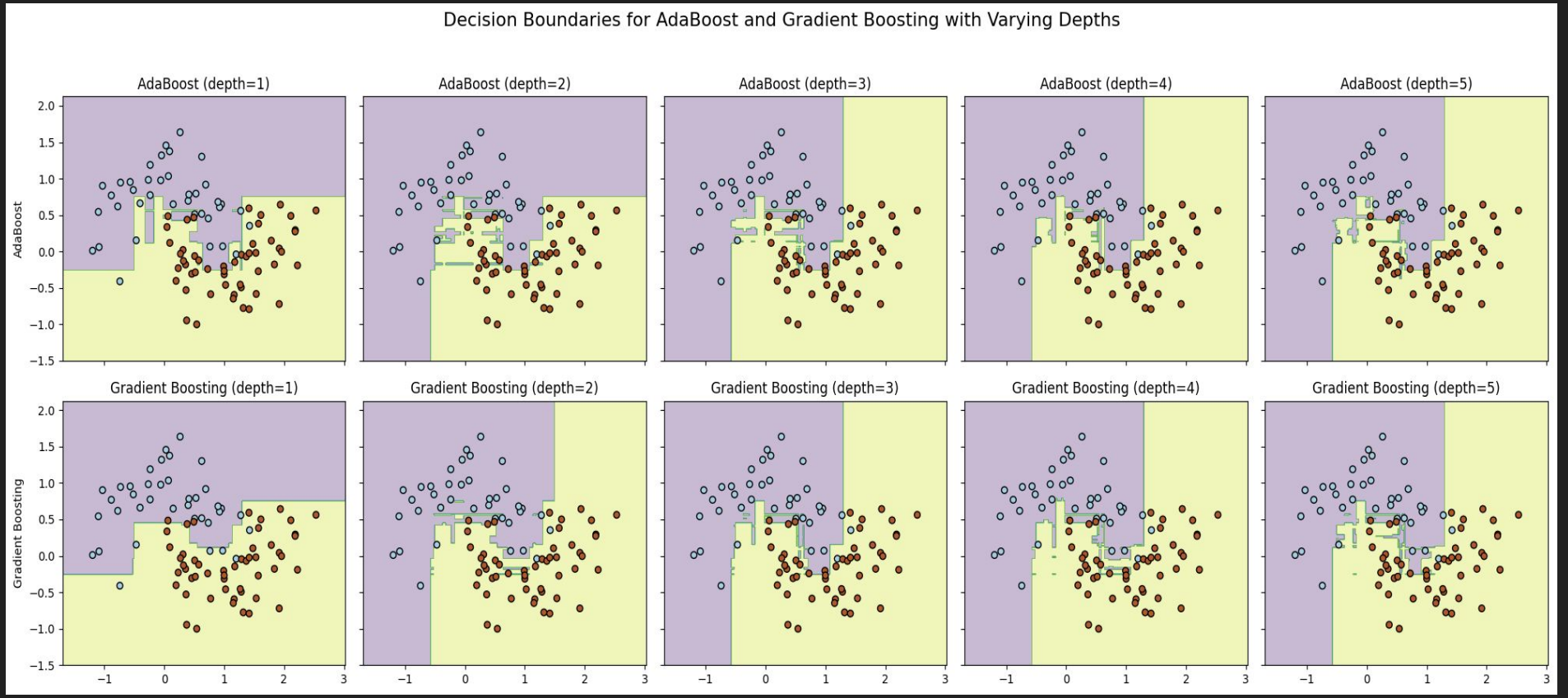
Gradient boosting

1. Инициализация ансамбля с базовым прогнозом
2. Вычисление остаточной ошибки
3. Градиент как направление исправления ошибки
4. Обучение новой модели на градиенте
5. Добавление нового прогноза с регулируемым шагом обучения
6. Повторение и постепенное улучшение



Ловушки градиентного спуска (в контексте деревянного бустинга)

1. **Количество моделей (n_estimators):** определяет, сколько моделей добавится в ансамбль. Чем больше шагов, тем точнее ансамбль, но слишком большое количество шагов может привести к переобучению.
2. **Глубина деревьев (max_depth):** неглубокие деревья предотвращают переобучение, поскольку сосредотачиваются на исправлении конкретных ошибок, не подстраиваясь под каждый отдельный пример.
3. **Скорость обучения (learning rate):** определяет, насколько сильно новые модели корректируют общий прогноз. Малый шаг требует большего количества моделей, но снижает риск переобучения.


Decision Boundaries for AdaBoost and Gradient Boosting with Varying Depths
AdaBoost (depth=1)
AdaBoost (depth=2)
AdaBoost (depth=3)
AdaBoost (depth=4)
AdaBoost (depth=5)
Gradient Boosting (depth=1)
Gradient Boosting (depth=2)
Gradient Boosting (depth=3)
Gradient Boosting (depth=4)
Gradient Boosting (depth=5)
AdaBoost
Gradient Boosting

Adaboost

- Присваивает бОльшие веса для ошибок на предыдущем шаге
- Чувствителен к выбросам и шуму
- будет фокусироваться на трудных примерах и будет иметь более резкие изменения в решающих плоскостях

Gradient boosting

- Строит каждое дерево таким образом, чтобы минимизировать ошибку предыдущего дерева с помощью градиентного спуска.
- Менее чувствителен к шуму
- будет плавно улучшать модель, минимизируя ошибку на каждом шаге и будет иметь более гладкие и менее резкие разделительные линии.

Можем ли мы строить градиентный бустинги над чем-то другим?

Можем! … но иногда возникает вопро - зачем?)

Почему?

Что такое линейная комбинация линейный моделей?

Это линейная модель

Современные градиентные бустинги

Catboost - Модель от Yandex оптимизирован для категориальных переменных и вообще много что умеет под капотом

LightGBM - Довольно быстрая модель от Microsoft

XGBoost - Еще один вариант бустинга до кучи

Blending

Берем идею от бэггинга, но делаем взвешенное голосование

Как их найти?

Откладываем часть выборки и обучаем еще одну модель для объединения результатов предыдущих

Blending

Плюсы:

- получаем более объективное голосование (в отличие от бейзлайна, где все модели вносят одинаковый вклад)
- Прост в применении

Минусы:

- Мы должны отдавать самое ценное - данные
- Линейная комбинация линейных моделей - это линейная модель(

Stacking

Схож с cross-validation (K-fold)

- Берем модель и обучаем её на одной части выборки, не показывая ей другую часть
- Получаем предсказания для той части, которую мы скрыли
- Profit - мы предобработали данные за счет модели, а не ручками и получили новые признаки
- Обучаем на этих признаках новую модель

Stacking

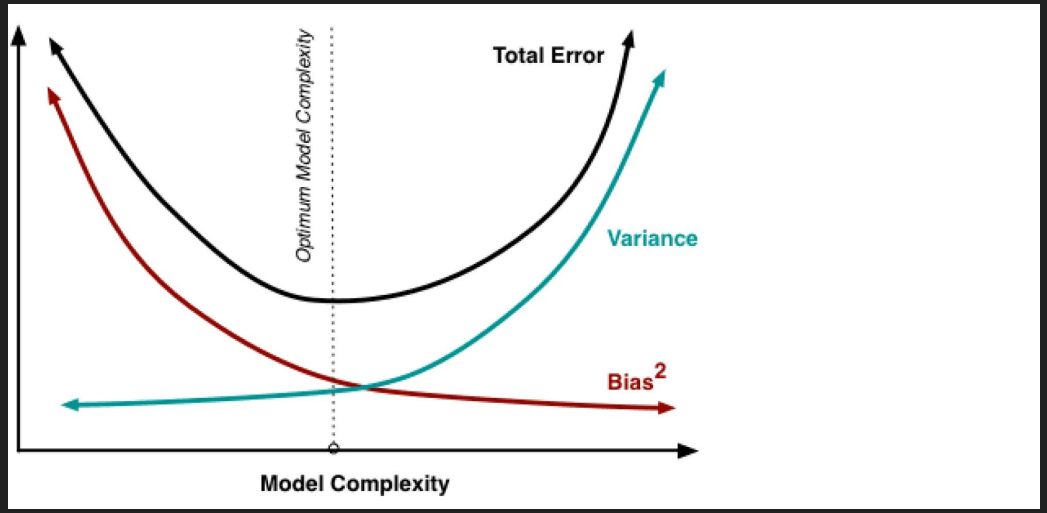
Плюсы:

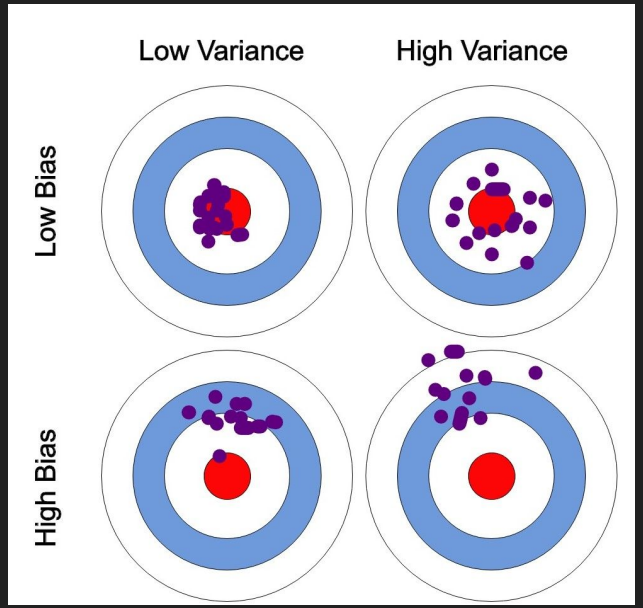
- Можем делать это бесконечно
- Хорош для ансамблей
- Спойлер: это любая нейронка с FC слоем в конце

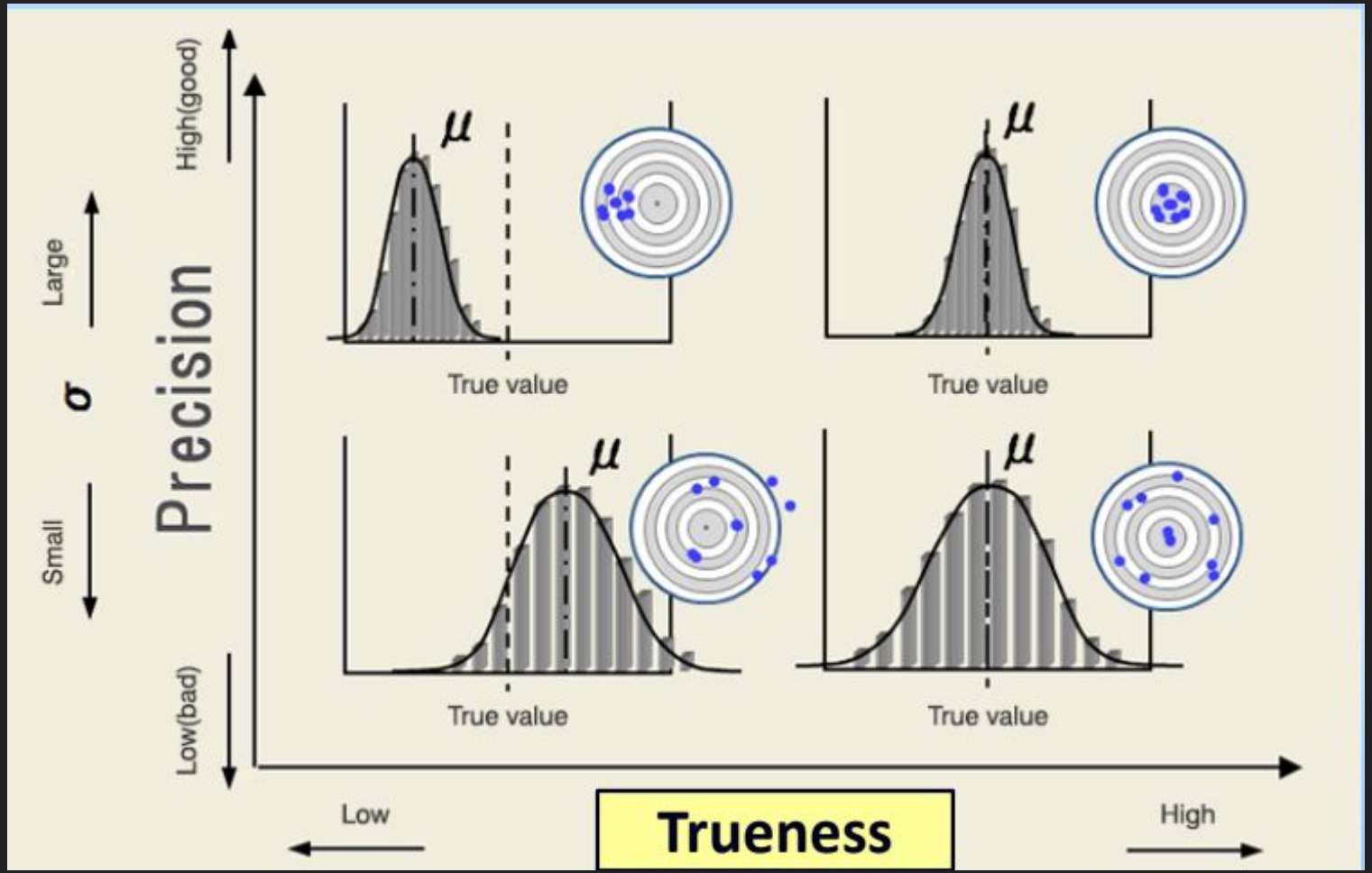
Минусы:

- Теряем интерпретируемость модели
- Нестабильность за счет разнородности признаков

Bias-variance tradeoff






Precision
High(good)
Low(bad)
Large
σ
Small
μ
True value
μ
True value
μ
True value
μ
True value
Low
Trueness
High

Методы оценки моделей

Cross-validation

Leave-one-out

Bootstrap

Отложенная выборка