

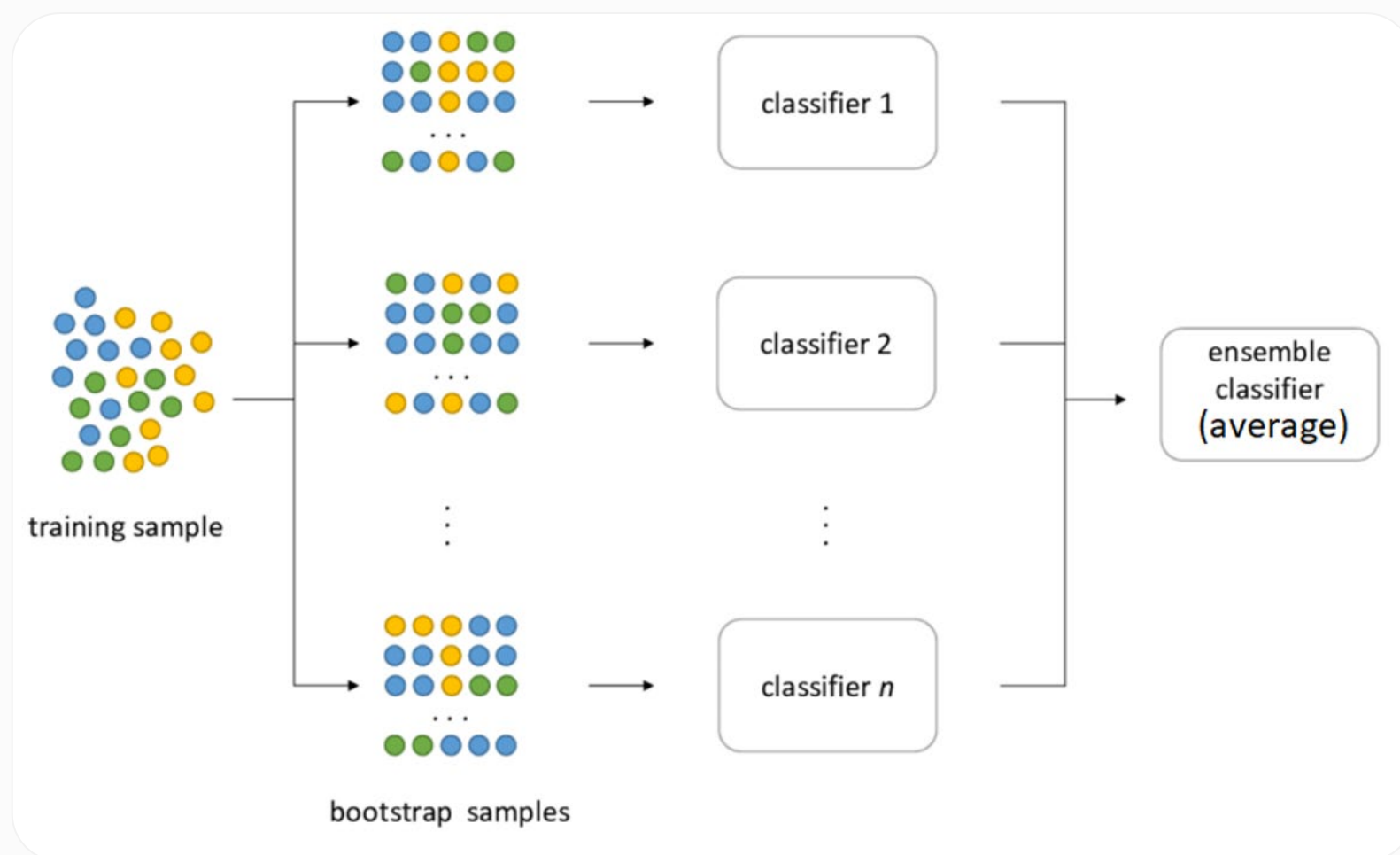
Случайный лес

Елена Кантонистова

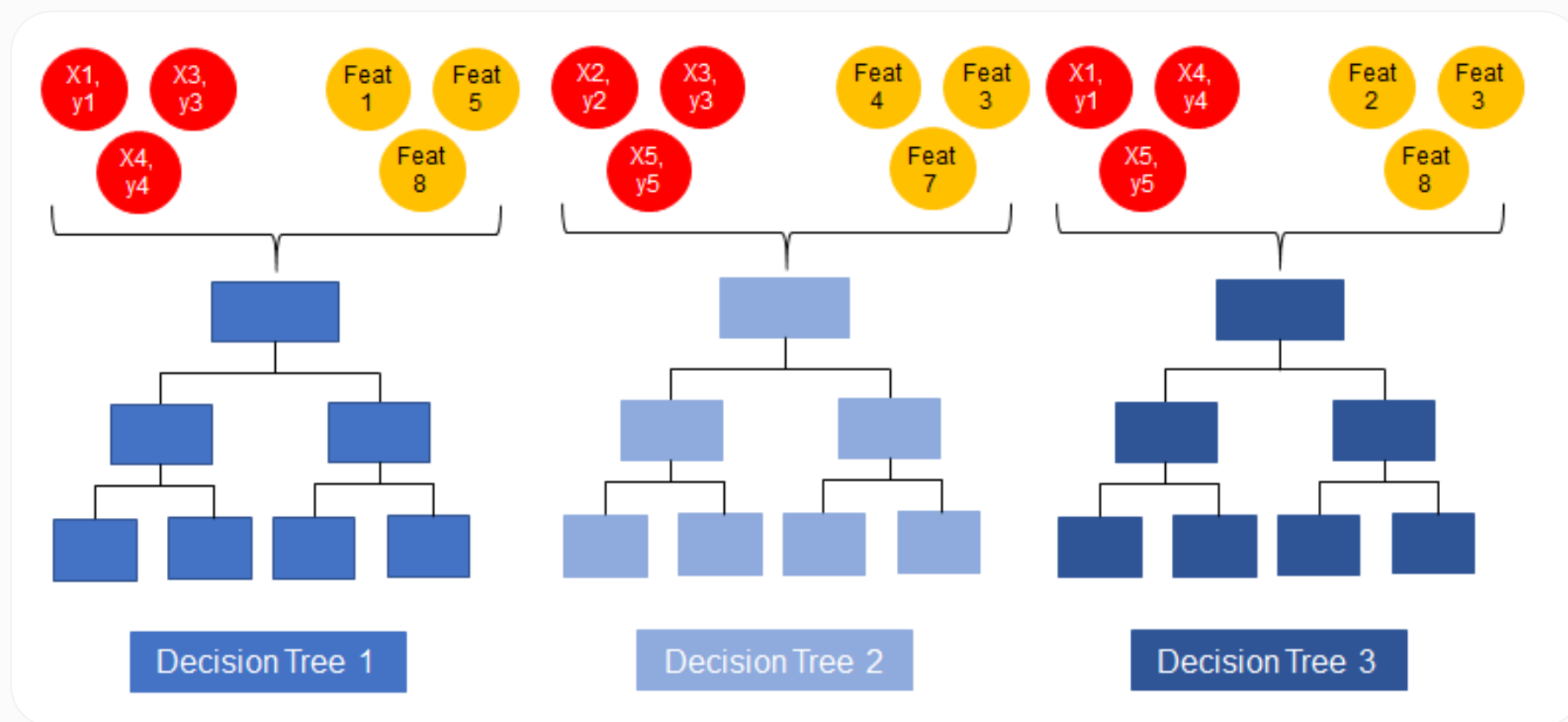
Skillbox

Бэггинг в задаче регрессии

$$a(x) = \frac{1}{N} \sum_{j=1}^N b_j(x)$$

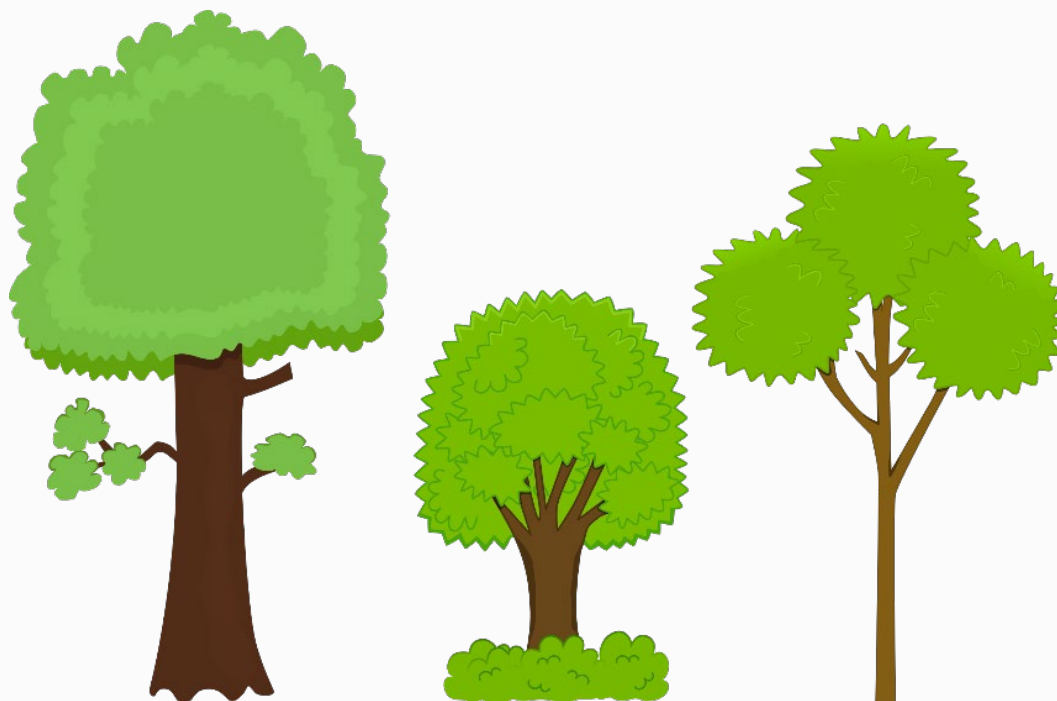


Subsampling



Случайный лес (random forest)

- В каждой вершине дерева будем искать разбиение не по всем признакам, а по подмножеству признаков
- Дерево строится до тех пор, пока в листе не окажется n_{min} объектов



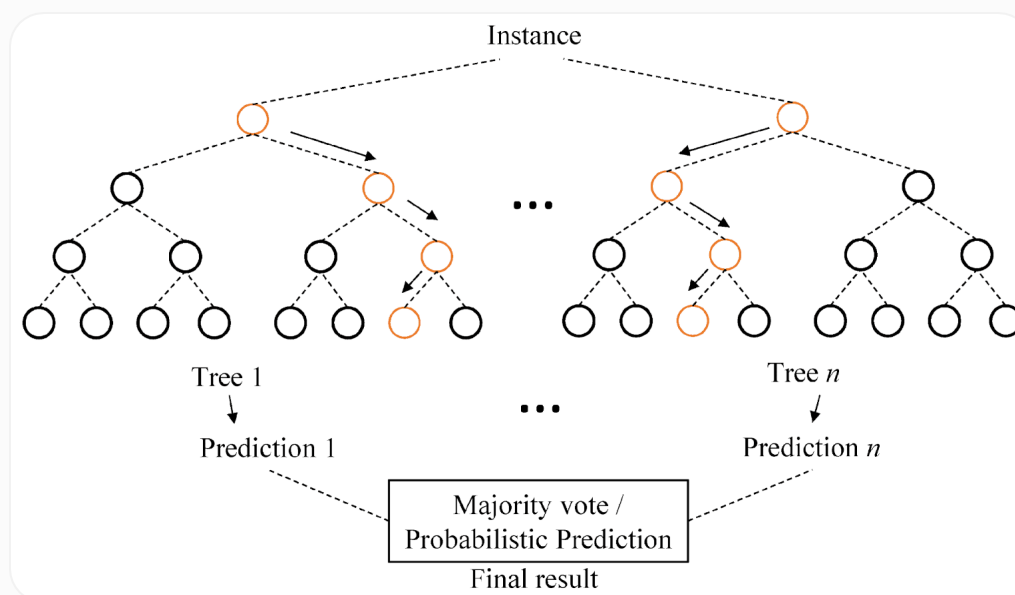
Random Forest

Алгоритм 3.1. Random Forest

- 1: **для** $n = 1, \dots, N$
 - 2: Сгенерировать выборку \tilde{X}_n с помощью бутстрэпа
 - 3: Построить решающее дерево $b_n(x)$ по выборке \tilde{X}_n :
 - дерево строится, пока в каждом листе не окажется не более n_{\min} объектов
 - при каждом разбиении сначала выбирается m случайных признаков из p , и оптимальное разделение ищется только среди них
 - 4: Вернуть композицию $a_N(x) = \frac{1}{N} \sum_{n=1}^N b_n(x)$
-

Случайный лес для классификации

- Для предсказания класса в задаче жёсткой классификации выбирается самый частый предсказанный деревьями класс
- В задаче мягкой классификации итоговая композиция усредняет эти вероятности и предсказывает класс, имеющий наибольшую вероятность



Вы узнали

- ✓ Что такое случайный лес и чем он отличается от бэггинга над решающими деревьями
- ✓ О применении случайных лесов для решения задач классификации и регрессии
- ✓ О большом количестве гиперпараметров у случайного леса, подбор которых значительно влияет на качество модели