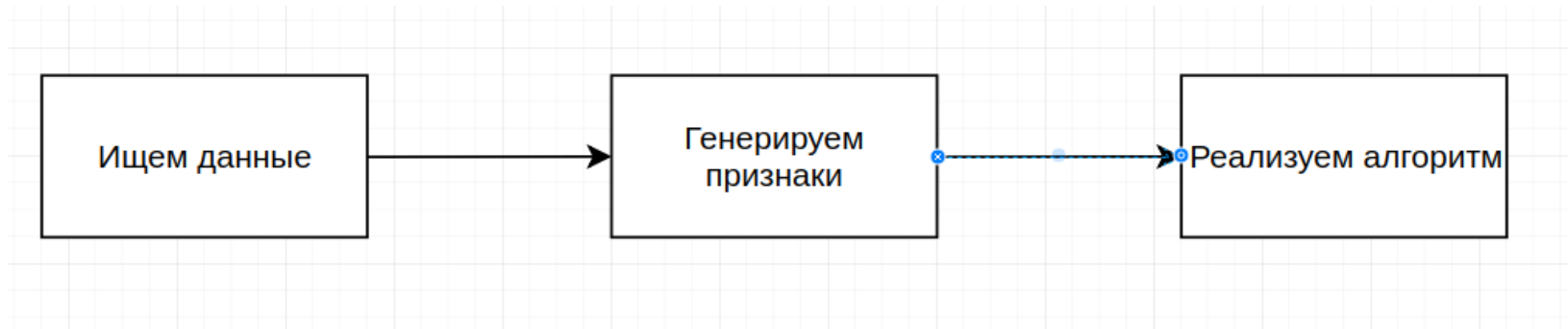


Задачи машинного обучения

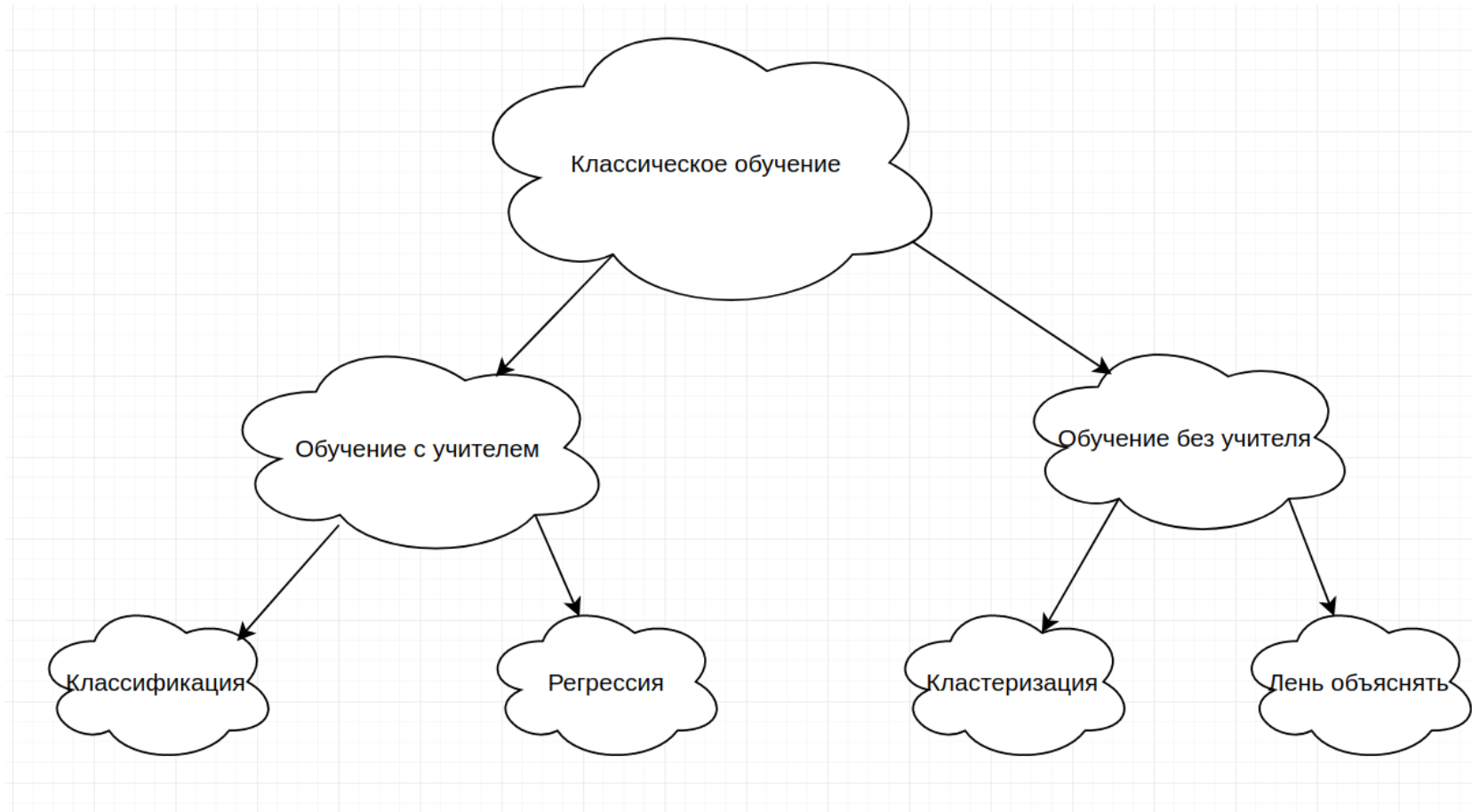
Что, откуда и куда?



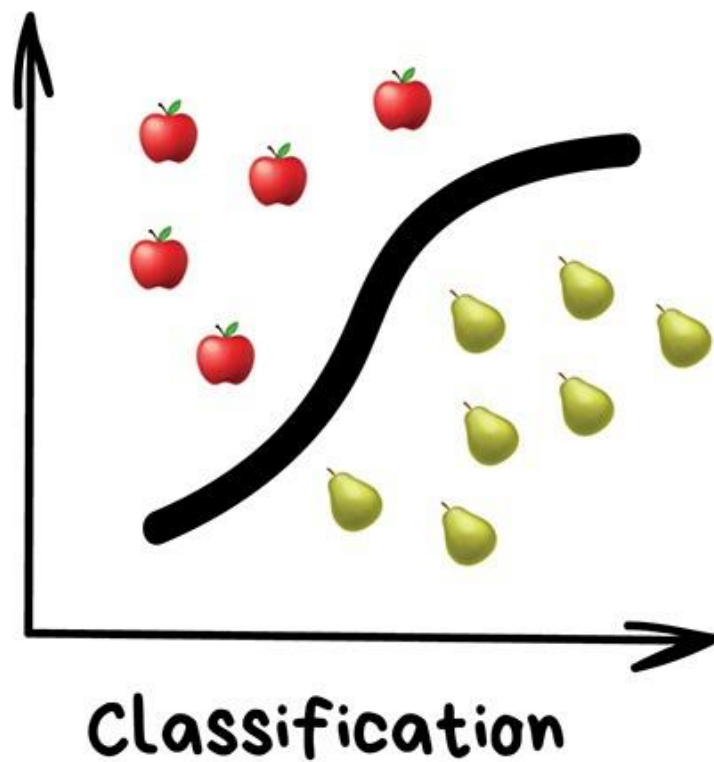
Виды машинного обучения



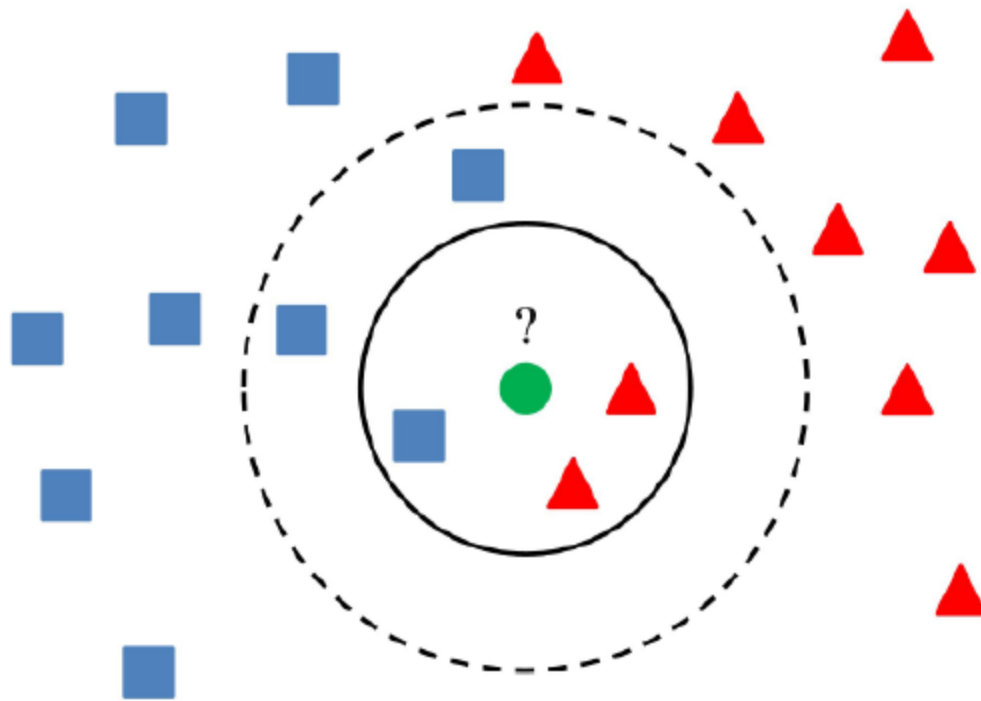
Классическое обучение



Классическое обучение. Классификация.



Классификация. kNN



Классификация. Наивный Байес.

привет...	1829
валера ...	1710
нет ...	1191
куда ...	1012
небо ...	985
огурцы ...	873
говорить...	747
третий ...	739

нормальные
письма

виагра ...	1552
казино ...	1492
100% ...	1320
кредит...	1184
скидка ...	985
нажми ...	873
free ...	747
доход ...	739

спам-письма

672 раза

«КОТИК»

13 раз

Простейший спам-фильтр
(использовались года до 2010)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

формула Байеса



не спам

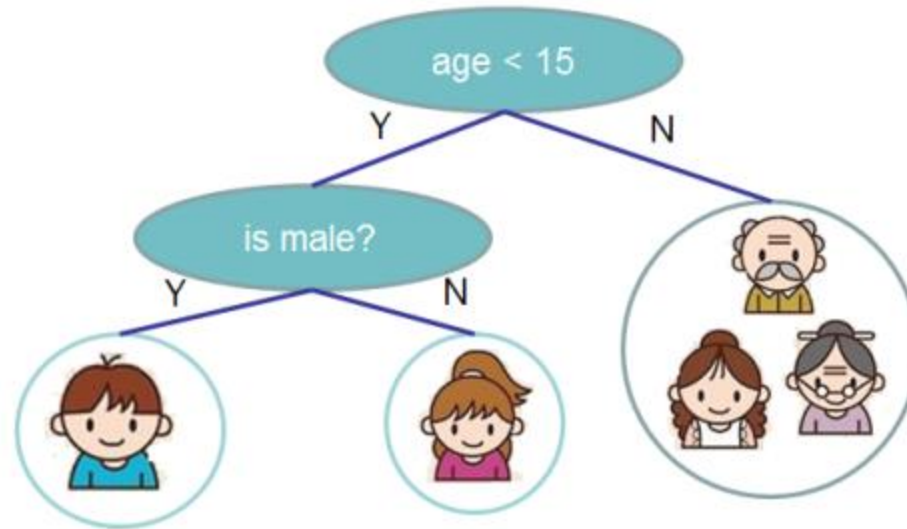
Наивный Байес

Классификация. Дерево решений.

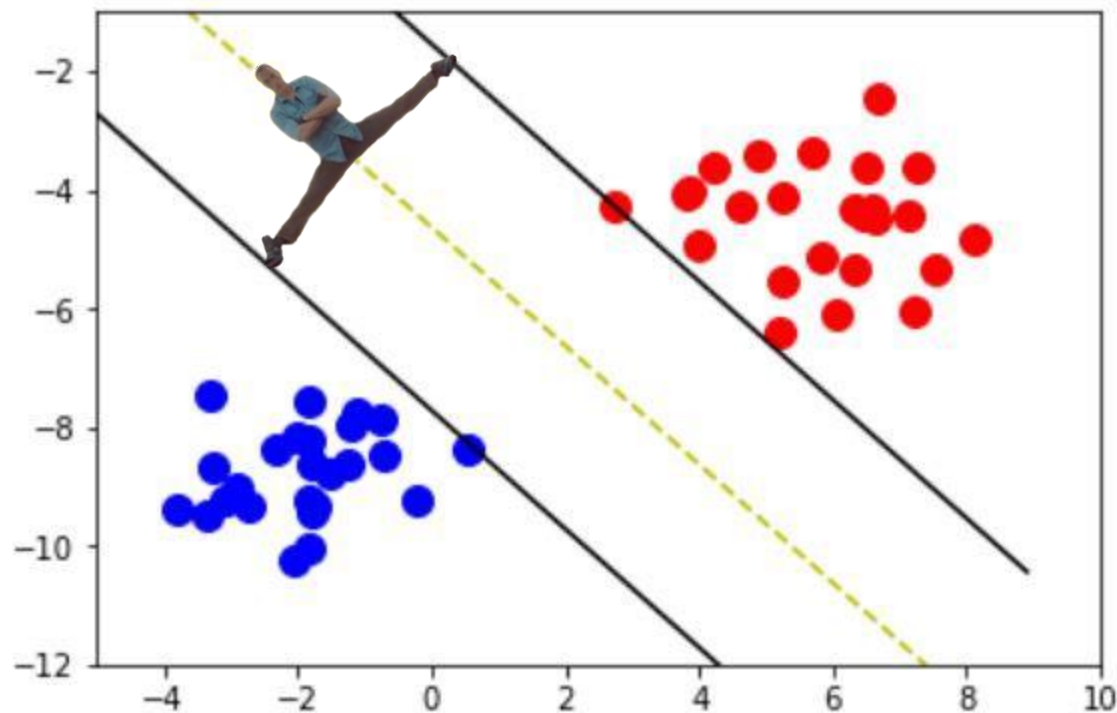
Input: age, gender, occupation, ...



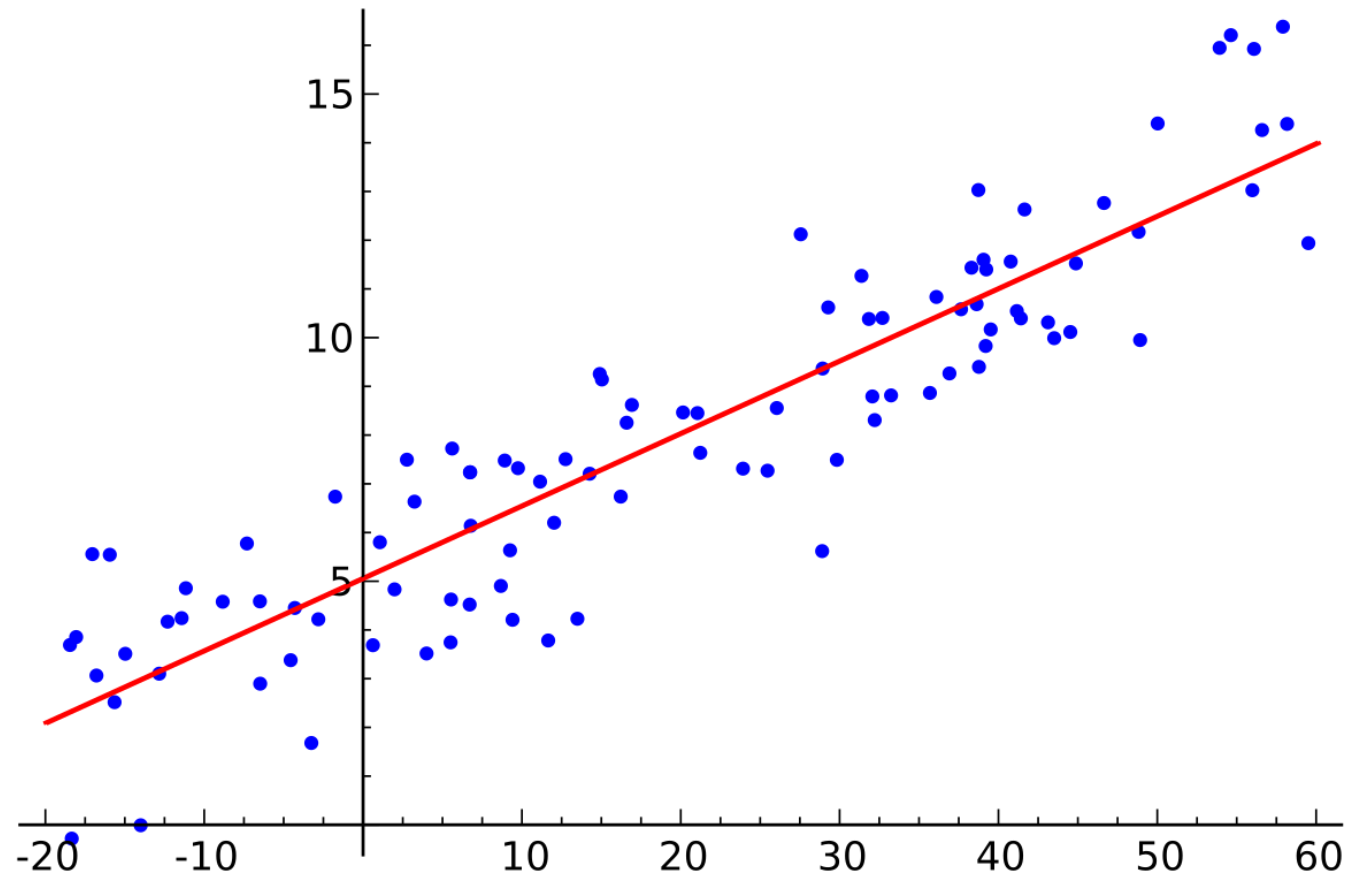
Does the person like computer games



Классификация. Метод опорных векторов.

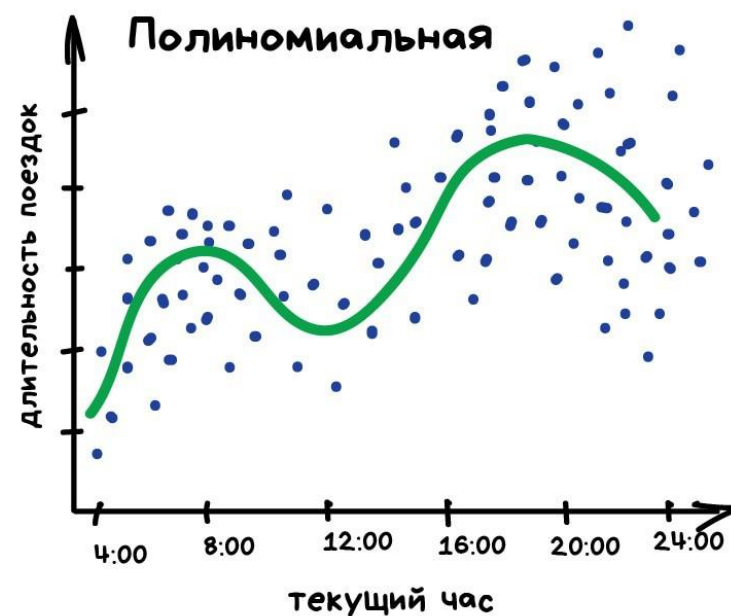
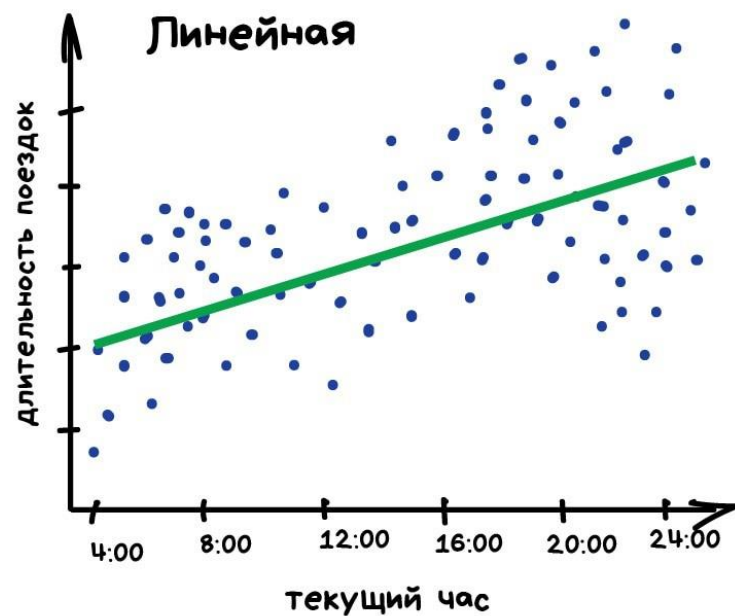


Классическое обучение. Регрессия.



Полиномиальная регрессия.

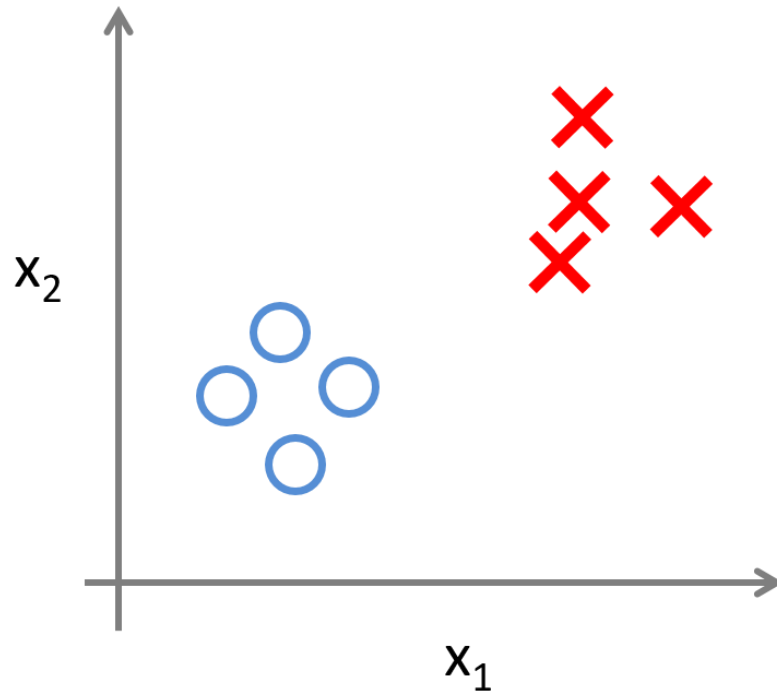
Предсказываем пробки



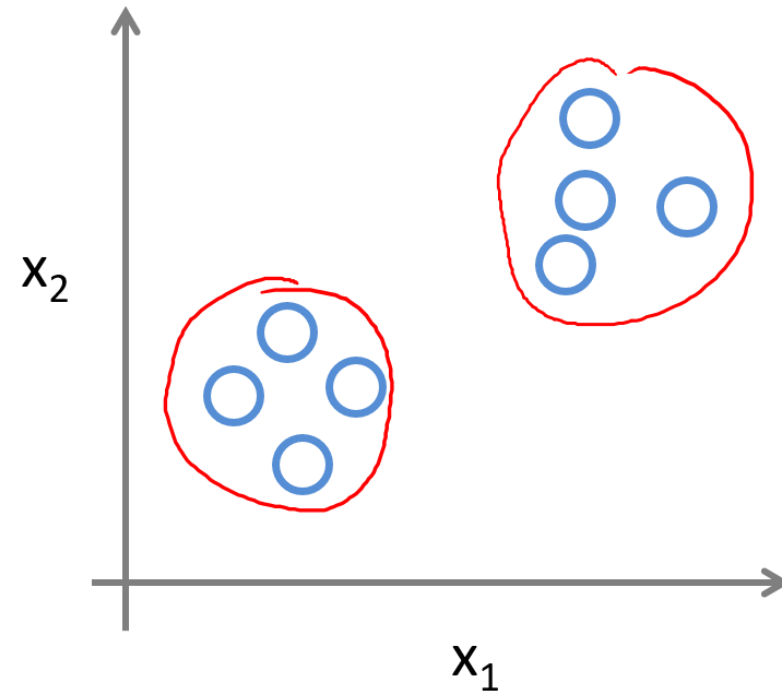
Регрессия

Обучение без учителя

Supervised Learning

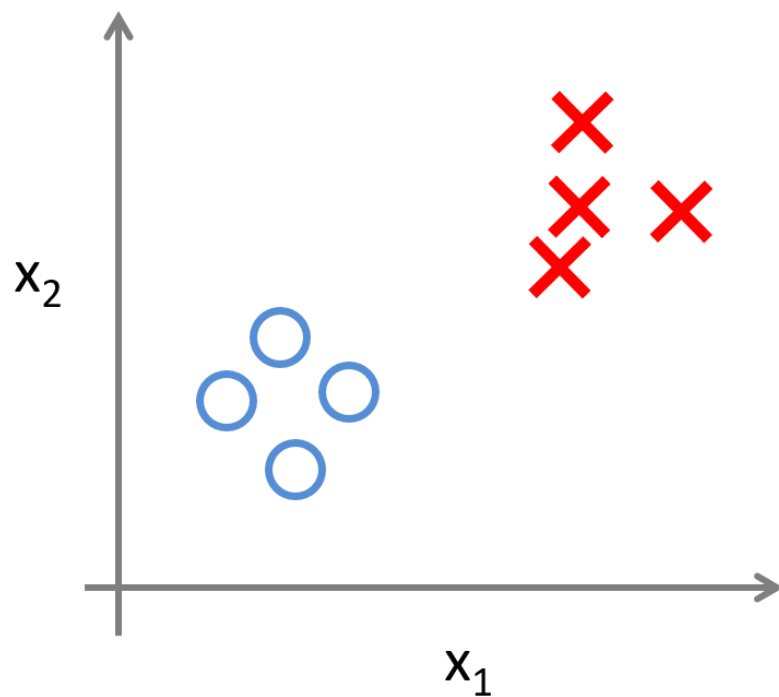


Unsupervised Learning

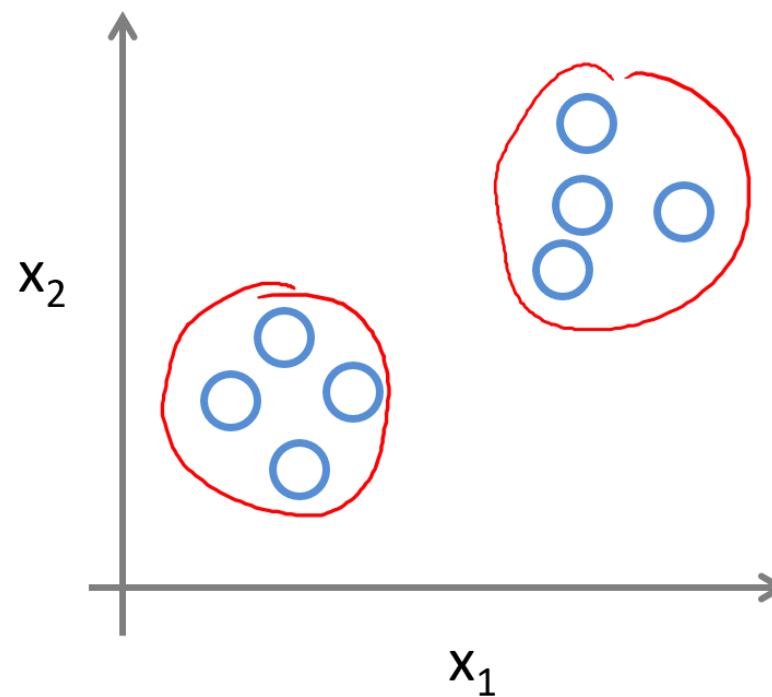


Кластеризация.

Supervised Learning

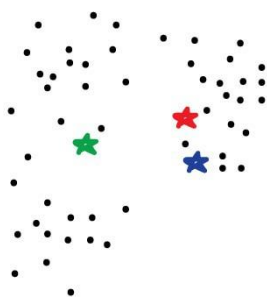


Unsupervised Learning

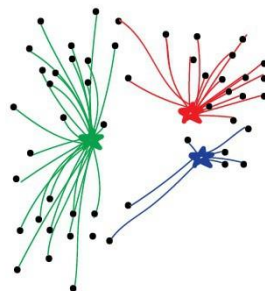


K-means

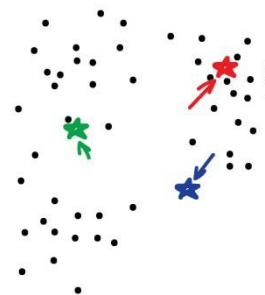
Ставим три ларька с шаурмой оптимальным образом
(иллюстрируя метод К-средних)



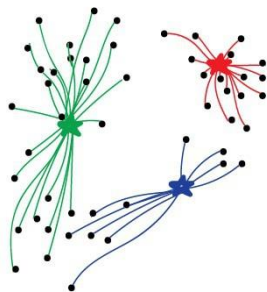
1. Ставим ларьки с шаурмой в случайных местах



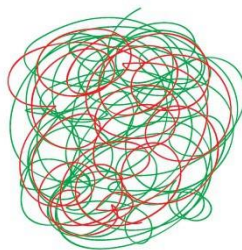
2. Смотрим в какой кому ближе идти



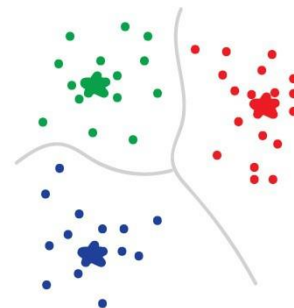
3. Двигаем ларьки ближе к центрам их популярности



4. Снова смотрим и двигаем

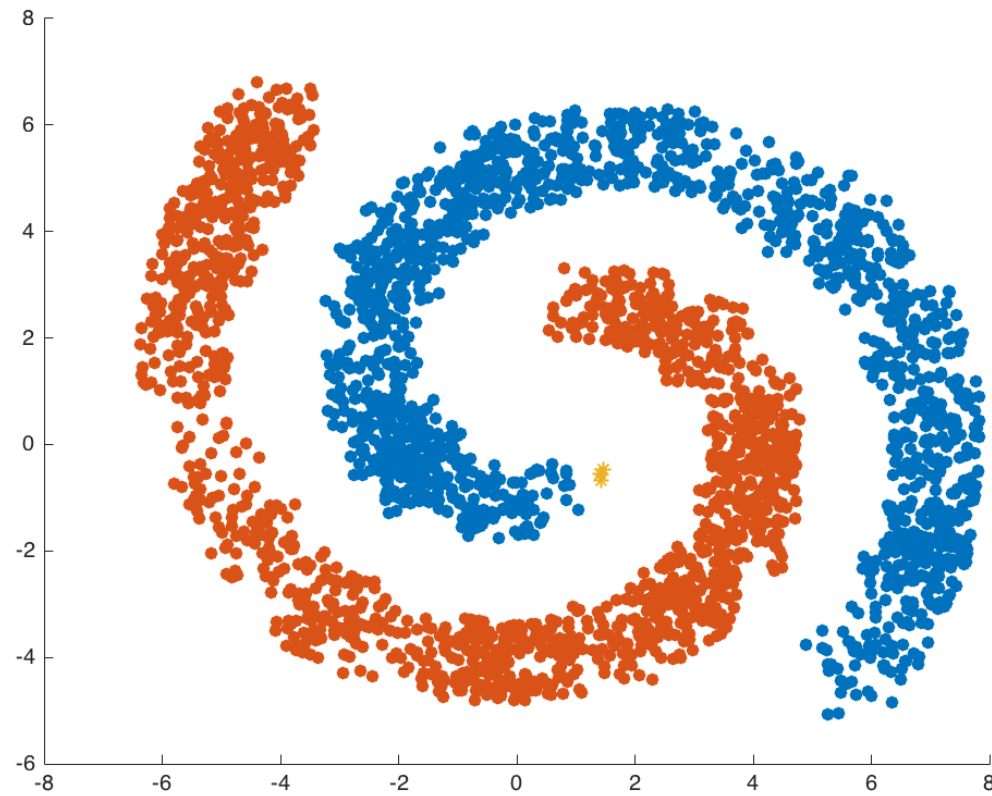


5. Повторяем много раз



6. Готово, вы великолепны!

DBSCAN



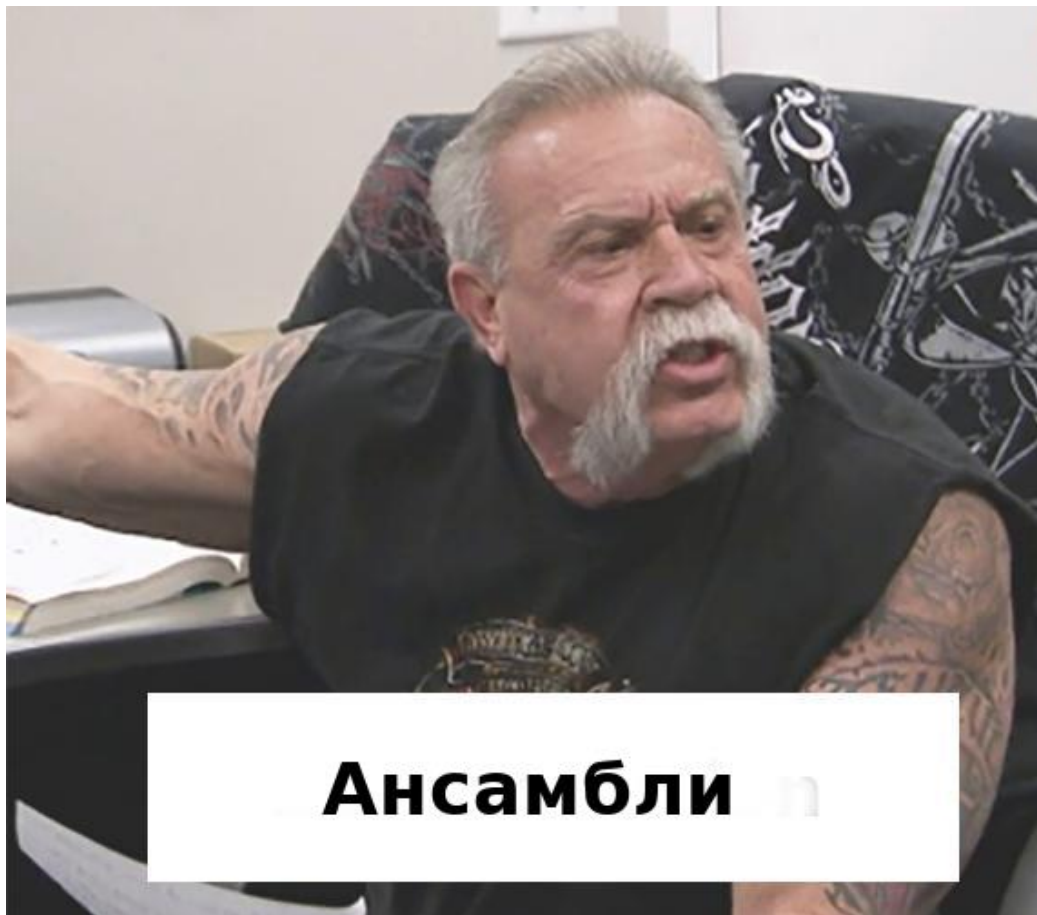
Обучение с подкреплением.



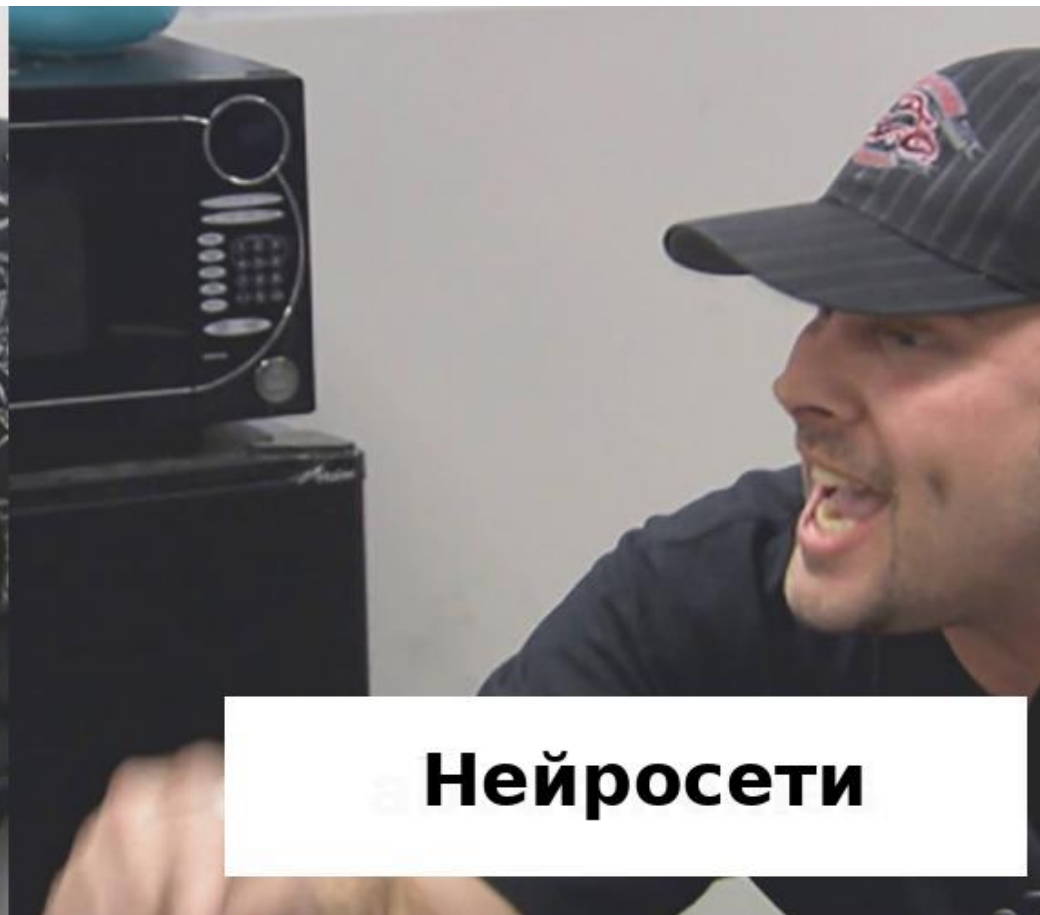
Q-learning, DQN



Ансамбли.

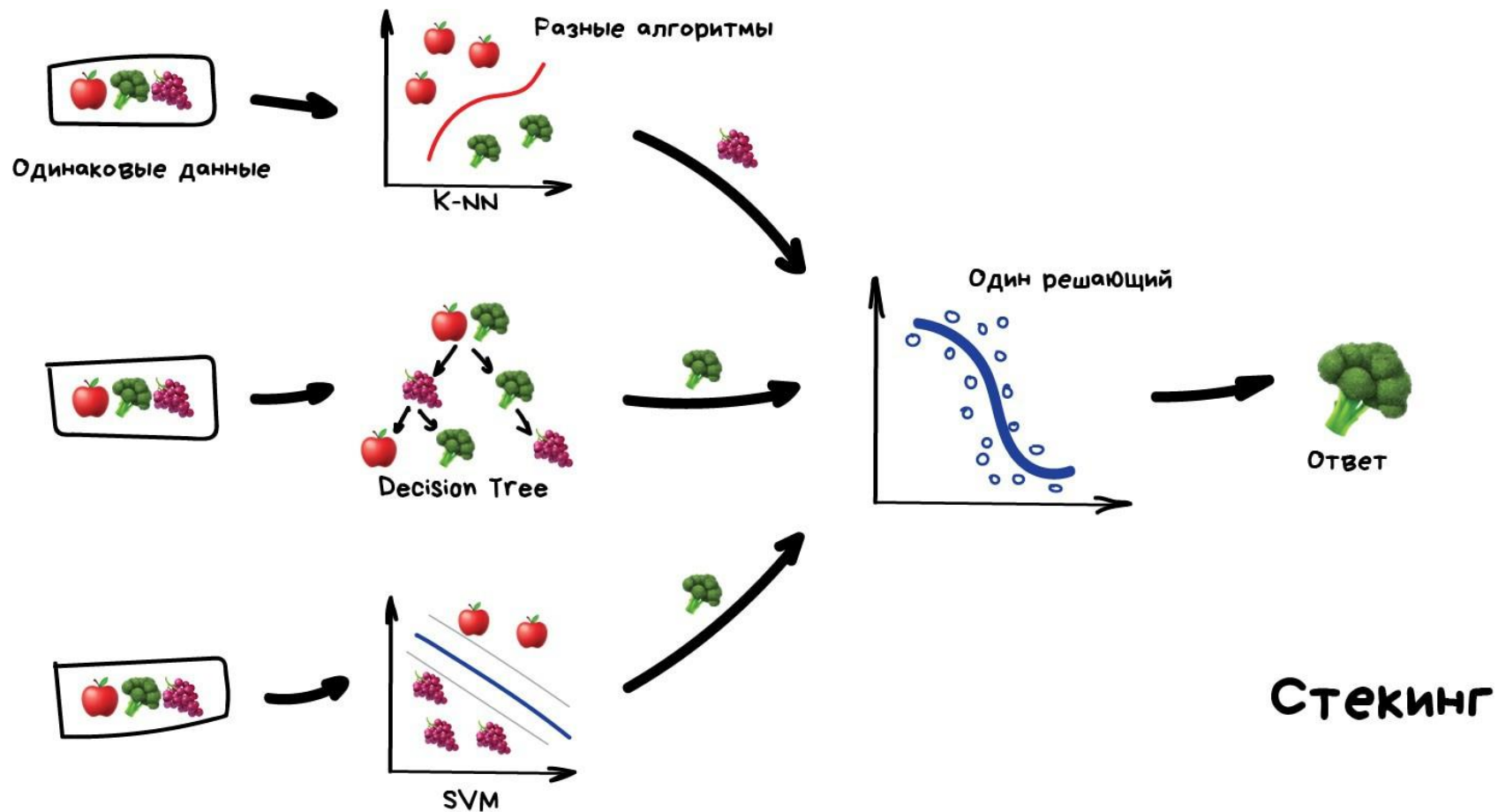


Ансамбли

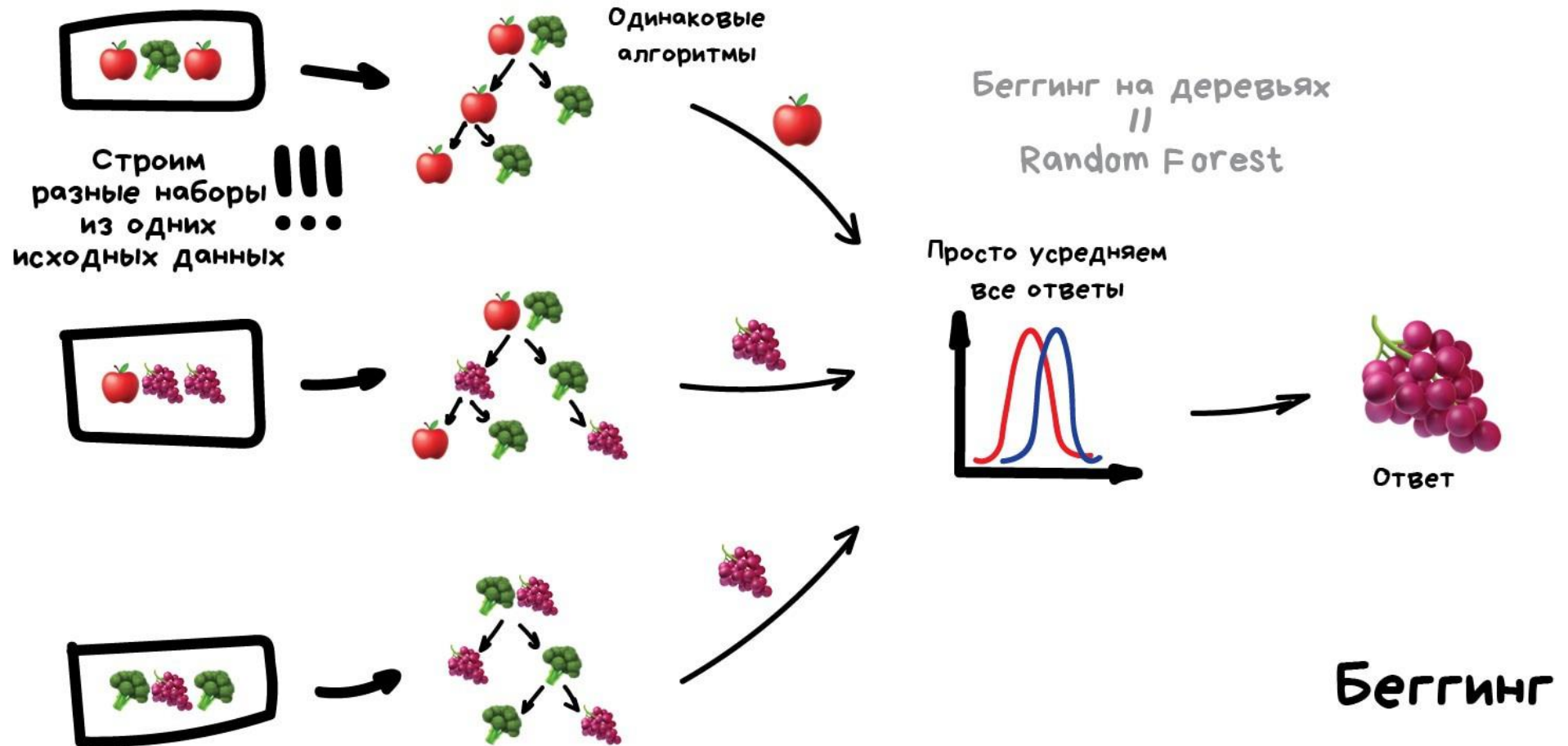


Нейросети

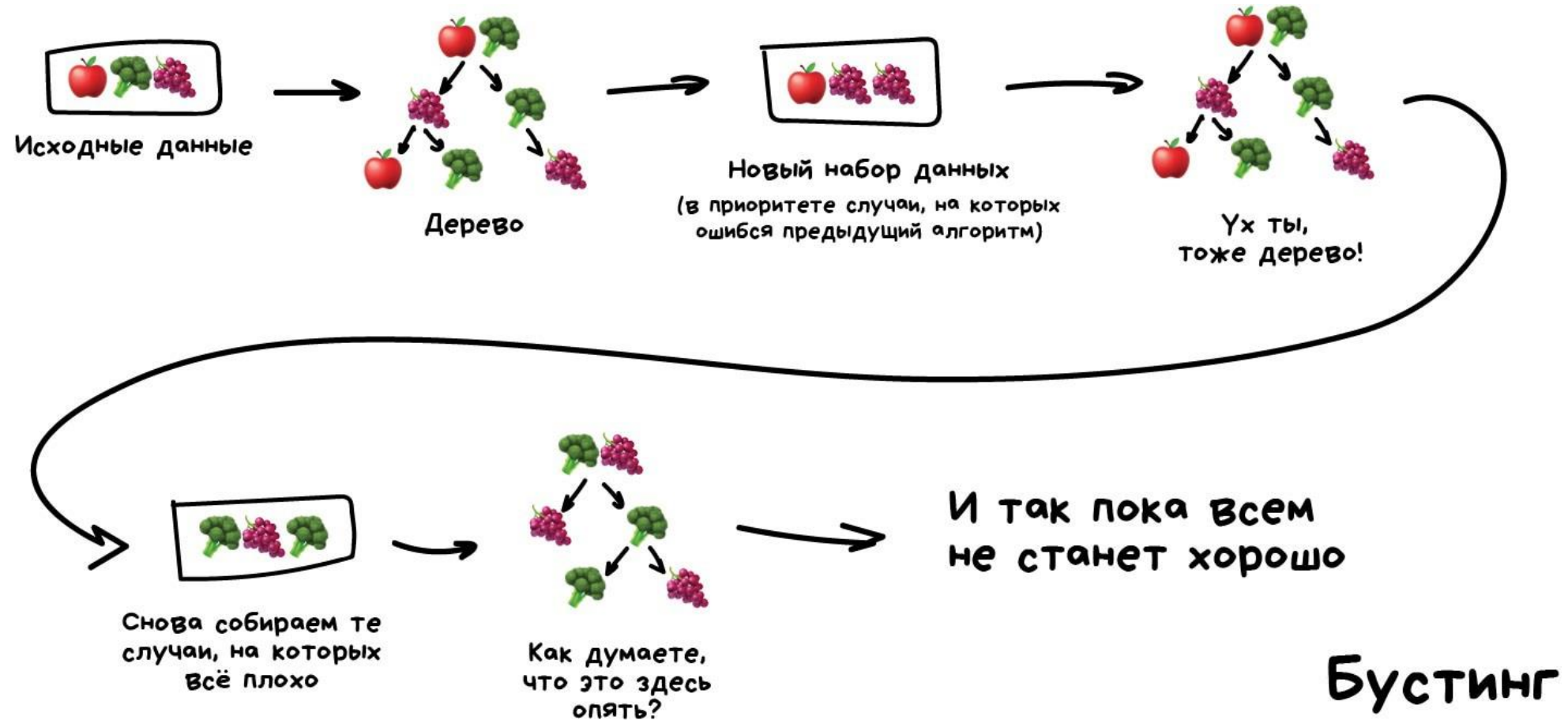
Ансамбли. Стекинг



Ансамбли. Беггинг



Ансамбли. Бустинг



Ансамбли. Бустинг

XGBoost

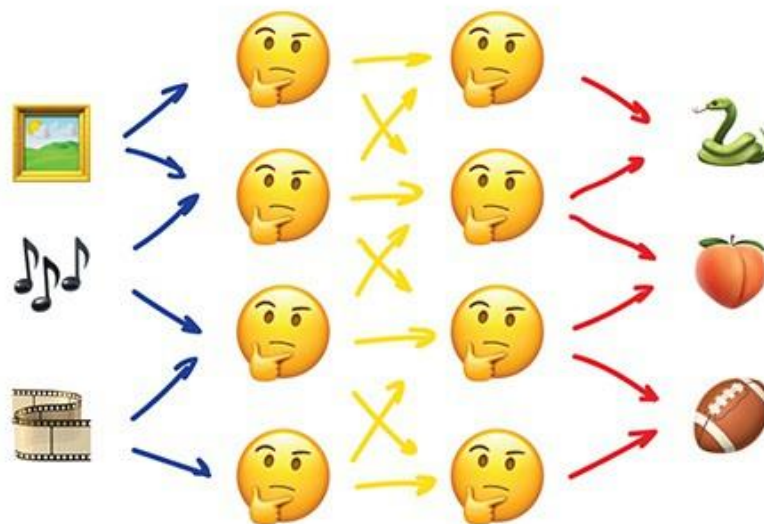
LightGBM



CatBoost

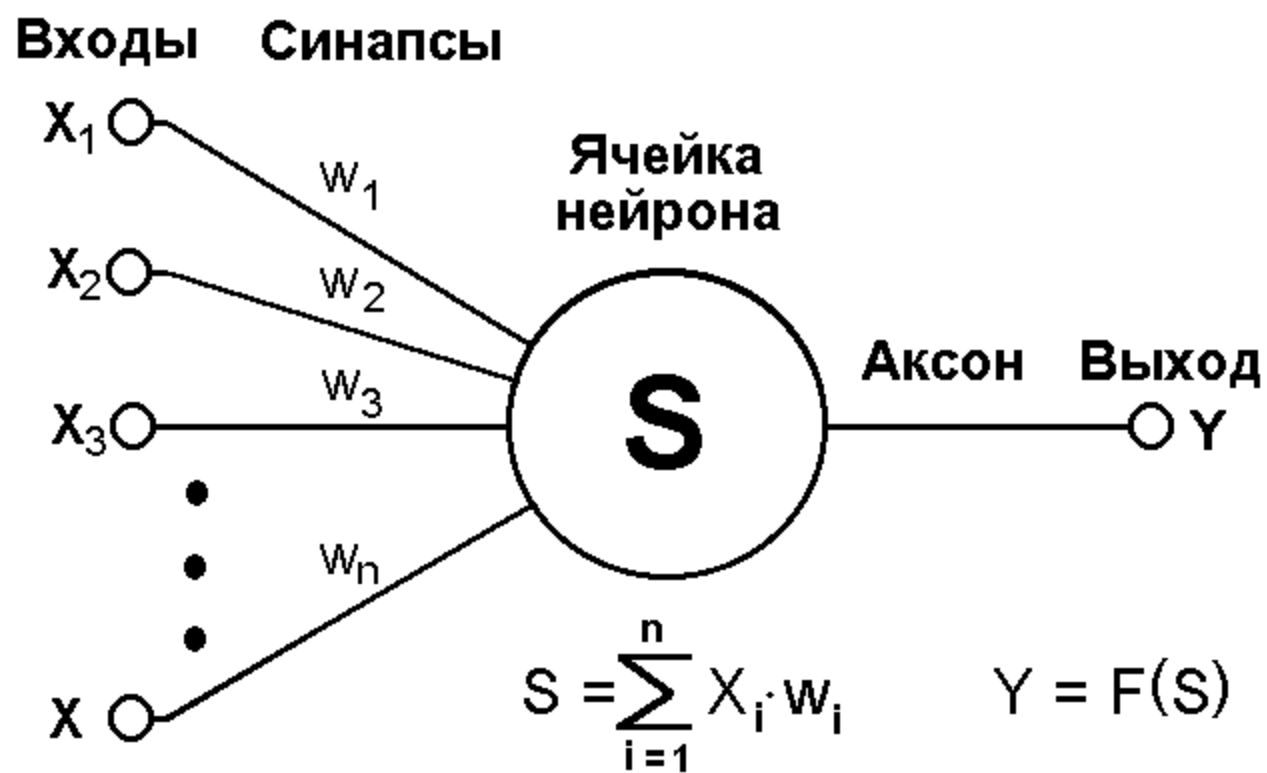


Нейросети

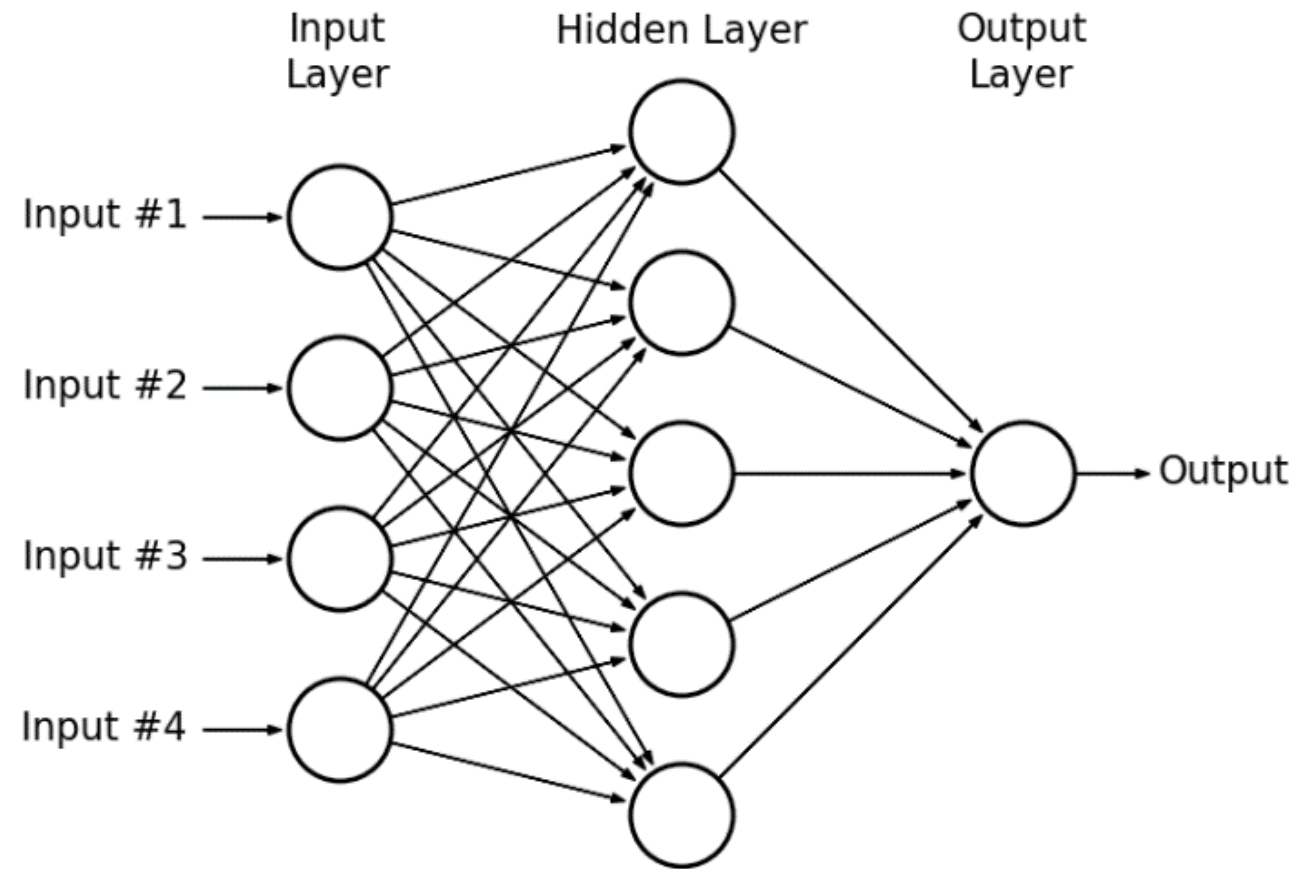


Neural Networks

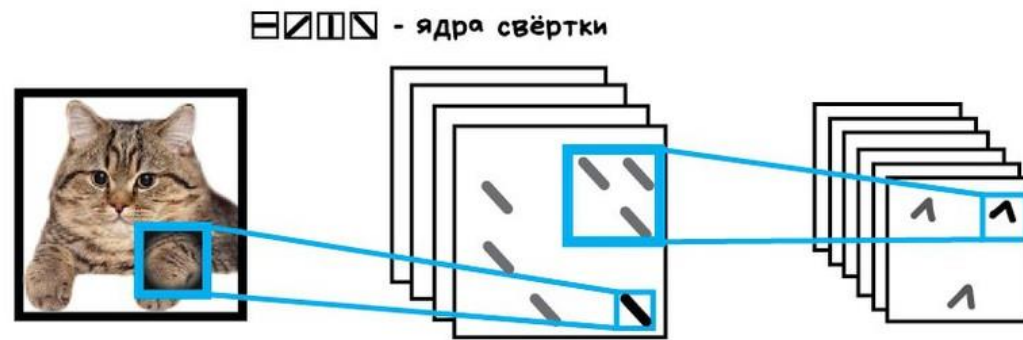
Нейрон



Многослойный перцептрон



Свёрточные нейросети

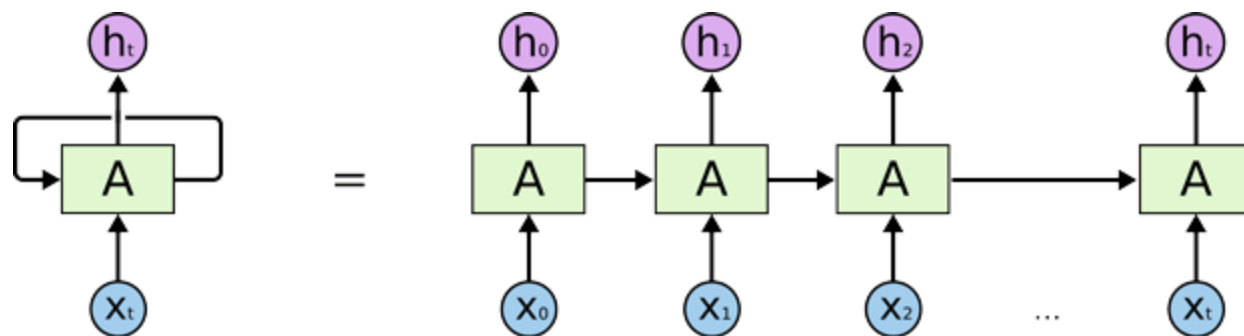


Сеть сама учится искать важные признаки,
собирая их из простых палочек



Свёрточная Нейросеть (CNN)

Реккурентные нейросети



Метрики

Метрики. Классификация.

- Accuracy - процент правильных ответов
- Пока достаточно знать только её :)

Метрики. Регрессия.

Mean squared error

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n e_t^2$$

Root mean squared error

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$$

Mean absolute error

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |e_t|$$
