ФОРМИРОВАНИЕ МОДЕЛИ

Балансировка

- Кошка1
- Кошка2
- Кошка3
- НЕкошка1
- НЕкошка2
- ….
- НЕкошка97

	Кот	неКот
Кот	0	0
неКот	3	97

Precision (кот) = 0
Recall(кот) = 0
Accuracy = 0.97

Балансировка

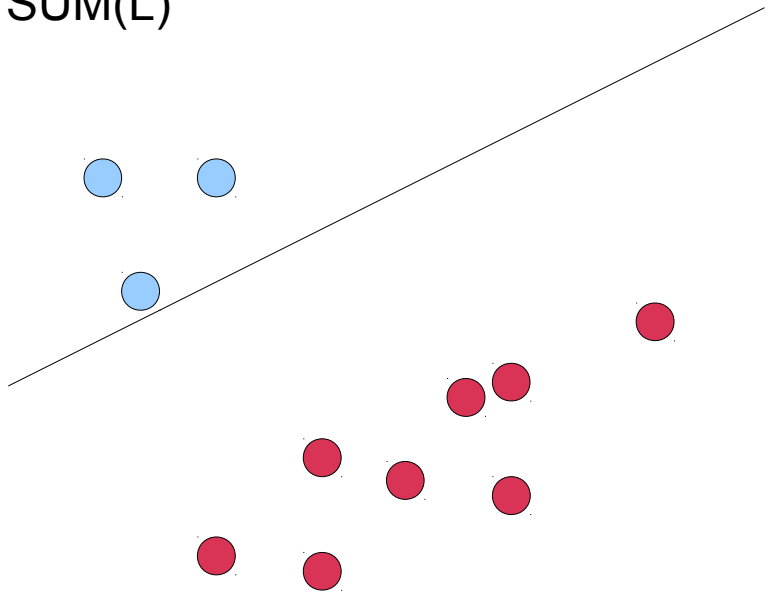
Кошка1

- Кошка2
- Кошка3
- НЕкошка1
- НЕкошка2
- ….
- НЕкошка97

L(y(Xi,W)) = [y(Xi,W)==y_truei]

yi= XiW

Q = SUM(L)



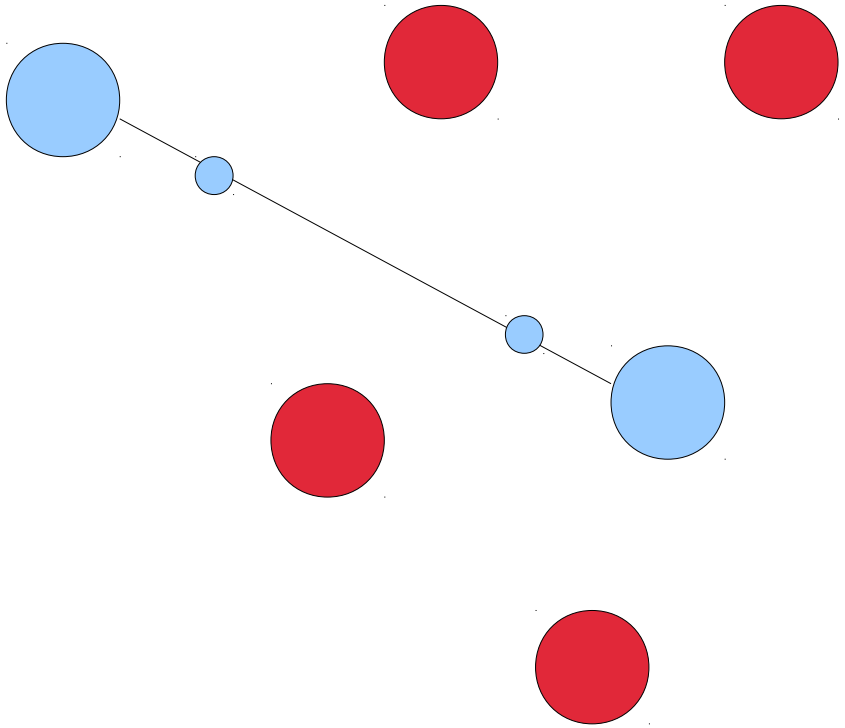
Балансировка

- Собрать больше данных
- Выбрать подходящую метрику качества
- Попробовать разные модели, одни модели более устойчивы к несбалансированным данным, чем другие
- Штраф за ошибки при прогнозе меньшего класса
- Undersampling и Oversampling
- Создание синтетических примеров для меньшего класса

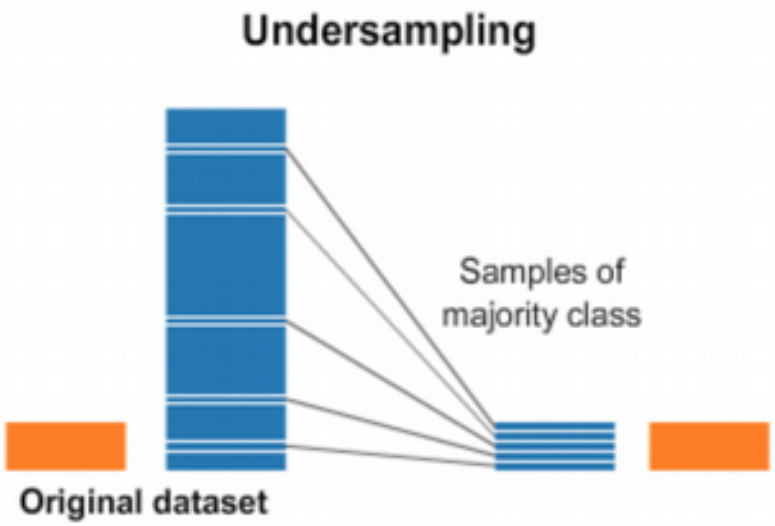
https://basegroup.ru/community/articles/imbalance-datasets

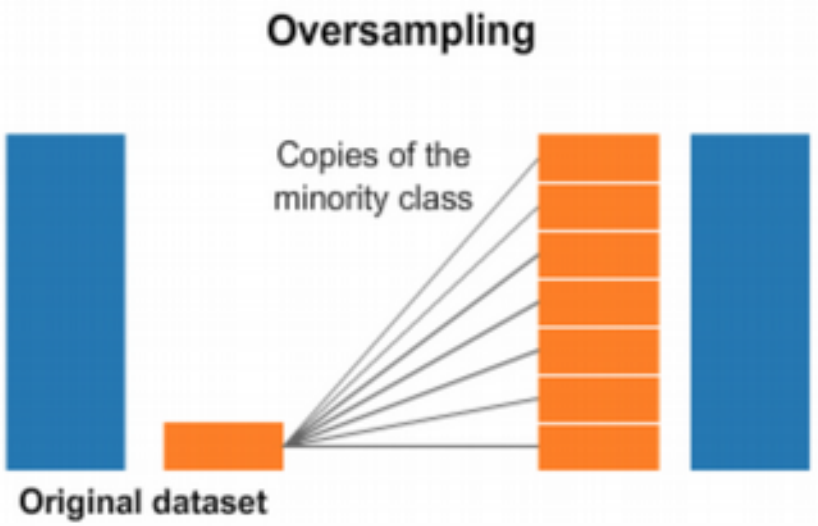
Балансировка

SMOTE

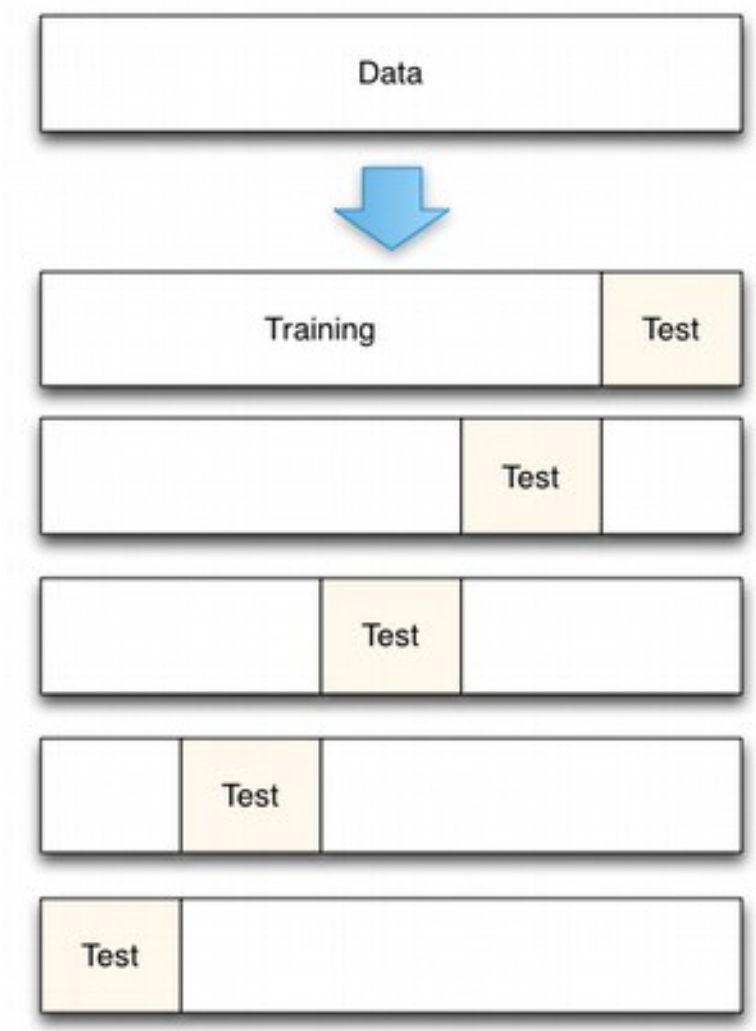


Балансировка





Кросс-Валидация

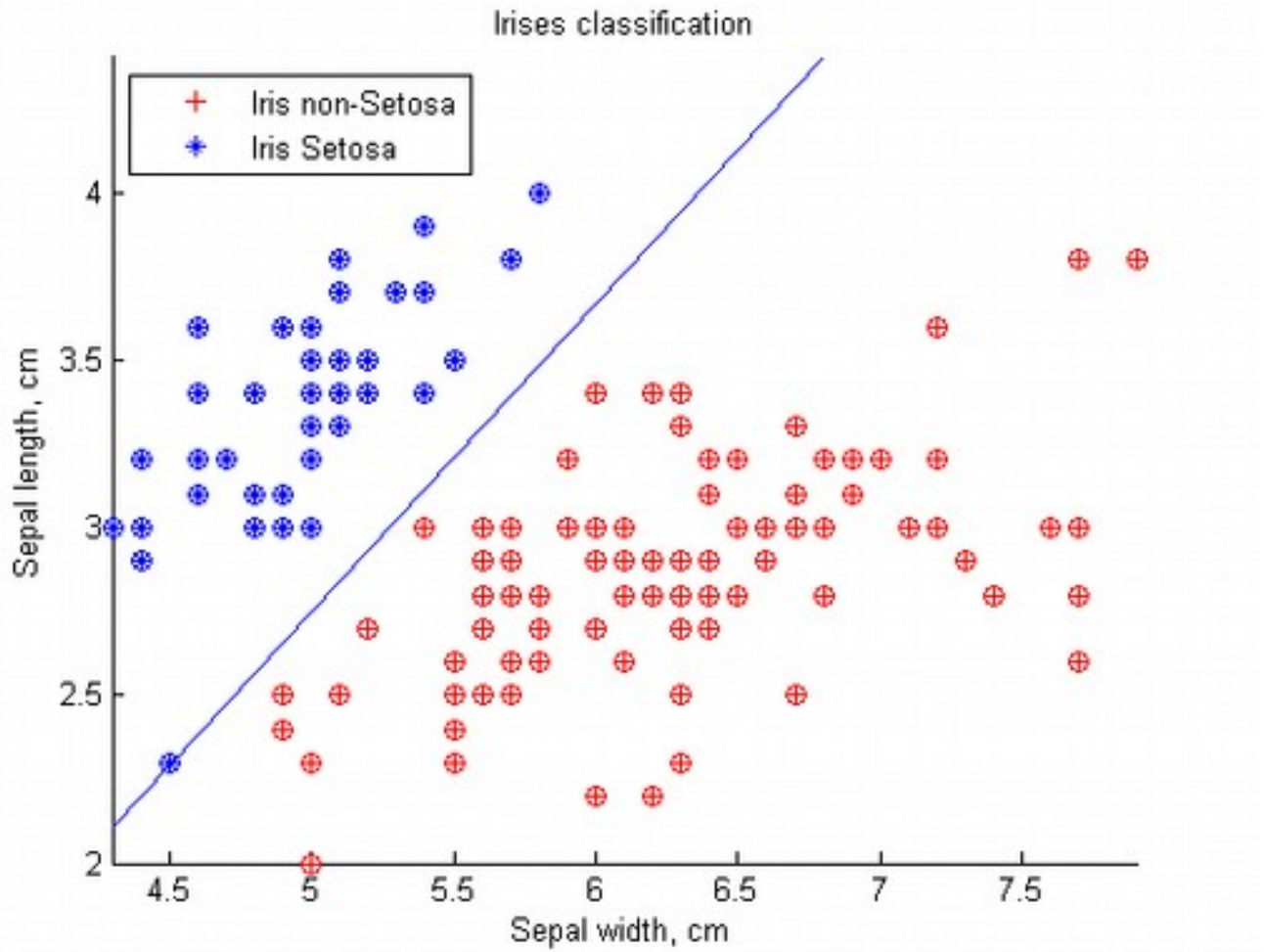


Кошка2

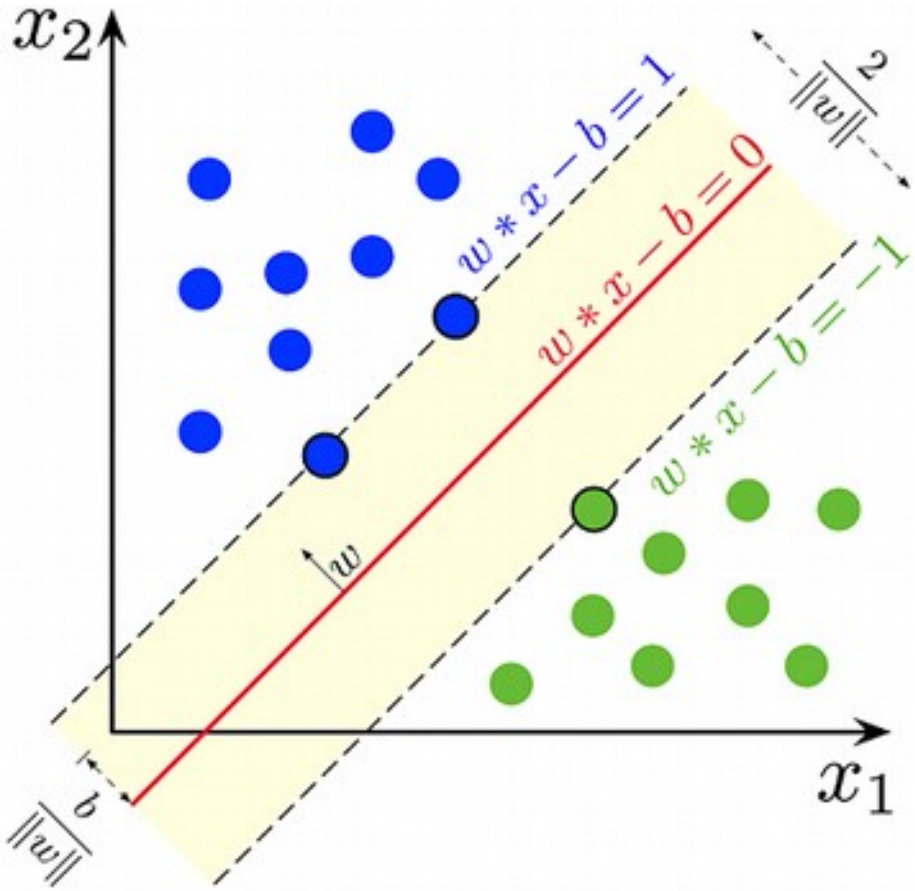
- Кошка3
- НЕкошка1
- НЕкошка2
- ….
- НЕкошка97

- Кошка1
- Кошка3
- НЕкошка1
- НЕкошка2
- ….
- НЕкошка97

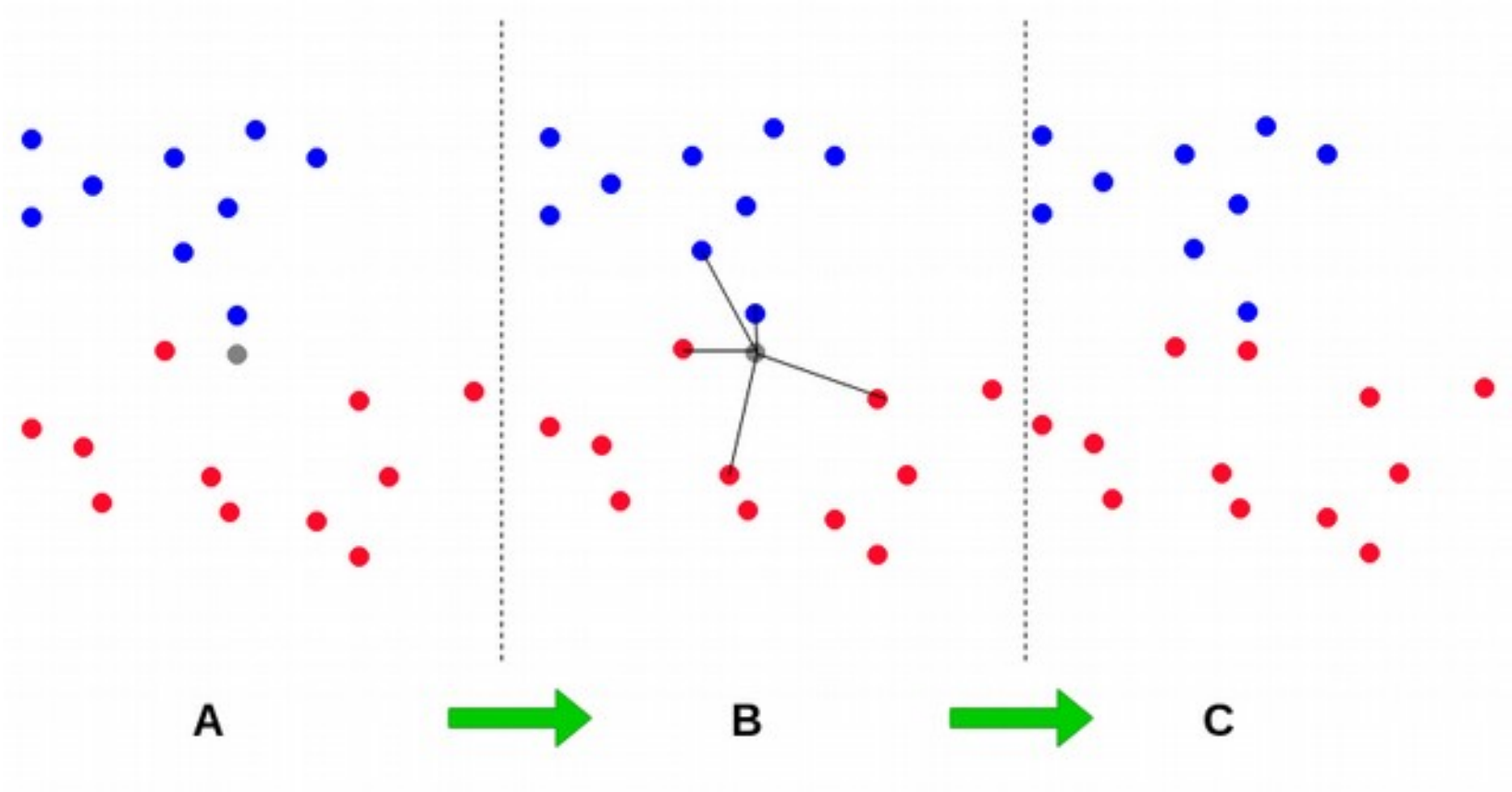
Классификаторы



Классификаторы - SVM



Классификаторы - KNN



Классификаторы — Случайный лес