

ДСА

Дремов Дмитрий, 3-е место

План

- Подготовка обучающего множества
- Фильтрация
- Признаки
- Ранжирование
- Замыкание

Обучающее множество

- Бьем на компоненты связности
- Выбираем компоненты так, чтобы они были ровно размером с тестовое множество по количеству вершин
- Также добавляем “шум” как и в тестовом множестве.
- Два “фолда” ~ 80%

Фильтрация

- Ближайшие соседи отдельно по каждому фонду
- Идеи те же что и у @bendyna
- Изначально более жесткая фильтрация: 58% полноты и 3.7% точности
- В итоге еще более жесткая фильтрация до 53% полноты и 5%+ точности

Признаки - куда

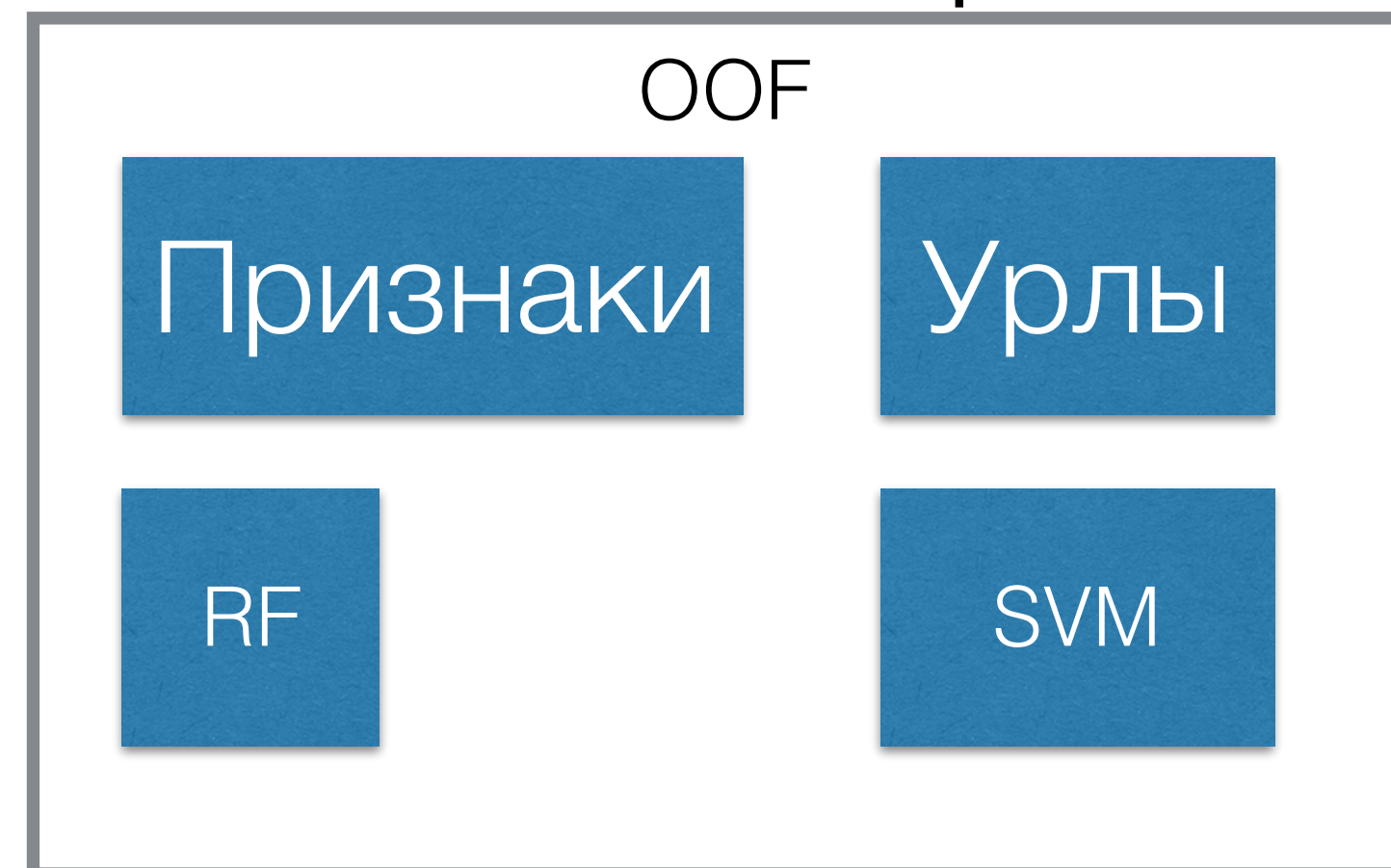
- Только urls
- Пути “A/B/C?D” : {“A”, “A/B”, “A/B/C”, “A/B/C/D”}
- SVD
- Расстояния L1, L2, L_max, KLD, cosine, cosine-non-norm

Признаки - когда

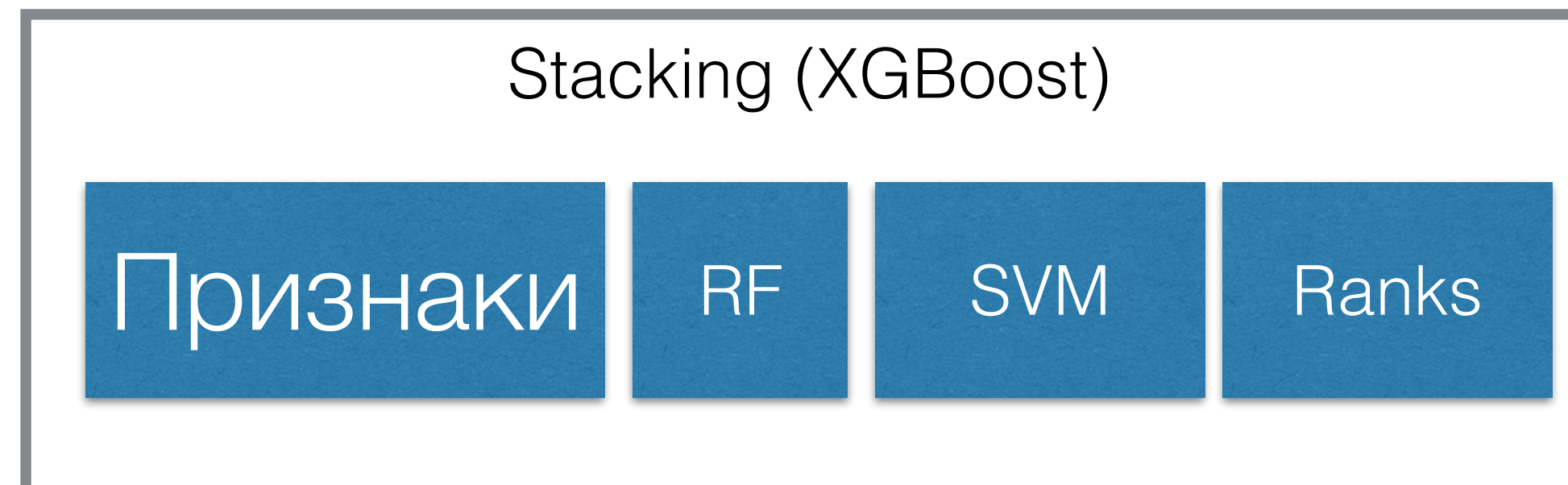
- Распределение посещений внутри дня
- Распределение времен между последовательными посещениями
- Окошки от минимального до максимального времени

Ранжирование

Базовые алгоритмы



Stacking (XGBoost)

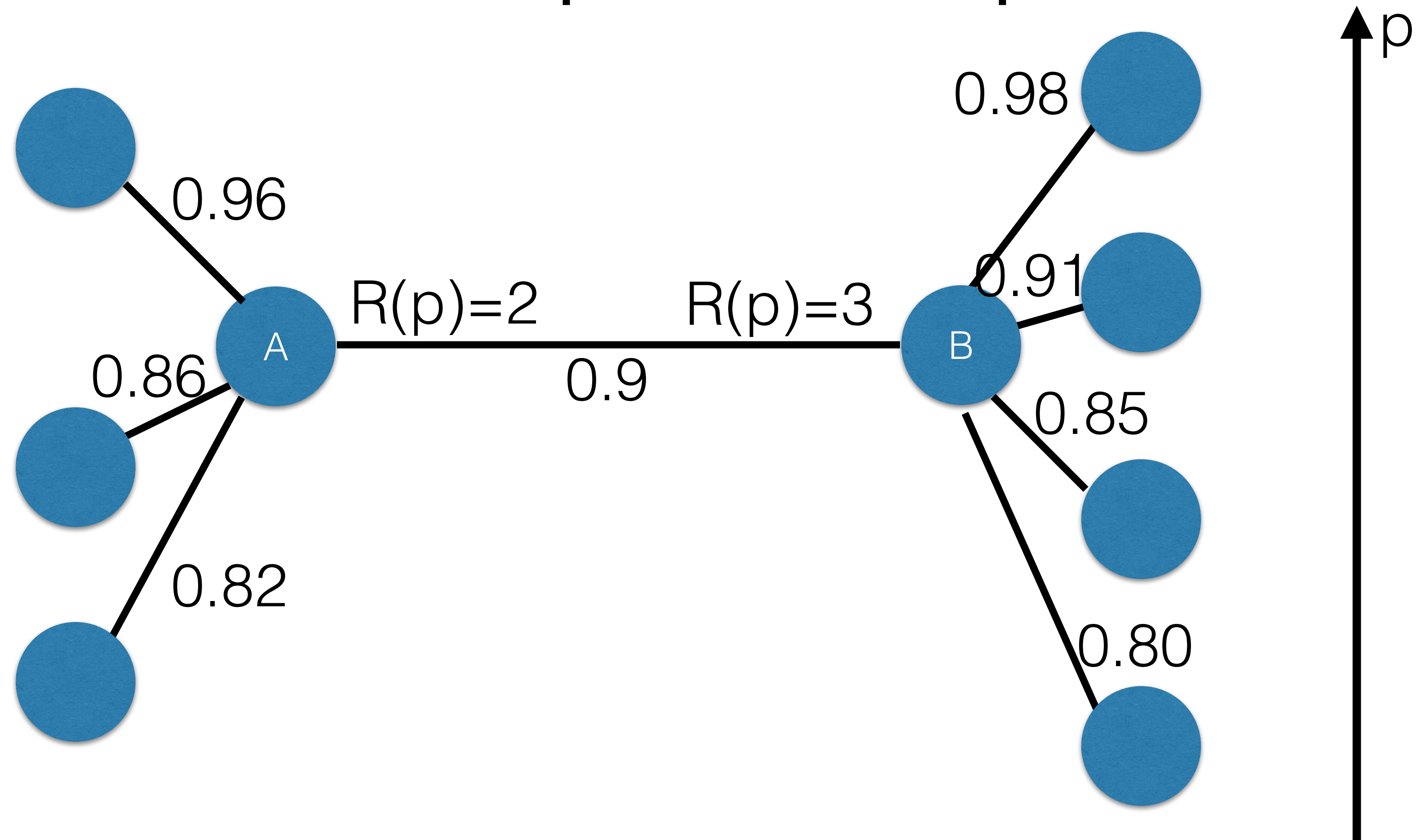


Ранжирование СТЕКИНГ

- Базовые признаки
- Предсказание RF
- Предсказание SVM (N моделей на N соседей)
- “Ранг” признака по ребрам $+0.012$ (отдельно по каждому фолду)

Ранжирование

Мета признак “ранг”



Замыкание +0.004

