

МЕТОДЫ РАНЖИРОВАНИЯ

ПОДХОДЫ К РАНЖИРОВАНИЮ



- Pointwise
- Pairwise
- Listwise

- › $y(q, d) \in \mathbb{R}$ — релевантность документа d запросу q
- › $a(q, d)$ — оценка релевантности
- › Будем предсказывать $y(q, d)$ методами регрессии
- › Например:

$$\sum_{i=1}^{\ell} (a(q, d) - y(q, d))^2 \rightarrow \min$$

- Минимизируем количество дефектных пар:

$$\sum_{x_i < x_j} [a(x_j) - a(x_i) < 0] \rightarrow \mathbf{min}$$

- › Минимизируем количество дефектных пар:

$$\sum_{x_i < x_j} L(a(x_j) - a(x_i)) \rightarrow \mathbf{min}$$

- › $L(M)$ — гладкая функция
- › $L(M) = \log(1 + e^{-M})$ — метод RankNet

- › RankNet, шаг стохастического градиентного спуска для линейной модели:

$$w := w + \eta \frac{1}{1 + \exp(\langle w, x_j - x_i \rangle)} (x_j - x_i)$$

- › Как оптимизировать NDCG вместо $L(M)$?

- › Домножим стохастический градиент по паре (x_i, x_j) на изменение **NDCG** при перестановке x_i и x_j местами:

$$w := w + \eta \frac{1}{1 + \exp(\langle w, x_j - x_i \rangle)} (x_j - x_i) \cdot |\Delta \text{NDCG}_{ij}|$$

- › Эмпирическое наблюдение: такая модификация действительно приводит к оптимизации **NDCG**

- › Три подхода: pointwise, pairwise, listwise
- › Наиболее часто используется попарный подход