МЕТОДЫ РАНЖИРОВАНИЯ

ПОДХОДЫ К РАНЖИРОВАНИЮ

<u>\МФТИ</u>,

- Pointwise
- Pairwise
- Listwise

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

$$\sum_{i=1}^{n} \left(a(q,d)-y(q,d)
ight)^2 o ext{min}$$

Минимизируем количество дефектных пар:

$$\sum_{x_i < x_j} [a(x_j) - a(x_i) < 0] o extstyle m{min}$$

> Минимизируем количество дефектных пар:

$$\sum_{x_i < x_i} L(a(x_j) - a(x_i)) o \mathsf{min}$$

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

RankNet, шаг стохастического градиентного спуска для линейной модели:

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

 $lacksymbol{
ho}$ Как оптимизировать NDCG вместо $oldsymbol{L}(oldsymbol{M})$?

LISTWISE-ПОДХОД

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

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

 Эмпирическое наблюдение: такая модификация действительно приводит к оптимизации NDCG

РЕЗЮМЕ

<u>\МФТИ</u>.

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