

1. Для предсказания класса по уровням ранжирования выгоднее брать максимальное значение, а для предсказания классов по вероятностям лучше брать среднее арифметическое.
2. Шум к выходным меткам добавляется для того, чтобы получать вероятность правильности метки была  $1-\epsilon$ , где  $\epsilon$  является шумом. Потому что в ином случае правильной меткой может быть любая другая метка.
3. Обзоры слишком короткие, что может дать недостаточно информации, чтобы нейронная сеть правильно предсказала значение.