

# Вопросы к экзамену по ММРО

ф-т ВМиК, кафедра ММП, весна 2017.

## Процедура экзамена

Вы получаете 2 билета и 30 минут на подготовку. Далее вы отвечаете и получаете дополнительные вопросы и задачи. В процессе подготовки к билетам и решения задач вы можете пользоваться заранее заполненным (перед экзаменом) листом А4 с основными моментами из лекций. Лист А4 должен быть написан от руки лично вами, его можно заполнять с 2х сторон. Ничем другим пользоваться нельзя. Для получения положительной оценки вы должны знать ответы на вопросы из теоретического минимума без использования А4.

## Билеты

1. Методы многоклассовой классификации бинарными классификаторами - подходы один против всех, каждый против каждого и подход на кодах, исправляющих ошибки.
2. Вывод разложения среднеквадратичной ошибки на смещение и дисперсию. Интуиция. При каких параметрах основных прогнозирующих алгоритмах в их прогнозах преобладает фактор дисперсии, а при каких - смещения?
3. Фиксированные схемы построения ансамблей, стэкинг моделей. Алгоритмы бэггинга, случайных подпространств, случайного леса и ExtraRandomTrees.
4. Алгоритм градиентного бустинга, его усовершенствование для деревьев в качестве базовых алгоритмов. Shrinkage, subsampling. Особенности шага бустинга для аппроксимации рядом Тейлора 2-го порядка.
5. Вывод алгоритма xgBoost. Чем он превосходит градиентный бустинг?
6. Отбор признаков по корреляции, взаимной информации и relief-критерию.
7. Алгоритм последовательного отбора признаков и его модификации. Алгоритм генетического отбора признаков.
8. Неравенство Йенсена. Когда оно обращается в равенство? Неотрицательность расстояния Кульбака-Лейблера (с док-вом). 3 эквивалентных критерия выпуклости и строгой выпуклости для дважды дифференцируемых функций (без док-ва).
9. Вывод EM-алгоритма в общем случае. Почему он приводит к монотонному неубыванию ф-ции правдоподобия. Что меняется в случае регуляризации и MAP оценки?
10. Вывод EM-алгоритма для моделирования смесями нормальных распределений. Упрощения ковариационной матрицы. Связь с методом K-средних.
11. Вывод EM-алгоритма для тематической модели pLSA. Генеративный процесс порождения коллекции тематической моделью LDA.
12. Кластеризация методом K-средних и K-медоид. Их вычислительная сложность. Иерархическая кластеризация сверху вниз.
13. Агломеративная кластеризация. Его сложность. Расстояния ближайшего соседа, дальнего соседа и усредненное расстояние. Пересчет матрицы попарных расстояний для каждого типа расстояния.
14. Алгоритм сеточной кластеризации и DB-scan.
15. Кластеризация по точкам максимума плотности и метод DENCLUE.

16. Алгоритм спектральной кластеризации. Лапласиан, его неотрицательная определенность.
17. Оценка качества кластеризации – коэффициент силуэта и индекс Калинского.
18. Анализ кластеризуемости данных по энтропии расстояния и статистике Хопкинса. Методы отбора признаков для кластеризации.
19. Многослойный персептрон. Основные ф-ции активации. Активации выходного слоя для задач регрессии и классификации. Идея аппроксимации нейросетями произвольных границ между классами и произвольных гладких функций.
20. Алгоритм обратного распространения ошибки (backpropagation algorithm)
21. Нелинейное снижение размерности: многомерное шкалирование, Isomap, Maximum variance unfolding.
22. Нелинейное снижение размерности: диффузионные карты, автокодировщик.
23. Нелинейное снижение размерности: локально-линейное вложение (LLE), Laplacian eigenmaps.
24. Baseline-алгоритм для коллаборативной фильтрации. Алгоритмы user-user и item-item. Какой из них применим в онлайн режиме?
25. Алгоритм разреженного сингулярного разложения для коллаборативной фильтрации.

## Теоретический минимум

1. Методы многоклассовой классификации бинарными классификаторами - подходы один против всех, каждый против каждого.
2. Разложение среднеквадратичной ошибки на смещение и дисперсию.
3. Стэкинг моделей.
4. Бэггинг и случайный лес.
5. Алгоритм градиентного бустинга.
6. Взаимная информация.
7. Алгоритм последовательного отбора признаков
8. EM-алгоритм в общем случае.
9. Генеративный процесс порождения коллекции тематической моделью pLSA и LDA.
10. Кластеризация методом K-средних. Иерархическая кластеризация сверху вниз.
11. Агломеративная кластеризация.
12. Кластеризация DB-scan.
13. Оценка качества кластеризации – коэффициент силуэта и индекс Калинского.
14. Многослойный персептрон. Основные ф-ции активации. Активации выходного слоя для задач регрессии и классификации. Автокодировщик.
15. Алгоритмы многомерного шкалирования и Isomap.
16. Алгоритмы user-user и item-item для коллаборативной фильтрации. Какой из них применим в онлайн режиме?