

Программа экзамена
по курсу «Прикладная статистика
в машинном обучении»

Д. Деркач, А. Артемов
Д. Мусаткина, В. Белавин

16 декабря 2019 г.

Глава 1

Правила проведения экзамена

1. Экзамен письменный и состоит из трёх частей.
2. **Теоретический минимум билета**
Теоретический минимум включает в себя 3 тестовых задания (3 балла).
3. **Теоретические вопросы билета**
Один вопрос по первой половине курса, второй – по второй половине курса.
Каждый вопрос стоит 2 балла(суммарно 4 балла).
4. **Практическое задание билета**
Случайная задача из курса. Стоимость – 3 балла.
5. **Правила проведения экзамена**
На подготовку билета отводится 120 минут.
6. Brace yourself.

Глава 2

Программа теоретического минимума

Тема 1: Оценки

1. Аксиомы Колмогорова. Байесовский и частотный подходы.
2. Теорема Байеса.
3. Метод моментов.
4. Метод максимального правдоподобия.
5. Оценка постериорного максимума.
6. Байесовский доверительный интервал. Маргинализация распределений.
7. Частотный интервал. Профильная функция правдоподобия.
8. Дельта-метод.

Тема 2: Тестирование гипотез

1. Постановка задачи тестирования гипотез.
2. ROC-кривая, площадь под кривой.
3. Критерий Вальда.
4. p -значение.
5. Критерий Неймана-Пирсона.
6. Лемма Неймана-Пирсона.
7. Критерий согласия Колмогорова-Смирнова.
8. Двухвыборочный критерий Стьюдента.
9. Двухвыборочный критерий знаков.
10. Множественное тестирование. Постановка задачи.
11. FWER. Коррекция Бонферрони.
12. FWER. Метод Холма.
13. FDR. Метод Бенджамини-Хохберга.

Тема 3: ANOVA

1. F-test и его свойства.
2. Постановка задачи метода ANOVA и его применение.
3. Условия применимости метода ANOVA и способы достижения условий применимости.
4. Анализ Краскела-Уоллиса: мотивация, алгоритм, свойства.
5. Апостериорные(post-hoc) тесты.
6. Анализ контрастов.
7. Fixed Effect Two-ways ANOVA.
8. ANOVA со случайными эффектами.
9. Дополнительные версии ANOVA: ANCOVA, MANOVA, ANOVA_{rm}.

Тема 4: GLM

1. Линейная регрессия и МНК.
2. Оценки дисперсии в МНК.
3. Прогнозирование и доверительные интервалы.
4. Оценка качества регрессии.
5. Множественная регрессия.
6. Методы выбора модели.
7. Регрессия с бинарными откликами и её преобразования.
8. Экспоненциальное семейство распределений.
9. Обобщённые линейные модели.
10. Оптимизация в GLM.
11. Доверительные интервалы и тест Вальда в GLM.
12. Оценка качества GLM.

Тема 5: Бутстрап

1. Стандартная постановка задачи и идея бутстрапа.
2. Оценка дисперсии на основе бутстрапа.
3. Бутстрапная оценка смещения.
4. Параметрический бутстрап.
5. Доверительное оценивание на основе бутстрапа.
6. Бэггинг.
7. Моделирование выборок из больших совокупностей.
8. Метод складного ножа.
9. Cross-Validation Leave-one-out.
10. Exact & Asymptotic LR Test Inversion.

Тема 6: GP & BO

1. Постановка задачи в гауссовых процессах и расчёт маргинального правдоподобия.
2. Постановка задачи в гауссовых процессах и использование для предсказаний на X_{test} .
3. Ядра используемые в гауссовых процессах и их влияние на предсказательную силу.
4. Обучение гауссова процесса.
5. Нестационарные Гауссовы процессы.
6. Байесовская оптимизация: постановка задачи, связь с гауссовыми процессами.
7. Варианты acquisition functions.