

**Федеральное государственное автономное образовательное  
учреждение высшего образования  
«Московский физико-технический институт  
(государственный университет)»**

**УТВЕРЖДАЮ  
Ректор МФТИ**

\_\_\_\_\_ **Н.Н. Кудрявцев**

**Программа государственного экзамена по специальности**

**по дисциплине:** **Интеллектуальный анализ данных**  
**по направлению:** Прикладные математика и физика (магистратура)  
**магистерская программа:** Интеллектуальный анализ данных  
**кафедра:** **Интеллектуальные системы**  
Физтех-школа прикладной математики и информатики

**курс:** 6  
**квалификация:** магистр

Семестр, формы государственной итоговой аттестации: 11 (Осенний) - Экзамен

**Программу составил:** К.В. Воронцов, д. ф.-м. наук, профессор РАН

**Программа обсуждена на заседании кафедры**

28 ноября 2018 г.

**СОГЛАСОВАНО:**

Заведующий кафедрой

К.В. Рудаков

Начальник учебного управления

И.Р. Гарайшина

Директор Физтех-школы

А.М. Райгородский

## 1. Цели и задачи

1.1 Государственный экзамен по специальности проводится в осеннюю экзаменационную сессию для студентов 6 курса магистратуры, в полном объеме выполнивших учебный план и не имеющих академических задолженностей по специальным дисциплинам, входящим в учебный план.

1.2 Целью государственного экзамена является определение соответствия результатов освоения обучающимися основной образовательной программы требованиям действующего образовательного стандарта.

1.3. Задачами государственного экзамена являются:

- оценка степени усвоения обучающимися теоретических положений общепрофессиональных и специальных дисциплин;
- оценка умения применять полученные знания и приобретенные навыки для решения конкретных задач;
- оценка знания и понимания перспектив развития отрасли.

## 2. Перечень примерных вопросов, выносимых на государственный экзамен, оценочные средства (виды и примеры контрольных заданий)

2.1 Перечень вопросов, выносимых на государственный экзамен:

1. Функциональная полнота систем функций алгебры логики. Замкнутые классы. Теорема Поста.
2. Предмет комбинаторики. Производящие функции. Коэффициенты и вычеты.
3. Машины Тьюринга, нормальные алгоритмы Маркова, примитивно-рекурсивные функции. Тезис Чёрча.
4. Группы и кольца. Теоремы о гомоморфизме групп, колец.
5. Байесовская теория классификации. Линейный дискриминант Фишера. Метод окна Парзена.
6. Линейные алгоритмы классификации. Логистическая регрессия и метод опорных векторов. Мультиколлинеарность и регуляризация.
7. Логические методы классификации. Решающие деревья и леса. Критерии информативности. Методы синтеза информативных конъюнкций.
8. Многослойные нейронные сети. Метод обратного распространения ошибки.
9. Внутренние и внешние критерии. Скользящий контроль и оценки обобщающей способности. Методы отбора признаков.
10. Методы построения композиций классификаторов. Бустинг и бэггинг.
11. Методы понижения размерности. Метод главных компонент. Неотрицательные матричные разложения.
12. Методы кластеризации: графовые, иерархические, статистические.
13. EM-алгоритм для разделения смесей распределений, классификации, кластеризации.
14. Общий вид алгоритмов, синтезируемых методами алгебраического подхода.
15. Основная схема построения алгебраических расширений. Требования к семействам отображений.
16. Декартовы и свободные произведения и прямые суммы систем множеств.
17. Симметрические и функциональные универсальные ограничения.
18. Условия полноты для моделей и семейств операций.
19. Задачи геометрического поиска и критерии его эффективности. Алгоритмы локализации точки в простом и выпуклом многоугольнике.
20. Локализация точки в планарном подразбиении при массовом запросе. Метод полос. Метод Киркпатрика.
21. Выпуклая оболочка конечного множества точек на плоскости. Алгоритм Джарвиса. Алгоритм Грэхема.

22. Алгоритмическая парадигма плоского заметания. Алгоритм вычисления пересечений конечного множества отрезков на плоскости, основанный на плоском заметании.
23. Разбиение Вороного и триангуляция Делоне. Свойства и взаимосвязь. Построение триангуляции Делоне методом жадной триангуляции.
24. Диаграмма Вороного линейных сегментов. Построение скелета многоугольной фигуры на основе диаграммы Вороного сегментов.
25. Двухкомпонентный случайный процесс со скрытой и наблюдаемой компонентами. Классификация задач оценивания скрытой компоненты по Норберту Винеру.
26. Скрытая марковская модель сигнала. Теорема о марковском свойстве апостериорного скрытого случайного процесса.
27. Оценивание скрытого марковского процесса для аддитивной функции потерь. Обобщенная процедура фильтрации-интерполяции случайного процесса со скрытой марковской компонентой.
28. Классическая процедура динамического программирования для минимизации парно-сепарабельных целевых функций. Функции Беллмана.
29. Оценивание скрытого марковского процесса для сингулярной функции потерь. Обобщенная процедура динамического программирования.
30. Линейная нормальная нестационарная модель сигнала со скрытой компонентой. Параметрическая процедура фильтрации-интерполяции. Фильтр-интерполятор Калмана-Бьюси.
31. Рекуррентные нейронные сети (RNN), проблема затухания градиентов. Управляемые рекуррентные блоки (GRU) сети долгой краткосрочной памяти (LSTM).
32. Задача языкового моделирования n-gram модели, нейросетевые модели.
33. Задача разметки последовательностей. Гибридная модель biLSTM-CRF.
34. Векторные представления слов. Дистрибутивная гипотеза и методы матричного разложения. Основные модели векторных представлений (Word2Vec, GloVe, fastText).
35. Векторные представления предложений и текстов. Основные модели (TF-IDF, усреднение, нейронные модели). Механизм Self-attention.
36. Тематическое моделирование. Модели PLSA и LDA.
37. Тематическое моделирование. Один из способов вывода EM-алгоритма, регуляризация и модальности, примеры регуляризаторов.
38. Машинный перевод как задача преобразования последовательностей. Статистические и нейросетевые методы. Способы работы с большим словарем.
39. Сверточные нейронные сети в применении к классификации текстов.
40. Глубокие контекстуальные представления слов (ELMO, BERT).

2.2 Государственный экзамен включает вопрос по тематике НИР и магистерской диссертации обучающегося.

2.3 Государственный экзамен включает следующие виды контрольных заданий: доклад, содержащий ответы на вопросы из перечня, дополнительные вопросы по специальности.

### 3. Перечень литературы для подготовки к государственному экзамену

1. Журавлев Ю. И., Флеров Ю. А. Дискретный анализ. Учебное пособие. – М.: МФТИ, 1999.
2. Грэхем Р., Кнут Д., Паташник О. Конкретная математика. Основание информатики. – М.: Мир, 1998.
3. Рассел С., Норвиг П. Искусственный интеллект: современный подход. М., Издательский дом «Вильямс». 2006 – 1408 с.
4. Воронцов К. В. Математические методы обучения по прецедентам. 2012. <http://www.machinelearning.ru>, «Машинное обучение (курс лекций, К.В.Воронцов)».
5. Мальцев А. И. Алгебраические системы. М.: Наука, 1970. 392 с.
6. Рудаков К. В. Алгебраическая теория универсальных и локальных ограничений для алгоритмов распознавания. Диссертация на соискание учёной степени доктора физико-математических наук. М.: ВЦ РАН, 1991.

7. Методы компьютерной обработки изображений. Под ред. В.А.Сойфера. М., Физматлит, 2003. – 780 с.
8. Местецкий Л. М. Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры // Москва, Физматлит, 2009, 288 с.
9. Моттль В. В., Мучник И. Б. Скрытые марковские модели в структурном анализе сигналов. М.: Наука, Физматлит, 1999, 351 с.
10. Липский В. Комбинаторика для программистов. – М.: Мир, 1988.
11. Эрдеш П., Спенсер Дж. Вероятностные методы в комбинаторике. М.: Мир. 1976.
12. Холл М. Комбинаторика. – М.: Мир, 1970.
13. Ф.Препарата, М.Шеймос. Вычислительная геометрия: введение. «Мир», Москва, 1989.
14. А.Ахо, Дж.Хопкрофт, Дж.Ульман. Построение и анализ вычислительных алгоритмов. «Мир», Москва, 1979.
15. М.Ласло. Вычислительная геометрия и компьютерная графика на С++. М., Бином, 1997.
16. Л.М.Местецкий. Вычислительная геометрия и графика. Конспект лекций.
17. Л.М.Местецкий. Непрерывная морфология бинарных изображений. М., Физматлит, 2009.
18. Dan Jurafsky and James H. Martin Speech and Language Processing (3rd ed. draft)
19. Stewen Bird et. al. Natural Language Processing with Python. 2-nd edition. 2016.
20. Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. Автоматическая обработка текстов на естественном языке и анализ данных. НИУ ВШЭ, 2017.
21. Ian Goodfellow, Yoshua Bengio, Aaron Courville. [Deep Learning](#), MIT Press, 2016.
22. Deep contextualized word representations, Peters и др., NAACL-HLT, 2018

#### 4. Рекомендации обучающимся по подготовке к государственному экзамену

4.1 Рекомендации обучающимся по подготовке к государственному экзамену: повторить материал лекций по прилагаемому списку литературы.

#### 5. Методика и критерии оценки государственного экзамена

5.1 Методика и критерии оценки государственного экзамена.

По итогам государственного экзамена выставляется оценка за по 10-балльной шкале.

Оценка	Критерии
Отлично (10)	Правильный, четкий и уверенный ответ на вопросы билета и дополнительные вопросы.
Отлично (9)	Даны правильные ответы на вопросы билета и дополнительные вопросы после некоторых оговорок и запинок.
Отлично (8)	Даны ответы на вопросы билета и дополнительные вопросы после небольших исправлений и наводящих вопросов экзаменаторов.
Хорошо (7)	Даны ответы на вопросы билета, но нет верного ответа на один из дополнительных вопросов.
Хорошо (6)	Есть недочеты в ответе на один из вопросов билета и нет верного ответа на один из дополнительных вопросов.
Хорошо (5)	Есть недочеты в ответах на вопросы билета и нет верного ответа на один из дополнительных вопросов.

Удовлетворительно(4)	Есть недочеты в ответах на вопросы билета или нет ответа ни на один из дополнительных вопросов.
Удовлетворительно(3)	Нет ответа на один из вопросов билета, но есть ответы на дополнительные вопросы (возможно с недочетами).
Неудовлетворительно (2)	Нет ответа на один из вопросов билета и на дополнительные вопросы.
Неудовлетворительно (1)	Нет ответа ни на один из вопросов билета.

Итоговая оценка за государственный экзамен по специальности определяется государственной экзаменационной комиссией.

## **6. Процедура проведения государственного экзамена**

6.1 Порядок проведения государственного экзамена осуществляется в соответствии с локальными нормативными актами МФТИ.

6.2 Продолжительность государственного экзамена: время на подготовку 1 час, время на ответ 20 минут.

6.4 Использование технических средств не допускается.

6.5 Состав экзаменационного билета: 3 вопроса из перечня.



