

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

**ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ
ОТДЕЛЕНИЕ ЭКОНОМИКИ
Кафедра Системного анализа и информатики**

УТВЕРЖДЕНА
на заседании кафедры
Системного анализа и информатики
Протокол от «1» сентября 2017г. № 1

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Индекс Б1.В.ДВ.1.2 «Байесовские методы»

по направлению подготовки 38.04.01 «Экономика»

направленность «Системы больших данных в экономике»

квалификация магистр

очная форма обучения

Год набора - 2017

Москва, 2017г.

Автор(ы)—составитель(и):

к.т.н., доцент кафедры Системного анализа и информатики Стефановский Д.В.

Заведующий кафедрой

Системного анализа и информатики, к.т.н., доцент, Маруев С.А.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине соотнесенных с планируемыми результатами освоения образовательной программы.....	4
2. Объем и место дисциплины в структуре образовательной программы.....	5
3. Содержание и структура дисциплины.....	5
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине.....	7
5. Методические указания для обучающихся по освоению дисциплины	13
6. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине	16
6.1. Основная литература	16
6.2. Дополнительная литература	16
6.3. Учебно-методическое обеспечение самостоятельной работы	16
6.4. Нормативные правовые документы	16
6.5. Интернет-ресурсы	17
6.6. Иные источники	17
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы	17

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

1.1. Дисциплина Б1.В.ДВ.1.2 «Байесовские методы» обеспечивает совершенствование следующими компетенциями:

Код компетенции	Наименование компетенции	Код этапа совершенствования компетенции	Наименование этапа освоения компетенции
ПК-9	Способность анализировать и использовать различные источники информации для проведения экономических расчетов	ПК-9.2У	Способен к разработке, проверке, оценке используемых моделей

1.2. В результате освоения дисциплины у студентов должны быть сформированы:

Профессиональные действия	Код этапа освоения компетенции	Результаты обучения
Проведение аналитического исследования в соответствии с согласованными требованиями	ПК-9.2У	<p>Знать:</p> <ul style="list-style-type: none"> Алгоритмы машинного обучения: обучение с учителем, обучение без учителя, полууправляемое обучение, обучение с подкреплением Определять теоретические верхние оценки переобученности: сложность, разделимость, устойчивость Решать проблемы переобучения и недообучения алгоритма Формировать предложения по использованию результатов анализа Машинное обучение: классификация, кластеризация, обнаружение выбросов, фильтрация Методы и модели классификации: логистическая регрессия, деревья решений, предредукция, постредукция, модели, основанные на правилах, наивный байесовский алгоритм, теорема Байеса, усиление энтропии информации Фильтрация шумовых выбросов, виды шумовых выбросов: глобальный, контекстуальный, коллективный Анализ изображений: тепловые карты, анализ сетей, анализ пространственных данных, анализ временных рядов Методы идентификации шаблонов Методы оценки моделей: оценка качества построенной модели по тестовой выборке и анализ обобщающих способностей алгоритма Распределенный анализ данных Анализ данных в реальном времени

		Уметь: <ul style="list-style-type: none"> • Осуществлять поиск информации о новых и перспективных методах анализа больших данных, сравнительный анализ методов
--	--	--

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

3 ЗЕ, 32 ак. часа на контактную работу с преподавателем, 76 ак. часов на самостоятельную работу обучающихся;

Место дисциплины в структуре ОП ВО

- Б1.В.ДВ.1.2 «Байесовские методы», 2 курс, 3 семестр;
- дисциплина реализуется после изучения дисциплин:
 - математический анализ (в объеме бакалавриата),
 - линейная алгебра (в объеме бакалавриата),
 - экономическая информатика (в объеме бакалавриата),
 - дискретная математика (в объеме бакалавриата),
 - теория вероятности и математическая статистика (в объеме бакалавриата),
- форма промежуточной аттестации – зачет с оценкой.

3. Содержание и структура дисциплины

Таблица 2.

№ п/п	Наименование тем (разделов)	Объем дисциплины, ак. час/час						Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л	ЛР	ПЗ	КСР		
Очная форма обучения								
Тема 1	Байесовский подход к теории вероятностей. Примеры байесовских рассуждений. Сопряжённые распределения, аналитический байесовский вывод, экспоненциальный класс распределений	13	2	2			9	Опрос
Тема 2	Байесовский выбор модели. Решение задачи выбора модели по Байесу. Обоснованность модели	13	2	2			9	Опрос
Тема 3	Метод релевантных векторов для задачи регрессии	14	2	2			10	Опрос

№ п/п	Наименование тем (разделов)	Объем дисциплины, ак. час/час						Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л	ЛР	ПЗ	КСР		
Очная форма обучения								
Тема 4	Метод релевантных векторов для задачи классификации и ЕМ-алгоритм. Байесовский метод главных компонент	14	2	2			10	Опрос
Тема 5	Вариационный подход как приближенный способ байесовского вывода	13	2	2			9	Опрос
Тема 6	Методы Монте Карло по схеме марковский цепей (MCMC)	14	2	2			10	Опрос
Тема 7	Непараметрические байесовские методы. Процессы Дирихле	14	2	2			10	Опрос
Тема 8	Латентное размещение Дирихле (LDA). Тематические модели для коллекций текстов	13	2	2			9	Опрос ДЗ (1-8), КР (1-8)
Промежуточная аттестация				-			-	Зачет с оценкой
Всего:		108/81	16/12	16/12			76/57	

Примечание – формы текущего контроля успеваемости: контрольная работа (КР), домашнее задание (ДЗ)

Содержание дисциплины

Тема 1. Байесовский подход к теории вероятностей. Примеры байесовских рассуждений. Сопряжённые распределения, аналитический байесовский вывод, экспоненциальный класс распределений

Различия между байесовскими и частотными подходами в теории вероятностей. Пример байесовских рассуждений. Правила суммирования и произведения вероятностей. Формула Байеса. Условная независимость. Два подхода к теории вероятностей: частотный подход и Байесовский подход. Связь между байесовским подходом и булевой логикой. Пример вероятностных рассуждений.

Тема 2. Байесовский выбор модели. Решение задачи выбора модели по Байесу. Обоснованность модели.

Полный байесовский вывод. Пример использования априорных знаний. Сопряженные распределения. Вероятностная модель обучения. Иерархическая схема Байеса.

Принцип наибольшей обоснованности. Обоснованность модели. Примеры использования.

Тема 3. Метод релевантных векторов для задачи регрессии

Обобщенные линейные модели. Метод максимума правдоподобия. Введение регуляризации. Метод релевантных векторов. Метод релевантных векторов для задачи регрессии.

Метод максимума правдоподобия (логистическая регрессия). Оптимизация функции правдоподобия (IRLS). Метод релевантных векторов. Приближение Лапласа. Вычисление обоснованности. Метод релевантных векторов для задачи классификации.

Тема 4. Метод релевантных векторов для задачи классификации и ЕМ-алгоритм. Байесовский метод главных компонент.

Метод оптимизации Ньютона. ЕМ-алгоритм в общем виде. ЕМ-алгоритм для разделения гауссовской смеси. Логистическая и мультиномиальная регрессия. Метод релевантных векторов для задачи классификации.

Задача уменьшения размерности в данных. Метод главных компонент. ЕМ-алгоритм для PCA. Учет пропусков в данных. Байесовский метод главных компонент. Вероятностная смесь главных компонент. Применение смеси главных компонент.

Тема 5. Вариационный подход как приближенный способ байесовского вывода

Гамма-распределение. Вывод в вероятностных моделях. Примеры вероятностных моделей: линейная, логистическая регрессии, смесь нормальных распределений. Нижняя оценка обоснованности. Минимизация в семействе факторизованных распределений. Пример применения вариационного подхода для модели линейной регрессии.

Дивергенция Кульбака-Лейблера. Вариационный ЕМ-алгоритм. Распределение Джеффриса. Вариационная линейная регрессия.

Тема 6 Методы Монте Карло по схеме марковский цепей (МСМС)

Методы Монте Карло в байесовском подходе. Простейшие методы генерации. Идея МСМС. Теоретические свойства марковских цепей. Схема Метрополиса-Хастингса. Схема Гиббса. Применение схемы Гиббса для дискретной марковской сети. Оценка нормировочной константы распределения с помощью схемы Гиббса.

Тема 7. Непараметрические байесовские методы. Процессы Дирихле

Распределение Дирихле и его свойства. Генерация выборки из распределения Дирихле. Набор гамма-распределений. “Ломка палки” (Stick-breaking). Урновая схема.

Процессы Дирихле. Условные распределения. Представления процесса Дирихле.

Смесь распределений с априорным распределением, заданным процессом Дирихле. Схема МакИчерна. Коллапсированная схема МакИчерна. Вариационный вывод.

Обобщения процесса Дирихле. Приложения процессов Дирихле и их расширений.

Тема 8. Латентное размещение Дирихле (LDA). Тематические модели для коллекций текстов.

Распределение Дирихле. Модель LDA. Вариационный подход для вывода в модели LDA. Схема Гиббса для вывода в модели LDA. Способы применения тематической модели LDA.

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине

4.1. Формы и методы текущего контроля успеваемости.

4.1.1. В ходе реализации дисциплины Б1.В.ДВ.1.2 «Байесовские методы» используются следующие методы текущего контроля успеваемости обучающихся:

Тема	Методы текущего контроля успеваемости
Тема 1	Опрос
Тема 2	Опрос
Тема 3	Опрос
Тема 4	Опрос
Тема 5	Опрос
Тема 6	Опрос
Тема 7	Опрос
Тема 8	Опрос, домашнее задание по темам 1-8, контрольная работа по темам 1-8

4.1.2. Зачет проводится с применением следующих методов (средств):

Зачет с оценкой проводится в виде письменной контрольной работы.

4.2. Материалы текущего контроля успеваемости обучающихся

Типовые оценочные материалы по теме 1

Опрос:

- Раскройте различие между байесовскими и частотными подходами в теории вероятностей на примерах.
- Раскройте суть байесовских рассуждений, приведите пример.
- Раскройте суть правила суммирования и произведения вероятностей, продемонстрируйте правило на примере.
- Приведите пример использования формулы Байеса, дайте определение условной независимости.
- Раскройте суть частотного и Байесовского подхода к теории вероятности.
- Покажите связь между байесовским подходом и булевой логикой и приведите примеры вероятностных рассуждений.

Типовые оценочные материалы по теме 2

Опрос:

- Раскройте суть полного байесовского вывода и продемонстрируйте ее на примере.
- Раскройте суть сопряженных распределений и продемонстрируйте ее на примере.
- Раскройте особенности и суть вероятностной модели обучения.
- Раскройте особенности и суть иерархическая схема Байеса, приведите пример иерархической схемы Байеса.
- Раскройте суть и принципы наибольшей обоснованности модели, приведите примеры использования.

Типовые оценочные материалы по теме 3

Опрос:

- Раскройте особенности и суть обобщенных линейных моделей, приведите пример.
- Раскройте особенности и суть метода максимума правдоподобия, приведите пример.
- Раскройте особенности и суть метода релевантных векторов для задачи регрессии, приведите пример.
- Раскройте особенности и суть метода релевантных векторов для задачи классификации, приведите пример.

Типовые оценочные материалы по теме 4

Опрос:

- Раскройте особенности и суть метода оптимизации Ньютона, приведите пример.
- Раскройте особенности и суть ЕМ-алгоритма в общем виде и приведите пример.

- Раскройте суть и особенности метода релевантных векторов для задачи классификации.
- Раскройте суть задачи уменьшения размерности в данных, приведите пример.
- Раскройте суть метода главных компонент, приведите пример.
- Раскройте суть метода вероятностной смеси главных компонент, приведите пример.

Типовые оценочные материалы по теме 5

Опрос:

- Раскройте суть вероятностных моделей, приведите пример линейной регрессионной вероятностной модели.
- Раскройте суть нижней оценка обоснованности, минимизации в семействе факторизованных распределений. Приведите пример применения вариационного подхода для модели линейной регрессии.
- Дивергенция Кульбака-Лейблера. Вариационный EM-алгоритм. Распределение Джеффриса. Вариационная линейная регрессия.

Типовые оценочные материалы по теме 6

Опрос:

- Методы Монте Карло в байесовском подходе.
- Покажите на примере простейшие методы генерации.
- Покажите на примере теоретические свойства марковских цепей.
- Покажите на примере использование марковских сетей.
- Раскройте суть схемы Метрополиса-Хастингса. Покажите на примере
- Покажите на примере применение схемы Гиббса для дискретной марковской сети.
- Оценка нормировочной константы распределения с помощью схемы Гиббса.

Типовые оценочные материалы по теме 7

Опрос:

- Распределение Дирихле и его свойства. Генерация выборки из тьраспределения Дирихле. Покажите на примере.
- Набор гамма-распределений. “Ломка палки” (Stick-breaking). Урновая схема. Покажите на примерах.
- Процессы Дирихле. Условные распределения. Представления процесса Дирихле. Покажите на примере использование процессов Дирихле.
- Смесь распределений с априорным распределением, заданным процессом Дирихле. Покажите на примере.
- Схема МакИчерна. Коллапсированная схема МакИчерна. Покажите на примере.
- Обобщения процесса Дирихле. Приведите примеры приложений процессов Дирихле и их расширений.

Типовые оценочные материалы по теме 8

Опрос:

- Распределение Дирихле. Модель LDA. Покажите на примере.
- Вариационный подход для вывода в модели LDA. Покажите на примере.
- Схема Гиббса для вывода в модели LDA. Покажите на примере.
- Способы применения тематической модели LDA. Покажите на примере.

Домашнее задание:

При выполнении домашнего задания студентам предлагается разработать программу, решающую одну из задач использование одного из методов, предложенных на лекциях.

Студенты самостоятельно выбирает источники данных для разрабатываемой программы. Если преподаватель считает выбранные источники достаточными для

успешного выполнения задания, то студент может приступить к выполнению задания. В противном случае студент должна устранить все замечания преподавателя и повторно представить тему на согласование.

После согласования темы, студентом проводится анализ с разработкой необходимых программ и алгоритмов.

Например, домашняя работа по теме «Латентное размещение Дирихле(LDA)» представляется в письменной форме и содержит следующие разделы:

1. Для набора объектов производится расчет основных показателей LDA.
2. Подбирается и, если необходимо, модифицируется алгоритм(ы) рассматриваемый(ые) в рамках темы.
3. Представляются в письменном виде результаты машинного обучения, модифицированные алгоритмы и программы, также указываются параметры качества параметра ошибок.

Контрольная работа:

Вопрос 1. На примере набора данных предложенных преподавателем (матрица размерностью не более 5x40) покажите использование метода максимума правдоподобия.

Вопрос 2. Расскажите о методе главных компонент и расскажите в каких случаях его надо использовать.

4.3. Оценочные средства для промежуточной аттестации.

4.3.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы. Показатели и критерии оценивания компетенций с учетом этапа их формирования.

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПК-9	Способность анализировать и использовать различные источники информации для проведения экономических расчетов	ПК-9.2У	Способен к разработке, проверке, оценке используемых моделей

Этап освоения компетенции	Показатель оценивания	Критерий оценивания
ПК-9.2У	Способен к разработке, проверке, оценке используемых моделей	Понимает и объясняет содержание и последовательность выполнения этапов разработки, проверки и оценки используемых моделей

4.3.2. Типовые оценочные средства.

Вопросы к зачету:

- Раскройте различие между байесовскими и частотными подходами в теории вероятностей на примерах.
- Раскройте суть байесовских рассуждений, приведите пример.

- Раскройте суть правила суммирования и произведения вероятностей, продемонстрируйте правило на примере.
- Приведите пример использования формулы Байеса, дайте определение условной независимости.
- Раскройте суть частотного и Байесовского подхода к теории вероятности.
- Покажите связь между байесовским подходом и булевой логикой и приведите примеры вероятностных рассуждений.
- Раскройте суть полного байесовского вывода и продемонстрируйте ее на примере.
- Раскройте суть сопряженных распределений и продемонстрируйте ее на примере.
- Раскройте особенности и суть вероятностной модели обучения.
- Раскройте особенности и суть иерархическая схема Байеса, приведите пример иерархической схемы Байеса.
- Раскройте суть и принципы наибольшей обоснованности модели, приведите примеры использования.
- Раскройте особенности и суть обобщенных линейных моделей, приведите пример.
- Раскройте особенности и суть метода максимума правдоподобия, приведите пример.
- Раскройте особенности и суть метода релевантных векторов для задачи регрессии, приведите пример.
- Раскройте особенности и суть метода релевантных векторов для задачи классификации, приведите пример.
- Раскройте особенности и суть метода оптимизации Ньютона, приведите пример.
- Раскройте особенности и суть ЕМ-алгоритма в общем виде и приведите пример.
- Раскройте суть и особенности метода релевантных векторов для задачи классификации.
- Раскройте суть задачи уменьшения размерности в данных, приведите пример.
- Раскройте суть метода главных компонент, приведите пример.
- Раскройте суть метода вероятностной смеси главных компонент, приведите пример.
- Раскройте суть вероятностных моделей, приведите пример линейной регрессионной вероятностной модели.
- Раскройте суть нижней оценка обоснованности, минимизации в семействе факторизованных распределений. Приведите пример применения вариационного подхода для модели линейной регрессии.
- Дивергенция Кульбака-Лейблера. Вариационный ЕМ-алгоритм. Распределение Джеффриса. Вариационная линейная регрессия.
- Методы Монте Карло в байесовском подходе.
- Покажите на примере простейшие методы генерации.
- Покажите на примере теоретические свойства марковских цепей.
- Покажите на примере использование марковских сетей.
- Раскройте суть схемы Метрополиса-Хастингса. Покажите на примере
- Покажите на примере применение схемы Гиббса для дискретной марковской сети.
- Оценка нормировочной константы распределения с помощью схемы Гиббса.
- Распределение Дирихле и его свойства. Генерация выборки из т-распределения Дирихле. Покажите на примере.
- Набор гамма-распределений. “Ломка палки” (Stick-breaking). Урновая схема. Покажите на примерах.
- Процессы Дирихле. Условные распределения. Представления процесса Дирихле. Покажите на примере использование процессов Дирихле.
- Смесь распределений с априорным распределением, заданным процессом Дирихле. Покажите на примере.
- Схема МакИчерна. Коллапсированная схема МакИчерна. Покажите на примере.

- Обобщения процесса Дирихле. Приведите примеры приложений процессов Дирихле и их расширений.
- Распределение Дирихле. Модель LDA. Покажите на примере.
- Вариационный подход для вывода в модели LDA. Покажите на примере.
- Схема Гиббса для вывода в модели LDA. Покажите на примере.
- Способы применения тематической модели LDA. Покажите на примере.
- На примере набора данных предложенных преподавателем (матрица размерностью не более 5x40) покажите использование метода максимума правдоподобия.
- Расскажите о методе главных компонент и расскажите в каких случаях его надо использовать.

Шкала оценивания.

Оценка определяется по формуле: $\frac{1}{4}$ опрос + $\frac{1}{4}$ д.з. + $\frac{1}{4}$ к.р. + $\frac{1}{4}$ сдача зачета.

10- бальная шкала	Традиционн ая шкала	«Зачтено»/ «Не зачтено»	Определение
10	Отлично	Зачтено	Полные, глубокие и систематические знания, полный и правильный ответ на теоретический вопрос, полное и правильное решение задачи.
9	Отлично	Зачтено	Глубокие и систематические знания, правильный ответ на теоретический вопрос, правильное решение задачи.
8	Отлично	Зачтено	Систематические знания, правильный ответ на теоретический вопрос, правильное решение задачи.
7	Хорошо	Зачтено	Систематические знания, правильный ответ на теоретический вопрос с незначительными неточностями, правильное решение задачи.
6	Хорошо	Зачтено	Систематические знания, правильный ответ на теоретический вопрос с незначительными неточностями, правильное решение задачи с незначительными неточностями.
5	Удовлетвори тельно	Зачтено	Ответ на теоретический вопрос неполный, правильное решение задачи с незначительными неточностями.
4	Удовлетвори тельно	Зачтено	Ответ на теоретический вопрос неполный, решение задачи содержит арифметические ошибки, не влияющие на правильность хода решения задачи.
3	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос неполный, решение задачи содержит идеологические ошибки.
2	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос неверный и/или решение задачи содержит идеологические ошибки.
1	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос неверный и решение задачи отсутствует.
0	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос отсутствует и решение задачи отсутствует.

4.4. Методические материалы по проведению промежуточной аттестации

Зачет проводится в аудитории. Отсчет времени, отведенного на письменную работу, идет по завершении процедуры размещения студентов и раздачи заданий.

Студент обязан являться на письменный контроль в указанное в расписании время. В случае опоздания время, отведенное на письменный контроль знаний, не продлевается.

При себе студенты могут иметь только письменные принадлежности. Необходимую для выполнения работы бумагу выдает преподаватель.

Преподаватель раздает варианты работы, содержащий 2 вопроса. Листы с заданиями должны быть повернуты текстом вниз, чтобы студенты до окончания процедуры раздачи не могли начать выполнение работы. По окончании раздачи вариантов студентам разрешается перевернуть текст задания и одновременно приступить к выполнению работы. По окончании отведенного времени студенты одновременно заканчивают выполнение работы. Если работа завершена существенно раньше срока, то по разрешению преподавателя студент может покинуть аудиторию досрочно.

Мобильные телефоны должны быть выключены и убраны со столов, допускается использование калькуляторов, выполняющих только простые арифметические вычисления.

Во время проведения письменного контроля знаний студентам не разрешается пользоваться учебными программами, справочниками и прочими источниками информации.

Использование материалов, а также попытка общения с другими студентами или иными лицами, в том числе с применением электронных средств связи, несанкционированные перемещения и т.п. являются основанием для удаления студента из аудитории и последующего проставления в ведомость оценки «неудовлетворительно».

Во время проведения письменного контроля знаний студентам разрешается покинуть аудиторию только при условии сдачи работы в объеме, выполненном к моменту выхода из аудитории. Дальнейшее продолжение работы запрещается.

Ответы в работе без объяснений не засчитываются. Рисунки должны быть четкими, все линии графиков, используемых при ответах на вопросы задач, должны быть подписаны.

Продолжительность экзаменационной письменной работы 120 минут.

5. Методические указания для обучающихся по освоению дисциплины

Любой вид занятий, создающий условия для зарождения самостоятельной мысли, познавательной и творческой активности студента связан с самостоятельной работой. В широком смысле под самостоятельной работой понимают совокупность всей самостоятельной деятельности студентов как в учебной аудитории, так и вне ее, в контакте с преподавателем и в его отсутствие. Самостоятельная работа может реализовываться: непосредственно в процессе аудиторных занятий – на лекциях, практических и семинарских занятиях, при выполнении контрольных и лабораторных работ и др.; в контакте с преподавателем вне рамок аудиторных занятий – на консультациях по учебным вопросам, в ходе творческих контактов, при ликвидации задолженностей, при выполнении индивидуальных заданий и т.д.; в библиотеке, дома, в общежитии, на кафедре и других местах при выполнении студентом учебных и творческих заданий.

Лекции

Главное в период подготовки к лекционным занятиям – научиться методам самостоятельного умственного труда, сознательно развивать свои творческие способности

и овладевать навыками творческой работы. Для этого необходимо строго соблюдать дисциплину учебы и поведения. Четкое планирование своего рабочего времени и отдыха является необходимым условием для успешной самостоятельной работы. В основу его нужно положить рабочие программы изучаемых в семестре дисциплин. Каждому студенту следует составлять еженедельный и семестровый планы работы, а также план на каждый рабочий день. С вечера всегда надо распределять работу на завтрашний день. В конце каждого дня целесообразно подводить итог работы: тщательно проверить, все ли выполнено по намеченному плану, не было ли каких-либо отступлений, а если были, по какой причине это произошло. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Если что-то осталось невыполненным, необходимо изыскать время для завершения этой части работы, не уменьшая объема недельного плана.

Самостоятельная работа на лекции. Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность студента. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое существенное, основное и сделано это самим студентом. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Желательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях. Конспект лекции лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось присить их у однокурсников и тем самым не отвлекать их во время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

Семинар и проведение опроса

Каждый студент должен начать с ознакомления с планом занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованную к данной теме. На основе индивидуальных предпочтений студенту необходимо самостоятельно выбрать тему доклада по проблеме семинара и по возможности подготовить по нему презентацию. Если программой дисциплины предусмотрено выполнение практического задания, то его необходимо выполнить с учетом предложенной инструкции (устно или письменно). Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы семинара, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Работа с литературными источниками.

В процессе подготовки к семинарским занятиям, студентам необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических

изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у студентов свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем по каждой теме семинарского или практического занятия, что позволяет студентам проявить свою индивидуальность в рамках выступления на данных занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Методические указания по выполнению самостоятельной работы:

Тема 1. Введение в машинное обучение. Цели и основная проблематика машинного обучения.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Различия между байесовскими и частотными подходами в теории вероятностей. Пример байесовских рассуждений. Правила суммирования и произведения вероятностей. Формула Байеса. Условная независимость. Два подхода к теории вероятностей: частотный подход и Байесовский подход. Связь между байесовским подходом и булевой логикой. Пример вероятностных рассуждений.

Тема 2. Байесовский выбор модели. Решение задачи выбора модели по Байесу. Обоснованность модели.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Полный байесовский вывод. Пример использования априорных знаний. Сопряженные распределения. Вероятностная модель обучения. Иерархическая схема Байеса.

Принцип наибольшей обоснованности. Обоснованность модели. Примеры использования.

Тема 3. Метод релевантных векторов для задачи регрессии

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Обобщенные линейные модели. Метод максимума правдоподобия. Введение регуляризации. Метод релевантных векторов. Метод релевантных векторов для задачи регрессии.

Метод максимума правдоподобия (логистическая регрессия). Оптимизация функции правдоподобия (IRLS). Метод релевантных векторов. Приближение Лапласа. Вычисление обоснованности. Метод релевантных векторов для задачи классификации.

Тема 4. Метод релевантных векторов для задачи классификации и ЕМ-алгоритм. Байесовский метод главных компонент.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Метод оптимизации Ньютона. ЕМ-алгоритм в общем виде. ЕМ-алгоритм для разделения гауссовской смеси. Логистическая и мультиномиальная регрессия. Метод релевантных векторов для задачи классификации.

Задача уменьшения размерности в данных. Метод главных компонент. ЕМ-алгоритм для РСА. Учет пропусков в данных. Байесовский метод главных компонент. Вероятностная смесь главных компонент. Применение смеси главных компонент.

Тема 5. Вариационный подход как приближенный способ байесовского вывода

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Гамма-распределение. Вывод в вероятностных моделях. Примеры вероятностных моделей: линейная, логистическая регрессии, смесь нормальных распределений. Нижняя

оценка обоснованности. Минимизация в семействе факторизованных распределений. Пример применения вариационного подхода для модели линейной регрессии.

Дивергенция Кульбака-Лейблера. Вариационный ЕМ-алгоритм. Распределение Джеффриса. Вариационная линейная регрессия.

Тема 6 Методы Монте Карло по схеме марковский цепей (МСМС)

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Методы Монте Карло в байесовском подходе. Простейшие методы генерации. Идея МСМС. Теоретические свойства марковских цепей. Схема Метрополиса-Хастингса. Схема Гиббса. Применение схемы Гиббса для дискретной марковской сети. Оценка нормировочной константы распределения с помощью схемы Гиббса.

Тема 7. Непараметрические байесовские методы. Процессы Дирихле

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Распределение Дирихле и его свойства. Генерация выборки из распределения Дирихле. Набор гамма-распределений. “Ломка палки” (Stick-breaking). Урновая схема.

Процессы Дирихле. Условные распределения. Представления процесса Дирихле.

Смесь распределений с априорным распределением, заданным процессом Дирихле. Схема МакИчерна. Коллапсированная схема МакИчерна. Вариационный вывод.

Обобщения процесса Дирихле. Приложения процессов Дирихле и их расширений.

Тема 8. Латентное размещение Дирихле (LDA). Тематические модели для коллекций текстов.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Распределение Дирихле. Модель LDA. Вариационный подход для вывода в модели LDA. Схема Гиббса для вывода в модели LDA. Способы применения тематической модели LDA.

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Основная литература.

1. Ветров Д.П., Кропотов Д.А. Байесовские методы машинного обучения, учебное пособие по спецкурсу, 2007

6.2. Дополнительная литература.

1. Бринк Хенрик, Джозеф Ричардс, Феверолф Марк, Машинное обучение. – СПб.: Питер, 2017. -336 с.: ил. – (Серия «Библиотека программиста»). ISBN 978-5-496-02989-6
2. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2016. 302 с.

6.3. Учебно-методическое обеспечение самостоятельной работы.

Отдельное обеспечение не предусмотрено.

6.4. Нормативные правовые документы.

Не предусмотрены.

6.5. Интернет-ресурсы.

1. <http://citeseer.ist.psu.edu/> — основной источник знаний по Computer Science, по многим статьям есть полные тексты
2. <http://citeseer.cs.msu.su/> — русскоязычная электронная библиотека научных статей
3. <http://arxiv.org/> — библиотека электронных публикаций, в основном по физике, но доля «Computer Science» в последнее время стремительно увеличивается
4. <http://rexa.info/> — библиографическая поисковая система по статьям, авторам и грантам
5. <http://elibrary.ru/> — российская научная электронная библиотека
6. <http://iinwww.ira.uka.de/bibliography/index.html>
7. библиографическая база данных для работы с BibTeX
8. <http://www.gotai.net/> -- русскоязычный сайт об искусственном интеллекте
9. Math-Net.ru -- общероссийский математический портал

6.6. Иные источники.

Не предусмотрены.

7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Для лекций:

1. Персональный компьютер
2. Мультимедийный проектор
3. Доска, мел или маркеры

Для лабораторных занятий:

1. Компьютерный класс,
2. Виртуальная машина Ubuntu 15.04 b выше с установленным Postgresql и MongoDB
3. Мультимедийный проектор
4. Доска, маркеры
5. Компилятор R-2.15.1 – GNU - <http://www.r-project.org/> либо интегрированная среда разработки RStudio – GNU AGP - <http://www.rstudio.com/ide/>.
6. Jupyter Notebook - бесплатная интерактивная оболочка для языка программирования Python, позволяющая объединить код, текст и диаграммы.
7. Компилятор Scala – <http://www.scala-lang.org/>
8. Программный комплекс анализа новостного сайта - "Crawler-Persona"
9. База данных "Централизация государственных закупок в 2014 г".
10. База данных учебно-методических материалов по дисциплине "Макроэкономика".
11. База данных Бюджетная и социально-экономическая статистика субъектов Российской Федерации.