

**Санкт-Петербургский государственный университет**

**РАБОЧАЯ ПРОГРАММА  
УЧЕБНОЙ ДИСЦИПЛИНЫ**

Основы байесовского вывода (семинар), тр 7 сем  
Introduction to Bayesian Derivation

**Язык(и) обучения**

русский

Трудоемкость в зачетных единицах: 2

Регистрационный номер рабочей программы: 044992

## **Раздел 1. Характеристики учебных занятий**

### **1.1. Цели и задачи учебных занятий**

Сообщение сведений об основных концепциях машинного обучения и байесовского вывода в объёме, необходимом для общего развития и изучения смежных дисциплин физико-математического цикла. Усвоение основных идей, понятий и фактов машинного обучения, основных концепций байесовского вывода.

### **1.2. Требования подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)**

Владение курсами «Теоретическая информатика», «Дискретная теория вероятностей».

### **1.3. Перечень результатов обучения (learning outcomes)**

Обучающийся должен овладеть теоретическим материалом в объеме, предусмотренном программой, уметь применять полученные знания при решении теоретических и прикладных задач на основе анализа освоенных разделов; уяснить логику и технику построения теории байесовского вывода как фундамента для понимания современных исследований в области машинного обучения, а также для дальнейших самостоятельных исследований. Обучающийся также должен овладеть практическими навыками по основам анализа данных, построению и обучению моделей на одном из языков программирования, решению прикладных задач анализа данных.

### **1.4. Перечень и объём активных и интерактивных форм учебных занятий**

Промежуточная аттестация (зачет) 4 часа, семинары 30 часов.

## Раздел 2. Организация, структура и содержание учебных занятий

### 2.1. Организация учебных занятий

#### 2.1.1 Основной курс

Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся																	
Код модуля в составе дисциплины, практики и т.п.	Контактная работа обучающихся с преподавателем											Самостоятельная работа			Объём активных и интерактивных форм учебных занятий	Трудоёмкость	
	лекции	семинары	консультации	практические занятия	лабораторные работы	контрольные работы	коллоквиумы	текущий контроль	промежуточная аттестация	итоговая аттестация	под руководством преподавателя	в присутствии преподавателя	сам. раб. с использованием методических материалов	текущий контроль (сам.раб.)			промежуточная аттестация (сам.раб.)
<b>ОСНОВНАЯ ТРАЕКТОРИЯ</b>																	
<b>очная форма обучения</b>																	
Семестр 7		30						2				34		6		32	2
		10-25						2-50				1-1		1-1			
ИТОГО		30						2				34		6		32	2

Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся																	
Код модуля в составе дисциплины, практики и т.п.	Контактная работа обучающихся с преподавателем											Самостоятельная работа			Объём активных и интерактивных форм учебных занятий	Трудоёмкость	
	лекции	семинары	консультации	практические занятия	лабораторные работы	контрольные работы	коллоквиумы	текущий контроль	промежуточная аттестация	итоговая аттестация	под руководством преподавателя	в присутствии преподавателя	сам. раб. с использованием методических материалов	текущий контроль (сам.раб.)			промежуточная аттестация (сам.раб.)
<b>ОСНОВНАЯ ТРАЕКТОРИЯ</b>																	
<b>очная форма обучения</b>																	
Семестр 7																	
ИТОГО																	

Виды, формы и сроки текущего контроля успеваемости и промежуточной аттестации						
Код модуля в составе дисциплины, практики и т.п.	Формы текущего контроля успеваемости		Виды промежуточной аттестации		Виды итоговой аттестации (только для программ итоговой аттестации и дополнительных образовательных программ)	
	Формы	Сроки	Виды	Сроки	Виды	Сроки
<b>ОСНОВНАЯ ТРАЕКТОРИЯ</b>						
<b>очная форма обучения</b>						
Семестр 7			зачет, устно, традиционная форма	по графику промежуточной аттестации, по		

				графику промежуточной аттестации		
--	--	--	--	--	--	--

## 2.2. Структура и содержание учебных занятий

Период обучения (модуль): Семестр 7

№ п/п	Наименование темы (раздела, части)	Вид учебных занятий	Количество часов
1	"Линейная и логистическая регрессия"	Лекции	
		семинары	15
		в присутствии преподавателя	
		по методическим материалам	
2	"Модели машинного обучения"	Лекции	
		семинары	15
		в присутствии преподавателя	
		по методическим материалам	
3	Зачет	промежуточная аттестация (ауд)	2
		промежуточная аттестация (с.р.)	32

### Раздел 1: Линейная и логистическая регрессия

1. Введение. История искусственного интеллекта. Вспоминаем теорию вероятностей. Теорема Байеса и машинное обучение. Что умеет делать машинное обучение.
2. Правило Лапласа. Априорные распределения. Сопряжённые априорные распределения.
3. Наименьшие квадраты и ближайшие соседи. Линейная регрессия. Логистическая регрессия.
4. Статистическая теория принятия решений. Разложение bias-variance-noise. Оверфиттинг. Регуляризация: гребневая регрессия. Линейная регрессия по-байесовски.
5. Линейная регрессия: разные формы регуляризаторов. Лассо-регрессия. Эквивалентные ядра. Проклятие размерности.
6. Задачи классификации. Линейный дискриминант Фишера. Наивный байесовский классификатор: мультиномиальный и многомерный.
7. Логистическая регрессия: как обучать. Мультиклассовая логистическая регрессия. Аппроксимация по Лапласу. Пробит. Логистическая регрессия по-байесовски.

### Раздел 2: Модели машинного обучения

1. Метод опорных векторов (SVM). Трюк с ядрами.
2. Варианты SVM. SVM по-байесовски: relevance vector machines.
3. Кластеризация: иерархическая, методами теории графов. Алгоритм EM для кластеризации. Алгоритм EM в общем виде.

4. Скрытые марковские модели.
5. Комбинация моделей: усреднение, бутстрап, бэггинг. Бустинг: AdaBoost.
6. Обучение ранжированию: постановка задачи, RankBoost. LambdaRank.
7. Рекомендательные системы: метод ближайших соседей, сингулярное разложение матриц.

### **Раздел 3. Обеспечение учебных занятий**

#### **3.1. Методическое обеспечение**

##### **3.1.1 Методические указания по освоению дисциплины**

Посещение лекций и семинаров

##### **3.1.2 Методическое обеспечение самостоятельной работы**

Основная и дополнительная литература

##### **3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания**

Зачет проводится в устной форме. Для получения зачета студент должен сделать доклад на семинаре. Преподаватель задает вопросы по теме доклада, а также дополнительные вопросы по другим темам, изложенным в курсе.

В случае получения оценки «не зачтено» студент должен написать и представить преподавателю реферат по теме доклада, ответить на устные вопросы.

Оценка «зачтено» ставится за связное и математически корректное изложение материала доклада (при передаче – в письменной форме), знание основных вопросов по каждой теме.

Оценка «не зачтено» выставляется, если не выполняется условие для получения оценки «зачтено».

##### **3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)**

Период обучения (модуль): **Семестр 7**

**Список вопросов к зачету:**

1. Правило Лапласа. Априорные распределения. Сопряжённые априорные распределения.
2. Наименьшие квадраты и ближайшие соседи. Линейная регрессия. Логистическая регрессия. Разные формы регуляризаторов: гребневая регрессия, лассо-регрессия.
3. Статистическая теория принятия решений. Разложение bias-variance-noise. Оверфиттинг.
4. Линейная регрессия по-байесовски.

5. Задачи классификации. Линейный дискриминант Фишера. Наивный байесовский классификатор: мультиномиальный и многомерный.
6. Логистическая регрессия: как обучать. Мультиклассовая логистическая регрессия. Аппроксимация по Лапласу. Пробит. Логистическая регрессия по-байесовски.
7. Метод опорных векторов (SVM). Трюк с ядрами.
8. Варианты SVM. SVM по-байесовски: relevance vector machines.
9. Кластеризация: иерархическая, методами теории графов. Алгоритм EM для кластеризации.
10. Алгоритм EM в общем виде.
11. Скрытые марковские модели: три задачи.
12. Скрытые марковские модели: обучение, алгоритм Баума-Уэлша.
13. Комбинация моделей: усреднение, бутстрап, бэггинг. Бустинг: AdaBoost.
14. Обучение ранжированию: постановка задачи, RankBoost. LambdaRank.
15. Рекомендательные системы: метод ближайших соседей, сингулярное разложение матриц.

#### **Темы семинаров:**

1. Решение задач о теореме Байеса.
2. Решение задач о сопряжённых априорных распределениях.
3. Решение задач о линейной регрессии.
4. Практическое занятие: линейная регрессия, разные формы регуляризаторов.
5. Решение задач о классификаторах.
6. Практическое занятие: реализация наивного байесовского классификатора, основы обработки текстов.
7. Решение задач о лапласовских аппроксимациях.
8. Практическое занятие: стандартные классификаторы. Логистическая регрессия.
9. Практическое занятие: обучение различных SVM-моделей.
10. Практическое занятие: кластеризация. Сравнение разных алгоритмов.
11. Решение задач о расстоянии Кульбака-Лейблера и алгоритме EM.
12. Практическое занятие: скрытые марковские модели.
13. Практическое занятие: методы комбинации моделей. Бустинг.
14. Практическое занятие: обучение ранжированию.
15. Практическое занятие: рекомендательные системы.

### **3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса**

## **3.2. Кадровое обеспечение**

### **3.2.1 Образование и (или) квалификация штатных преподавателей и иных лиц, допущенных к проведению учебных занятий**

К чтению лекций должны привлекаться преподаватели, имеющие ученую степень доктора или кандидата наук (в том числе степень PhD, прошедшую установленную процедуру признания и установления эквивалентности) и/или ученое звание профессора или доцента.

### **3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом**

не требуется

## **3.3. Материально-техническое обеспечение**

### **3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий**

Стандартно оборудованные лекционные аудитории, должны вмещать поток в соответствии со списком студентов

### **3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования**

Для проведения занятий требуется проектор для показа слайдов, а также маркерная или меловая доска.

### **3.3.3 Характеристики специализированного оборудования**

не требуется

### **3.3.4 Характеристики специализированного программного обеспечения**

не требуется

### **3.3.5 Перечень и объёмы требуемых расходных материалов**

Маркеры или мел для доски, губка.

## **3.4. Информационное обеспечение**

### **3.4.1 Список обязательной литературы**

1. С.И. Николенко, А.Л. Тулупьев. Самообучающиеся системы. М., МЦНМО, 2009.

### **3.4.2 Список дополнительной литературы**

### **3.4.3 Перечень иных информационных источников**

1. Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, Information Science and Statistics series, 2006.

2. Kevin Murphy. Machine Learning: A Probabilistic Perspective, MIT Press, 2012.



3. David J. C. MacKay. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003.

#### **Раздел 4. Разработчики программы**

Николенко Сергей Игоревич, кандидат физ.-мат. наук, научный сотрудник ПОМИ РАН,  
snikolenko@gmail.com