Государственный комитет по высшему образованию
Московский физико-технический институт

УТВЕРЖДАЮ
Проректор по учебной работе
_____ Т. В. Кондранин
"___" _____ 200_ г.

Факультет управления и прикладной математики
Кафедра интеллектуальных систем

ПРОГРАММА

по курсу: МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБУЧЕНИЯ ПО ПРЕЦЕДЕНТАМ

по направлению 511656
курс 3, 4
семестр 6, 7
лекции 66 часов
практические (семинарские) занятия 16 часов
лабораторные занятия 0 часов

Диф. зачет 6 семестр
Экзамен 7 семестр

Программу составил: к.ф.-м.н. К. В. Воронцов

Программа обсуждена на заседании кафедры 14 сентября 2004 г.

Программа обсуждена и одобрена на методической комиссии факультета
"___" _____ 200_ г.

Председатель методической комиссии ФУПМ

чл.-корр.РАН Ю.А. Флеров

1 Часть I (6 семестр)

1.1 Задачи распознавания, классификации, прогнозирования

Постановка задач обучения по прецедентам. Понятие модели алгоритмов. Объекты и признаки. Задачи со стандартным описанием (матрица объекты–признаки). Типы шкал: бинарные, номинальные, порядковые, количественные. Функционалы качества. Понятие обобщающей способности. Примеры прикладных задач.

1.2 Байесовская теория решений

Функционал среднего риска. Ошибки I и II рода. Теорема об оптимальности байесовского решающего правила. Нормальный дискриминантный анализ. Геометрическая интерпретация. Линейные и квадратичные разделяющие поверхности: разбор случаев.
Литература: [1].

1.3 Линейный дискриминант Фишера

Принцип максимума правдоподобия, выборочные оценки. Подстановочный алгоритм, его недостатки и способы их преодоления. Проблема переобучения. Регуляризация ковариационной матрицы. Метод редукции размерности. Робастное оценивание.
Семинар: Решение задач на линейные дискриминанты. Экскурс в матричное дифференцирование.
Литература: [1], [17].

1.4 Разделение смеси распределений, EM-алгоритм

Модель смеси распределений. Теорема о смеси многомерных нормальных распределений. Критерий останова. Выбор начального приближения. Выбор числа компонентов смеси. *Иерархический EM-алгоритм.* Сети радиальных базисных функций (RBF) и их настройка с помощью EM-алгоритма.
Литература: [15].

1.5 Метрические алгоритмы классификации

Метод k ближайших соседей (kNN) и его обобщения. Подбор числа k по критерию скользящего контроля. Отбор эталонных объектов. *Алгоритмы быстрого поиска ближайших объектов. Проблема синтеза метрик. Выбор весов признаков (взвешенный kNN)*.
Непараметрическая оценка плотности распределения по Парзену-Розенблатту. Выбор функции ядра. Выбор ширины окна, переменная ширина окна. Метод потенциальных функций. Простейший алгоритм обучения: теорема сходимости, достоинства и недостатки, сопоставление с RBF.
Литература: [8], [9], [15].

1.6 Кластеризация (обучение без учителя)

Примеры прикладных задач. Графовые алгоритмы. Функционалы качества кластеризации. Статистические алгоритмы: EM и k-means. Агломеративные (иерархические) алгоритмы. Формула Ланса-Вильямса. Теорема о монотонности. Алгоритм построения дендрограммы. *Потоковые (субквадратичные) алгоритмы кластеризации.*
Задача многомерного шкалирования, визуализация кластерной структуры на картах сходства.
Литература: [1].

1.7 Непараметрическая регрессия

Задача одномерного сглаживания. Оценка Надарая-Ватсона. Обобщение на случай произвольного метрического пространства. Выбор функции ядра. Выбор ширины окна сглаживания. Сглаживание с переменной шириной окна.

Интерполяция и аппроксимация с помощью сплайнов. Алгоритм построения кубических сплайнов, метод прогонки.
Литература: [16].

1.8 Многомерная линейная регрессия

Принцип наименьших квадратов. Линейная, полиномиальная и криволинейная регрессия. Нормальная система, псевдообращение, ортогональные разложения. Модифицированная ортогонализация Грама-Шмидта. Недостатки линейного МНК. Робастная регрессия. Проблема мультиколлинеарности. Проблема переобучения. Гребневая регрессия. Отбор признаков путём ортогонализации с выбором главного элемента, критерии выбора и останова. Лассо Тибширани. Шаговая регрессия (метод добавлений-удалений). *Линейная неубывающая регрессия, симплекс-метод.*
Литература: [2], [13], [19].

1.9 Нелинейная регрессия

Нелинейная параметрическая регрессия, методы Ньютона-Рафсона и Ньютона-Гаусса. Неквадратичные функции потерь. Одномерные нелинейные преобразования признаков: метод последовательных сглаживаний Хасти-Тибширани и метод гистограмм.
Литература: [19].

1.10 Логистическая регрессия

Линейный пороговый классификатор. «Наивное» сведение задачи классификации к задаче регрессии, его недостатки. Гладкие аппроксимации пороговой функции потерь, в том числе на основе принципа максимума правдоподобия. Метод наименьших квадратов с итеративным пересчетом весов. *Подгонка порога решающего правила по критерию числа ошибок I и II рода, кривая ошибок (lift curve), отказы от классификации.*
Примеры прикладных задач (кредитный скоринг, скоринговые карты).
Литература: [19].

1.11 Отбор информативных признаков

Внутренние и внешние критерии. Полный перебор. Поиск в глубину — метод ветвей и границ. Усечённый поиск в ширину — многорядный итерационный алгоритм МГУА. Метод добавлений-удалений. Случайный поиск с адаптацией. Генетические алгоритмы. Эвристики скрещивания, мутации и отбора алгоритмов.
Литература: [2], [9], [10].

1.12 Синтез информативных признаков

Линейные преобразования признакового пространства. Анализ главных компонент. Преобразование Карунена-Лоэва. *Анализ независимых компонент.*
Семинар: Вывод формул преобразования признакового пространства.
Литература: [2].

1.13 Прогнозирование временных рядов

Одномерные и многомерные временные ряды. Аддитивная модель временного ряда: тренд, сезонность, цикличность. Модель Бокса-Дженкинса. *Модель авторегрессии и интегрированного скользящего среднего.* Регрессионные модели.
Семинар: Примеры экономических приложений. Прогнозирование цен электроэнергии. Прогнозирование потребительского спроса.
Литература: [3], [14].

2 Часть II (7 семестр)

2.1 Персептрон и искусственные нейронные сети

Естественный нейрон и его математическая модель. Персептрон Розенблатта. Метод стохастического градиента. *Теорема сходимости (Новикова).* Проблема «исключающего или». Проблема полноты. Теоремы Колмогорова и Стоуна. Геометрическое доказательство полноты. Полнота двухслойных сетей в пространстве булевских функций.
Литература: [7].

2.2 Многослойные нейронные сети

Алгоритм обратного распространения ошибок. Недостатки алгоритма, способы их устранения. Проблема переобучения. Подбор структуры сети. Регуляризация (редукция) весов. Сети с радиальными базисными функциями.
Литература: [7].

2.3 Обучающееся векторное квантование (сети Кохонена)

Структура сети Кохонена. Конкурентное обучение. Самоорганизующиеся карты Кохонена. Применения: визуальный анализ данных, аппроксимация функций, *решение комбинаторных задач (на примере задачи коммивояжера)*. Примеры прикладных задач.
Литература: [7], [15].

2.4 Машины опорных векторов (SVM)

Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin). Случай линейной разделимости. Задача квадратичного программирования. Опорные векторы. Функции ядра (kernel functions), спрямляющее пространство, теорема Мерсера. Сопоставление SVM и нейронной сети. Случай отсутствия линейной разделимости. Алгоритм настройки SVM методом активных ограничений. SVM-регрессия.
Литература: [15], [18].

2.5 Алгоритмические композиции

Понятия базового алгоритма и корректирующей операции. Процесс последовательного построения базовых алгоритмов.
Комитеты большинства. Взвешенное голосование алгоритмов. Сопоставление с нейронной сетью. Понятия максимальной совместной подсистемы и минимального комитета. Последовательное построение комитета большинства. Верхняя оценка числа членов комитета. Комитеты старшинства (решающий список). Последовательное построение комитета старшинства.
Взвешенное голосование алгоритмов. Бустинг, алгоритм AdaBoost. Теорема о сходимости. *Теорема об увеличении зазора. Связь бустинга с градиентными методами и логистической регрессией. Обобщающая способность AdaBoost.* Разновидности алгоритмов бустинга. *Алгоритмы ComBoost и SvmBoost.* Бэггинг.
Нелинейная монотонная коррекция.
Понятие области компетентности алгоритма. Иерархические смеси алгоритмов. Процесс совместного обучения базовых алгоритмов. EM-алгоритм. Генетический алгоритм.
Литература: [22].

2.6 Логические алгоритмы классификации

Понятие логической закономерности. Эвристическое, энтропийное и комбинаторное определения информативности, их асимптотическая эквивалентность. Способы бинаризации признаков. Задача оптимального разбиения интервала значений признака на зоны.
Литература: [5], [20].

2.7 Решающие списки и деревья

Решающие списки. Жадный алгоритм синтеза списка. Разновидности решающих правил в списках: шары, гиперплоскости, гиперпараллелепипеды (конъюнкции).
Решающие деревья. Алгоритм синтеза дерева ID3. Недостатки алгоритма и способы их устранения. Проблема переобучения. Редукция решающих деревьев.
Алгоритм CART. Алгоритм C4.5. Решающий лес — список редуцированных деревьев. Бустинг над решающими деревьями.
Литература: [21], [4], [20].

2.8 Взвешенное голосование логических закономерностей

Классификация по принципу голосования. Алгоритмы синтеза конъюнктивных закономерностей КОРА и ТЭМП. Применение ТЭМП для синтеза решающего списка.
Проблема диверсификации конъюнкций. Алгоритм бустинга. Теорема сходимости.
Взвешенные решающие деревья (*alternating decision tree*).
Примеры прикладных задач: кредитный скоринг, прогнозирование ухода клиентов.
Литература: [8], [9], [22].

2.9 Алгоритмы вычисления оценок

Структура АВО. Тупиковые тесты и тупиковые представительные наборы. Проблема оптимизации АВО. Применение бустинга для оптимизации АВО.
Литература: [11].

2.10 Задачи с большим числом классов

Применение логических алгоритмов и композиций для сведения к задаче с 2 классами: каждый против всех, каждый против каждого, турнир на выбывание, дерево классов. Алгоритм ECOC. Примеры задач: распознавание символов, речи.
Литература: [15].

2.11 Статистическая теория обучения

Проблема переобучения. Принцип ограничения сложности модели. Функционалы качества обучения. Функция роста и ёмкость. Теорема Вапника-Червоненкиса. Метод структурной минимизации риска. Ёмкость семейства линейных решающих правил. Принцип минимума длины описания. Достаточная длина обучающей выборки. Причины завышенности оценок Вапника-Червоненкиса. Эффект локализации семейства алгоритмов. *Оценки, зависящие от данных. Принцип самоограничения сложности.*
Литература: [5].

2.12 Оценки обобщающей способности некоторых методов обучения

Обобщающая способность линейных комбинаций (машины опорных векторов, взвешенное голосование, нейронные сети, бустинг). Методы явной максимизации зазора. *Профиль разделимости выборки.*
Обобщающая способность метрических алгоритмов классификации. Профиль компактности выборки. Алгоритм разделения объектов на шумовые, периферийные и опорные.
Понятие стабильности метода обучения. Методы, локальные по Деврою.
Обобщающая способность процедуры выбора модели по критерию скользящего контроля. Профиль множества моделей.
Обобщающая способность методов, основанных на поиске логических закономерностей. Связь информативности и обобщающей способности. Профиль информативности.
Литература: [6].

Литература

[1] Айвазян С. А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Классификация и снижение размерности. — М. Финансы и статистика. 1989.

[2] Айвазян С. А., Енюков И.С., Мешалкин Л.Д. Исследование зависимостей. — М. Финансы и статистика. 1985.

[3] Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. — М.: Юнити, 1998.

[4] Вагин В. Н., Головина Е. Ю., Загорянская А. А, Фомина М. В. Достоверный и правдоподобный вывод в интеллектуальных системах. — М.: Физматлит. 2004.

[5] Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука. 1979.

[6] Воронцов К. В. Комбинаторный подход к оценке качества обучаемых алгоритмов // Математические вопросы кибернетики.
http://**www.ccas.ru/frc/papers/voron04mpc.pdf**

[7] Головко. В. А. Нейронные сети: обучение, организация и применение. — М.: ИПРЖР. 2001.

[8] Загоруйко Н. Г. Прикладные методы анализа данных и знаний. — Новосибирск: ИМ СО РАН, 1999.

[9] Загоруйко Н. Г., Ёлкина В. Н., Лбов Г. С. Алгоритмы обнаружения эмпирических закономерностей. — Новосибирск: Наука, 1985.

[10] Ивахненко А. Г., Юрачковский Ю. П. Моделирование сложных систем по экспериментальным данным. — М.: Радио и связь, 1987.

[11] Журавлев Ю. И. Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. — 1978. — Т. 33. — С. 5–68.

[12] Казанцев В. С. Задачи классификации и их программное обеспечение. — М. Наука. 1990.

[13] Лоусон Ч, Хенсон Р. Численное решение задач метода наименьших квадратов. — М. Наука. 1986.

[14] Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика: начальный курс. М.: Дело. 2004.

[15] Мерков А. Б. Основные методы, применяемые для распознавания рукописного текста. Лаборатория распознавания образов МЦНМО. 2004.
http://**www.recognition.mccme.ru/pub/RecognitionLab.html/methods.html.**

[16] Хардле В. Прикладная непараметрическая регрессия. — М.: Мир. 1993.

[17] Шурыгин А. М. Прикладная стохастика: робастность, оценивание, прогноз. — М. Финансы и статистика. 2000.

[18] Burges C. J. C. A tutorial on support vector machines for pattern recognition // Data Mining and Knowledge Discovery. — 1998. — Vol. 2, no. 2. — Pp. 121–167. http://citeseer.ist.psu.edu/burges98tutorial.html.

[19] Hastie T., Tibshirani R. Generalized additive models. Chapman and Hall. London. 1990. http://citeseer.ist.psu.edu/hastie95generalized.html.

[20] Martin J. K. An exact probability metric for decision tree splitting and stopping // Machine Learning. — 1997. — Vol. 28, no. 2-3. — Pp. 257–291. http://citeseer.ist.psu.edu/martin97exact.html.

[21] Marchand M., Shawe-Taylor J. Learning with the set covering machine // Proc. 18th International Conf. on Machine Learning. — Morgan Kaufmann, San Francisco, CA, 2001. — Pp. 345–352. http://citeseer.ist.psu.edu/452556.html.

[22] Schapire R. The boosting approach to machine learning: An overview // MSRI Workshop on Nonlinear Estimation and Classification, Berkeley, CA. — 2001. http://citeseer.ist.psu.edu/schapire02boosting.html.