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
специализация «Интеллектуальный анализ данных»

ПРОГРАММА

по курсу: Машинное обучение

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

Аннотация

Машинное обучение возникло на стыке прикладной статистики, оптимизации, дискретного анализа, и за последние 30 лет оформилось в самостоятельную математическую дисциплину. Методы машинного обучения составляют основу ещё более молодой дисциплины — интеллектуального анализа данных (data mining).

В курсе рассматриваются основные задачи обучения по прецедентам: классификация, кластеризация, регрессия, понижение размерности. Изучаются методы их решения, как классические, так и новые, созданные за последние 10–15 лет. Упор делается на глубокое понимание математических основ, взаимосвязей, достоинств и ограничений рассматриваемых методов. Отдельные теоремы приводятся с доказательствами. Все методы излагаются по единой схеме: исходные идеи и эвристики; их формализация и математическая теория; описание алгоритма в виде слабо формализованного псевдокода; анализ достоинств, недостатков и границ применимости; пути устранения недостатков; сравнение с другими методами; примеры прикладных задач.

Данный курс существенно расширяет и углубляет набор тем, рекомендованный международным стандартом ACM/IEEE Computing Curricula 2001 по дисциплине «Машинное обучение и нейронные сети» (machine learning and neural networks) в разделе «Интеллектуальные системы» (intelligent systems).

На материал данного курса существенно опираются последующие курсы, читаемые студентам на специализации «Интеллектуальный анализ данных».

От студентов требуются знания курсов линейной алгебры, математического анализа, теории вероятностей. Знание математической статистики, методов оптимизации и какого-либо языка программирования желательно, но не обязательно.

1 Часть I (6 семестр) [1]

1.1 Задачи обучения по прецедентам

Постановка задач обучения по прецедентам, типы задач. Понятия модели алгоритмов и метода обучения. Функционалы качества и принцип минимизации эмпирического риска. Понятие обобщающей способности. Скользящий контроль. Вероятностная постановка задачи и принцип максимума правдоподобия. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные. Примеры прикладных задач распознавания, классификации, кластеризации, прогнозирования.

1.2 Байесовская теория решений

Функционал среднего риска. Ошибки I и II рода. Теорема об оптимальности байесовского решающего правила. Задача восстановления плотности распределения. «Наивный» байесовский классификатор. Непараметрическое оценивание плотности распределения по Парзену-Розенблатту. Выбор функции ядра. Выбор ширины окна, переменная ширина окна. Робастное оценивание плотности.
Литература: [3].

1.3 Нормальный дискриминантный анализ

Параметрическое оценивание плотности вероятности. Нормальный дискриминантный анализ. *Матричное дифференцирование и оценки параметров нормального распределения.* Геометрическая интерпретация. Линейные и квадратичные разделяющие поверхности. Подстановочный алгоритм, его недостатки и способы их преодоления.

[1] Курсивом выделены темы, которые не входят в программу, но рассказывались в прошлые годы, либо, возможно, будут включены в будущем.

Литература: [3], [19].

1.4 Линейный дискриминант Фишера

Линейный дискриминант Фишера. Проблемы мультиколлинеарности и переобучения. Регуляризация ковариационной матрицы. *Метод редукции размерности Шурыгина.* Робастное оценивание.
Литература: [3], [19].

1.5 Разделение смеси распределений, EM-алгоритм

Модель смеси распределений. EM-алгоритм. Теорема о смеси многомерных нормальных распределений. Критерий останова. Выбор начального приближения. Выбор числа компонентов смеси. Стохастический EM-алгоритм. Сети радиальных базисных функций (RBF) и их настройка с помощью EM-алгоритма. *Иерархический EM-алгоритм.*
Литература: [17].

1.6 Метрические алгоритмы классификации

Метод k ближайших соседей (kNN) и его обобщения. Подбор числа k по критерию скользящего контроля. Обобщённый метрический классификатор. Метод потенциальных функций, градиентный алгоритм. Настройка весов объектов. Отбор эталонных объектов. *Алгоритмы быстрого поиска ближайших объектов. Проблема синтеза метрик. Выбор весов признаков (взвешенный kNN)*.
Литература: [10], [11], [3].

1.7 Кластеризация (обучение без учителя)

Примеры прикладных задач. Графовые алгоритмы: связные компоненты, кратчайший незамкнутый путь, Форель. Функционалы качества кластеризации. Статистические алгоритмы: EM и k-means. Агломеративные (иерархические) алгоритмы. Формула Ланса-Вильямса. Алгоритм построения дендрограммы. Свойства сжатия/растяжения, монотонности и редуктивности. Определение числа кластеров. *Потоковые (субквадратичные) алгоритмы кластеризации.*
Литература: [3], [10], [11].

1.8 Многомерное шкалирование

Многомерное шкалирование. Размещение одной точки методом Ньютона-Рафсона. Субквадратичный алгоритм. Визуализация: карты сходства и диаграммы Шепарда. *Совмещение многомерного шкалирования и иерархической кластеризации.* Примеры прикладных задач.
Литература: [3].

1.9 Непараметрическая регрессия

Локально взвешенный метод наименьших квадратов и оценка Надарая-Ватсона. Выбор функции ядра. Выбор ширины окна сглаживания. Сглаживание с переменной шириной окна. Проблема «выбросов» и робастная непараметрическая регрессия. Проблема «проклятия размерности» и проблема выбора метрики.
Интерполяция и аппроксимация с помощью сплайнов. Алгоритм построения кубических сплайнов, метод прогонки.
Литература: [4], [18].

1.10 Многомерная линейная регрессия

Принцип наименьших квадратов. Сингулярное разложение. Проблемы мультиколлинеарности и переобучения. Гребневая регрессия. Лассо Тибширани. *Линейная монотонная регрессия*

(симплекс-метод). Линейные преобразования признакового пространства. Метод главных компонент и декоррелирующее преобразование Карунена-Лоэва. Робастная регрессия.
Литература: [2], [4], [15].

1.11 Шаговая регрессия

Алгоритм модифицированной ортогонализации Грама-Шмидта, достоинства и недостатки. Отбор признаков в процессе ортогонализации, критерии выбора и останова. *Метод наименьших углов, его связь с лассо и шаговой регрессией.*
Литература: [2], [4], [15].

1.12 Нелинейная регрессия

Нелинейная параметрическая регрессия, применение методов Ньютона-Рафсона и Ньютона-Гаусса. Одномерные нелинейные преобразования признаков: метод обратной настройки (backfitting) Хасти-Тибширани, *метод гистограмм*. Обобщённые линейные модели. Неквадратичные функции потерь, примеры прикладных задач.
Литература: [2].

1.13 Логистическая регрессия

Линейный пороговый классификатор. «Наивное» сведение задачи классификации к задаче регрессии, его недостатки. Гладкие аппроксимации пороговой функции потерь. Обоснование логистической регрессии: теорема об экспонентных плотностях. Метод наименьших квадратов с итеративным пересчетом весов. Настройка порога решающего правила по критерию числа ошибок I и II рода, кривая ошибок (lift curve), *отказы от классификации*. Пример прикладной задачи: кредитный скоринг и скоринговые карты.
Литература: [2], [3].

1.14 Критерии качества регрессионной модели

Внутренние и внешние критерии. Скользящий контроль, критерии непротиворечивости и регуляризации. Критерии, основанные на оценках обобщающей способности: Вапника-Червоненкиса, Акаике (AIC), байесовский (BIC). Статистические критерии: коэффициент детерминации, критерий Фишера, анализ остатков.

1.15 Теория обобщающей способности

Функционалы скользящего контроля. Теорема Вапника-Червоненкиса. Функция роста и ёмкость. Емкость некоторых семейств алгоритмов. Метод структурной минимизации риска. Принцип минимума длины описания. Достаточная длина обучающей выборки. Причины завышенности оценок Вапника-Червоненкиса. Эффект локализации семейства алгоритмов. Оценки, зависящие от данных. Принцип самоограничения сложности. Декомпозиция ошибки на шум, смещение и вариацию. Понятие стабильности обучения. Методы эмпирического оценивания обобщающей способности.
Литература: [7].

1.16 Методы отбора информативных признаков

Методы отбора признаков: полный перебор, метод добавлений-удалений (шаговая регрессия), поиск в глубину (метод ветвей и границ), усечённый поиск в ширину, многорядный итерационный алгоритм МГУА, генетический алгоритм, случайный поиск с адаптацией.
Литература: [3], [11], [12].

1.17 Прогнозирование временных рядов

Одномерные и многомерные временные ряды. Аддитивная модель временного ряда: тренд, сезонность, цикличность. Модель Бокса-Дженкинса. Модель ARIMA — авторегрессии и интегрированного скользящего среднего. Фурье-модели. Регрессионные модели с отбором

информативных признаков. Адаптивные модели. Примеры экономических приложений. Прогнозирование грузоперевозок, потребительского спроса. Модель прерывистого спроса. Прогнозирование при несимметричном неквадратичном функционале качества.
Литература: *[5], [16].*

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). Связь однослойного персептрона с логистической регрессией и обоснование сигмоидной функции потерь. Проблема «исключающего или». Проблема полноты. Полнота двухслойных сетей в пространстве булевских функций. Теоремы Колмогорова, Стоуна, Горбаня (без доказательства).
Литература: [9].

2.2 Многослойные нейронные сети

Алгоритм обратного распространения ошибок. Недостатки алгоритма, способы их устранения. Проблема переобучения. Проблема «паралича» сети. Редукция весов. Подбор структуры сети. Метод оптимальной редукции сети (optimal brain damage).
Литература: [9].

2.3 Обучающееся векторное квантование (сети Кохонена)

Структура сети Кохонена. Конкурентное обучение, стратегии WTA и WTM. Самоорганизующиеся карты Кохонена. Применение для визуального анализа данных. Сети встречного распространения, их применение для кусочно-постоянной и гладкой аппроксимации функций. *Решение комбинаторных задач (на примере задачи коммивояжера)*.
Литература: [9], [17].

2.4 Машины опорных векторов (SVM)

Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin). Случай линейной разделимости. Задача квадратичного программирования. Опорные векторы. Случай отсутствия линейной разделимости. Функции ядра (kernel functions), спрямляющее пространство, теорема Мерсера. Способы построения ядер. Примеры ядер. Сопоставление SVM и нейронной RBF-сети. Обучение SVM методом активных ограничений. SVM-регрессия. *Обобщающая способность линейных комбинаций. Принцип максимизации зазора. Профиль разделимости выборки.*
Литература: [17], [20].

2.5 Линейные алгоритмические композиции

Понятия базового алгоритма и корректирующей операции. Процесс последовательного обучения базовых алгоритмов. Простое голосование (комитет большинства). Решающий список (комитет старшинства). Взвешенное голосование. Бустинг: алгоритм AdaBoost, теорема сходимости. Стохастические методы: бэггинг и метод случайных подпространств.
Литература: [13], [23].

2.6 Нелинейные алгоритмические композиции

Смеси экспертов, понятие области компетентности алгоритма. Выпуклые функции потерь. Методы построения смесей: последовательный и иерархический. *Построение смесей экспертов с помощью EM-алгоритма. Нелинейная монотонная коррекция.*

2.7 Метод комитетов

Комитеты большинства, простое и взвешенное голосование. Сопоставление с нейронной сетью. Понятия максимальной совместной подсистемы и минимального комитета. Алгоритм построения комитета большинства. Верхняя оценка числа членов комитета.
Литература: [14].

2.8 Адаптивные композиции алгоритмов прогнозирования

Выбор лучшей модели по скользящему контролю. Адаптивная линейная композиция алгоритмов прогнозирования. Регуляризация. Ограничение неотрицательности весов. Пример прикладной задачи: прогнозирование потребительского спроса.

2.9 Логические алгоритмы классификации

Понятие логической закономерности. Энтропийное и комбинаторное определения информативности, их асимптотическая эквивалентность. Разновидности закономерностей: шары, гиперплоскости, гиперпараллелепипеды (конъюнкции). Бинаризация признаков, алгоритм выделения информативных зон. «Градиентный» алгоритм синтеза конъюнкций, частные случаи: жадный алгоритм, стохастический локальный поиск, стабилизация, редукция. *Обобщающая способность методов, основанных на поиске логических закономерностей. Связь информативности и обобщающей способности. Профиль информативности.*
Литература: [7], [21].

2.10 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: предредукция и постредукция. Преобразование решающего дерева в решающий список. Решающий лес и бустинг над решающими деревьями. *Алгоритм CART. Алгоритм C4.5.*
Литература: [22], [6], [21].

2.11 Взвешенное голосование логических закономерностей

Классификация по принципу голосования. Проблема различности (диверсификации) закономерностей. Алгоритмы синтеза конъюнктивных закономерностей КОРА и ТЭМП. Применение ТЭМП для синтеза решающего списка. Алгоритм бустинга. Теорема сходимости. *Взвешенные решающие деревья (alternating decision tree).* Примеры прикладных задач: кредитный скоринг, прогнозирование ухода клиентов.
Литература: [10], [11], [23].

2.12 Алгоритмы вычисления оценок

Принцип частичной прецедентности. Структура АВО. Тупиковые тесты и тупиковые представительные наборы. Проблема оптимизации АВО. АВО как композиция метрических закономерностей. Применение бустинга для оптимизации АВО.
Литература: [13].

2.13 Поиск ассоциативных правил

Пример прикладной задачи: анализ рыночных корзин. Понятие ассоциативного правила и его связь с понятием логической закономерности. Алгоритм APriori, его недостатки и пути усовершенствования.

2.14 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: [17].

Основная литература

[1] Воронцов К. В. Математические методы обучения по прецедентам. 2007. http://www.ccas.ru/teaching.html.

[2] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer. 2001.

Дополнительная литература

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

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

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

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

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

[8] Воронцов К. В. Комбинаторный подход к оценке качества обучаемых алгоритмов // Математические вопросы кибернетики. http://www.ccas.ru/voron.

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

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

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

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

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

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

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

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

[17] 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.

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

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

[20] 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.

[21] 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.

[22] 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.

[23] 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.