

Аннотация дисциплины Б.1.2.12 Дисциплина. Перспективное программное обеспечение

Дисциплина "Перспективное программное обеспечение" изучается обучающимися по основной профессиональной образовательной программе "Разработка программных систем" направления подготовки "09.03.04 Программная инженерия".

Дисциплина изучается в 8 семестре. Общая трудоемкость дисциплины составляет 144/4 часов/з.ед. Самостоятельная работа заключается в выполнении работ, указанных в разделе 4.

В ходе изучения дисциплины осуществляется текущий контроль в форме технологии рейтингового контроля в соответствии с технологической карты дисциплины, размещенной на электронном курсе, а также промежуточный контроль в форме зачет.

Целью изучения дисциплины является формирование следующих компетенций:

1. ПК-7 Готовность к использованию методов и инструментальных средств исследования объектов профессиональной деятельности
2. ПК-9 Владение навыками моделирования, анализа и использования формальных методов конструирования программного обеспечения

В ходе изучения дисциплины последовательно рассматриваются темы:

1. Лекция №1. Основные понятия и примеры прикладных задач.
Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые, количественные.
Типы задач: классификация, регрессия, прогнозирование, ранжирование.
Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль.
Линейные модели регрессии и классификации. Метод наименьших квадратов. Полиномиальная регрессия.
Примеры прикладных задач.
2. Лекция №2. Метрические методы классификации и регрессии.
Гипотезы компактности и непрерывности.
Обобщённый метрический классификатор.
Метод ближайших соседей kNN и его обобщения. Подбор числа k по критерию скользящего контроля.
Метод окна Парзена с постоянной и переменной шириной окна.
Метод потенциальных функций и его связь с линейной моделью классификации.
Непараметрическая регрессия. Локально взвешенный метод наименьших квадратов. Ядерное сглаживание.
Оценка Надарая-Ватсона с постоянной и переменной шириной окна. Выбор функции ядра.
Задача отсева выбросов. Робастная непараметрическая регрессия. Алгоритм LOWESS.
Задача отбора эталонов. Понятие отступа. Алгоритм СТОЛП.
Задача отбора признаков. Жадный алгоритм построения метрики.
3. Лекция №3. Логические методы классификации.
Понятие логической закономерности.
Параметрические семейства закономерностей: конъюнкции пороговых правил, синдромные правила, шары, гиперплоскости.
Переборные алгоритмы синтеза конъюнкций: стохастический локальный поиск, стабилизация, редукция.
Двухкритериальный отбор информативных закономерностей, парето-оптимальный фронт в (p,n) -пространстве.

Решающее дерево. Жадная нисходящая стратегия «разделяй и властвуй». Алгоритм ID3. Недостатки жадной стратегии и способы их устранения. Проблема переобучения.

Вывод критериев ветвления. Мера нечистоты (impurity) распределения. Энтропийный критерий, критерий Джини.

Редукция решающих деревьев: предредукция и постредукция. Алгоритм C4.5.

Деревья регрессии. Алгоритм CART.

Небрежные решающие деревья (oblivious decision tree).

Решающий лес. Случайный лес (Random Forest).

4. Лекция №4. Градиентные методы обучения.

Линейный классификатор, модель МакКаллока-Питтса, непрерывные аппроксимации пороговой функции потерь.

Метод стохастического градиента SG.

Метод стохастического среднего градиента SAG.

Частные случаи: адаптивный линейный элемент ADALINE, персептрон Розенблатта, правило Хэбба.

Теорема Новикова о сходимости. Доказательство теоремы Новикова

Эвристики: инициализация весов, порядок предъявления объектов, выбор величины градиентного шага, «выбивание» из локальных минимумов.

Проблема мультиколлинеарности и переобучения, регуляризация или редукция весов (weight decay).

Вероятностная постановка задачи классификации. Принцип максимума правдоподобия.

Вероятностная интерпретация регуляризации, совместное правдоподобие данных и модели. Принцип максимума апостериорной вероятности.

5. Лекция №5. Метод опорных векторов.

Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin).

Случаи линейной разделимости и отсутствия линейной разделимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь.

Понятие опорных векторов.

Рекомендации по выбору константы C.

Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера.

Способы конструктивного построения ядер. Примеры ядер.

SVM-регрессия.

6. Лекция №6. Многомерная линейная регрессия.

Задача регрессии, многомерная линейная регрессия.

Метод наименьших квадратов, его вероятностный смысл и геометрический смысл.

Сингулярное разложение.

Проблемы мультиколлинеарности и переобучения.

Регуляризация. Гребневая регрессия через сингулярное разложение.

Методы отбора признаков: Лассо Тибширани, Elastic Net, сравнение с гребневой регрессией.

Метод главных компонент и декоррелирующее преобразование Карунена-Лоэва, его связь с сингулярным разложением.

Спектральный подход к решению задачи наименьших квадратов.

Задачи и методы низкоранговых матричных разложений.

7. Лекция №7. Нелинейная регрессия.

Метод Ньютона-Рафсона, метод Ньютона-Гаусса.

Обобщённая аддитивная модель (GAM): метод настройки с возвращениями (backfitting) Хасты-Тибширани.

Логистическая регрессия. Метод наименьших квадратов с итеративным пересчётом

весов (IRLS). Пример прикладной задачи: кредитный скоринг. Бинаризация признаков. Скоринговые карты и оценивание вероятности дефолта. Риск кредитного портфеля банка.

Обобщённая линейная модель (GLM). Экспоненциальное семейство распределений. Неквадратичные функции потерь. Метод наименьших модулей. Квантильная регрессия. Пример прикладной задачи: прогнозирование потребительского спроса.

Робастная регрессия, функции потерь с горизонтальными асимптотами.

8. Лекция №8. Прогнозирование временных рядов.

Задача прогнозирования временных рядов. Примеры приложений.

Экспоненциальное скользящее среднее. Модель Хольта. Модель Тейла-Вейджа.

Модель Хольта-Уинтерса.

Адаптивная авторегрессионная модель.

Адаптивная селективная модель. Адаптивная композиция моделей.

Локальная адаптация весов с регуляризацией.

Основными стратегическими образовательными технологиями являются: дискуссионные, исследовательские, лекционные занятия, практические и лабораторные занятия.

В рамках указанных технологий применяются тактические образовательные технологии: задания, игровое проектирование, информационные, классическая лекция, проблемная лекция.