

Теоретический минимум по курсу «Математические основы теории прогнозирования» 2012

Незнание ответа на *любой* из нижеследующих вопросов автоматически влечет неудовлетворительную оценку за экзамен. Эти вопросы являются лишь необходимым требованием. Знание ответов только на эти вопросы не обеспечивает положительной оценки за экзамен.

1. В каких случаях применяются методы распознавания, методы регрессионного анализа?
2. Что такое обобщающая способность? Методы оценки обобщающей способности.
Что такое скользящий контроль?
3. Что такое эффект переобучения? Привести пример переобучения.
4. Что такое Байесовский классификатор?
5. Структура алгоритма распознавания. Что такое распознающий оператор?
6. Каковы цели использования ROC анализа в задачах распознавания.
5. Что такое метод максимального правдоподобия? Привести пример функции правдоподобия.
6. Что такое метод наименьших квадратов?
7. Дать определение математического ожидания, ковариации, дисперсии, коэффициента корреляций. Что такое ковариационная матрица для многомерных случайных величин.
8. Записать формулу для одномерного/многомерного нормального распределения.
9. Записать формулу для биномиального распределения? Что оно описывает?
10. Что такое метод k-ближайших соседей?
11. Что такое Линейный дискриминант Фишера?
12. Что такое многослойный перцептрон?
13. Основная идея метода Опорных векторов. К решению какой оптимизационной задачи сводится МОВ.
14. Дать определение теста и тупикового теста. Общая схема тестового алгоритма.
15. Что такое представительный набор.
16. Изложить общую схему алгоритма вычисления оценок.
Дать определение опорного множества и характеристического вектора опорного множества. распознавания.
17. Каким образом вводятся алгебраические операции над алгоритмами распознавания? Что такое линейное и алгебраическое замыкание множества алгоритмов распознавания?
18. Общая схема решающего дерева.
19. Общая схема методов распознавания, основанных на голосовании по системам закономерностей.
20. Цели использования кластерного анализа. Что такое иерархические методы кластеризации?
21. Основные цели методов анализа выживаемости (надёжности).

Программа экзамена.

1. Область применения методов, основанных на обучении по прецедентам (машинном обучении). Примеры применения. Понятие обучающей выборки. Способ обучения, основанный на минимизации эмпирического риска.
2. Понятие обобщающей способности. Для каких алгоритмов достигается максимум обобщающей способности. Байесовский классификатор.
3. Способы оценки обобщающей способности. Скользящий контроль.
4. Эффект переобучения.
5. Теоретические подходы к исследованию обобщающей способности
Теория Вапника-Червоненкиса. Понятие равномерной сходимости частот ошибок к их вероятностям. Условия достижения равномерной сходимости. Функции роста и её свойства. Понятие ёмкости множества алгоритмов и связь её с обобщающей способностью.
6. Метод линейная машина. Способ обучения.
7. Линейная регрессия. Использование метода наименьших квадратов для оценки коэффициентов. Связь МНК с методом максимального правдоподобия. Оценка параметров одномерной регрессии.
8. Поиск коэффициентов многомерной регрессии с помощью МНК. Формула для регрессионных коэффициентов. Явление мультиколлинеарности.
9. Свойства оптимальных регрессий.
10. Трёхкомпонентное разложение обобщённой ошибки. Смысл шумовой составляющей, составляющей сдвига и дисперсионной составляющей. Bias-Variance дилемма.
11. Байесовские методы обучения. Аппроксимация с помощью многомерного нормального распределения. Способ обучения.
12. Линейный дискриминант Фишера. Способ обучения.
13. Метод k-ближайших соседей. Способ обучения.
14. Логистическая регрессия. Способ обучения.
15. Распознавания при заданной точности распознавания одного из классов. Оптимальное решение согласно лемме Неймана-Пирсона.
16. ROC анализ. Основные цели метода. Способ построения ROC кривых.
17. Принцип частичной прецедентности. Понятие тупикового теста. Общая схема тестового алгоритма. Обобщение для вещественнозначной информации.
18. Понятие тупикового представительного набора. Общая схема алгоритма распознавания, основанного на тупиковых представительных наборах. Обобщение для вещественнозначной информации.
19. Модель Алгоритмов вычисления оценок. Понятия опорного множества, функции близости, для вычисления оценок за классы. Компактные формулы для оценок в случае, когда признаки равноправны, а мощность опорных множеств фиксирована. Способы обучения для модели ABO.
20. Модель искусственного нейрона. Печептрон Розенблатта и метод его обучения, условие сходимости.
21. Многослойный перцептрон и его структура. Аппроксимирующая способность многослойных перцептронов. Метод обратного распространения ошибки.
22. Ядерные методы оценки плотности распределения классов.

23. Метод опорных векторов. Концепция максимального “зазора”. Сведение к задаче квадратичного программирования. Условия, налагаемые теоремой Каруша-Куна-Таккера. Двойственная задача квадратичного программирования. Опорные вектора и их роль в формировании распознающего алгоритма.
24. Обобщение исходного варианта метода опорных векторов на случай отсутствия линейной делимости. “Смягчение” условия линейной делимости с помощью введения дополнительных переменных. Основные отличия от исходного варианта метода.
25. Обобщение метода опорных векторов, позволяющее строить нелинейные разделяющие поверхности.
26. Решающие деревья. Методы обучения. Индексы неоднородности. Критерии остановки ветвления. Методы “подрезки”.
27. Коллективные методы. Обоснование. Ошибка выпуклой комбинации алгоритмов прогнозирования. Простые комитетные методы. Наивный Байесовский классификатор.
28. Логическая коррекция. Монотонный логический корректор.
29. Алгебраические методы коррекции. Введение алгебраических операций над распознающими алгоритмами. Линейное и алгебраическое замыкание. Полнота задачи распознавания относительно множества операторов. Условие существования корректного алгоритма в линейном замыкании и алгебраическом замыкании.
30. Линейное и алгебраическое замыкание над моделями алгоритмов. Корректность алгебраического замыкания модели АВО относительно регулярных задач.
31. Коллективные методы основанные на бутстрэп репликациях. Метод бэггинг.
32. Метод бустинг.
33. Методы, основанные на голосовании по системам логических закономерностей. Полные и частичные логические закономерности. Методы поиска. Коллективное решение.
34. Метод «Статистически взвешенные синдромы». Оптимальные разбиения в рамках фиксированных моделей. Коллективное решение.
35. Методы кластеризации. Цели кластерного анализа. Метод k-внутригрупповых средних. Алгоритм Форель. Иерархические методы кластеризации.
36. Коллективные решения в кластерном анализе. Сумматоры. Контрастные матрицы. Поиск оптимального коллективного решения.
37. Методы многомерной визуализации.
38. Метод главных компонент. Метод поиска и свойства главных компонент.
39. Методы анализа выживаемости (надёжности). Оценки кривых выживаемости по методу Каплан-Майера. Модель Кокса.
40. Прогнозирование временных рядов. Метод скользящего среднего. Использование методов машинного обучения при прогнозировании временных рядов. Поиск циклов.