

Вопросы к экзамену по курсу «Математические основы теории прогнозирования» 2010

1. Различные постановки задач машинного обучения (классификация, регрессия, кластеризация, идентификация, прогнозирование, поиск закономерностей). Основные проблемы в теории машинного обучения: переобучение, некорректность данных, малый объем обучающей информации.
2. Метод максимального правдоподобия. Его достоинства и недостатки. Нормальное распределение (одномерное и многомерное), его основные свойства.
3. Задача восстановления линейной регрессии. Метод наименьших квадратов. Решение несовместных СЛАУ.
4. Задача восстановления линейной регрессии. Вероятностная формулировка метода наименьших квадратов.
5. Логистическая регрессия. Вероятностная постановка. Регуляризация обучения. Итеративный метод наименьших квадратов.
6. Основные правила работы с вероятностями. Условная независимость случайных величин. Примеры.
7. Графические модели. Основные задачи, возникающие в анализе графических моделей.
8. Байесовские сети. Примеры.
9. Марковские сети. Примеры.
10. Скрытые марковские модели. Обучение СММ с учителем.
11. Алгоритм динамического программирования и его применение в скрытых марковских моделях.
12. EM-алгоритм. Его применение для задачи разделения гауссовской смеси.
13. EM-алгоритм и его применение в скрытых марковских моделях.
14. Условная независимость в скрытых марковских моделях. Алгоритм «вперед-назад».
15. Уменьшение размерности в данных. Метод главных компонент. Выбор размерности пространства.
16. Вероятностная формулировка метода главных компонент. Аналитическая точка максимального правдоподобия.

17. Вероятностный метод главных компонент. Поиск решения с помощью EM-алгоритма. Учет пропусков в данных.
18. Ограничения метода главных компонент. Вероятностная смесь главных компонент. Анализ независимых факторов.
19. Поиск максимального потока в сетях. Сведение задачи минимизации энергии марковской сети в случае бинарных переменных к задаче поиска наименьшего разреза графа.
20. Задача минимизации энергии марковской сети в случае K-значных переменных. Примеры. Ее решение с помощью альфа-расширения.
21. Алгоритм «Кора». Определить, к какому классу будет отнесен объект $(1, 1, \dots, 1)$, если обучающая выборка имеет следующий вид:

$$\begin{array}{l}
 K_1 \\
 K_2
 \end{array}
 \left\{ \begin{array}{l}
 (0, 1, 1, \dots, 1, 1), \\
 (1, 0, 1, \dots, 1, 1), \\
 \dots\dots\dots \\
 (1, 1, 1, \dots, 1, 0), \\
 (1, 1, \dots, 1, 1, 0), \\
 (1, 1, \dots, 1, 0, 1), \\
 \dots\dots\dots \\
 (0, 1, \dots, 1, 1, 1).
 \end{array} \right.$$

22. Тестовый алгоритм. Система уравнений, определяющих тупиковые тесты. Таблица обучения с двумя классами объектов. Основные упрощающие формулы при умножении левых частей тестовых уравнений.
23. Обоснование способа построения всех тупиковых тестов через приведение системы тестовых уравнений к неупрощаемой ДНФ.
24. Алгоритм вычисления оценок. Шаги определения, полный перечень параметров. Итоговая формула для числа голосов.
25. В алгоритме вычисления оценок написать формулу для числа голосов, если система опорных множеств состоит всех непустых подмножеств, а функция близости определяется только $\varepsilon_1, \dots, \varepsilon_n$.
26. Теорема о представлении алгоритма как композиции распознающего оператора и фиксированного решающего правила. Отмеченные точки в контрольной матрице. Построение корректного алгоритма, если найдена совокупность операторов, отмечающих все единичные точки информационных векторов контрольной матрицы.