

Лабораторная работа №3. Классификация

Цель работы: изучение методов решения задачи классификации.

Среда выполнения: SPSS Statistics, Statistica, Deductor.

Задание

1. Выбрать массив данных (ссылки № 3-4 из списка литературы), описать параметры. Примечание: тип задачи «классификация».
2. Провести дескриптивный анализ, определить, согласуются ли выборки с нормальным распределением.
3. Стандартизировать переменные (при необходимости).
4. Изучить пример решения задачи классификации (ссылка №2 из списка литературы).
5. Решить задачу классификации двумя методами (нейронная сеть, дерево решений, метод опорных векторов, дискриминантная функция и т.д.). Аргументировать выбор метода. Допустимо использовать две нейронные сети разных типов.
6. Оценить качество построенных моделей (в т.ч. точность, полноту классификатора). Провести сравнительный анализ решений.
7. Изменить параметры одного из методов (или параметры нейросети), оценить полученные результаты.
8. Интерпретировать результаты.
9. Оформить отчет.

Содержание отчета

1. Титульный лист.
2. Цель работы.
3. Описание исходных данных.
4. Результаты дескриптивного анализа.
5. Описание параметров метода (сети).
6. Сравнительный анализ решений и интерпретация результатов.

Список литературы и ссылки на материалы

1. Айвазян С.А. Методы эконометрики. – М.: Магистр: - ИНФРА-М, 2010. – 512 с.
2. Пример решения задачи классификации в Statistica. http://statsoft.ru/solutions/ExamplesBase/tasks/detail.php?ELEMENT_ID=261
3. Массивы данных. <https://www.kdnuggets.com/datasets/index.html>
4. Массивы данных. <http://archive.ics.uci.edu/ml/datasets.html>
5. Обзор методов классификации. <https://tproger.ru/translations/scikit-learn-in-python/>

6. Постановка задачи и конспекты лекций:
<http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%B%D0%B0%D1%81%D1%81%D0%B8%D1%84%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F>
7. Оценка качества классификатора
<http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html#:~:text=%D0%A2%D0%BE%D1%87%D0%BD%D0%BE%D1%81%D1%82%D1%8C%20%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D1%8B%20%D0%B2%20%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B0%D1%85%20%D0%BA%D0%BB%D0%B0%D1%81%D1%81%D0%B0,%D1%8D%D1%82%D0%BE%D0%B3%D0%BE%20%D0%BA%D0%BB%D0%B0%D1%81%D1%81%D0%B0%20%D0%B2%20%D1%82%D0%B5%D1%81%D1%82%D0%BE%D0%B2%D0%BE%D0%B9%20%D0%B2%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B5.>
8. Нейросети в Statistica
http://statsoft.ru/products/STATISTICA_Neural_Networks/first-steps.php

Вопросы к защите

1. Бинарная классификация. Многоклассовая классификация.
2. Методы решения задачи классификации. Метод k ближайших соседей. Метод опорных векторов. Деревья решений. Леса (джунгли). Bagging. Boosting.
3. Метрики качества классификации. Доля правильно распознанных ответов, точность, полнота. PR-, ROC-кривая. Оценка качества многоклассовой классификации.
4. Дискриминантный анализ. Дискриминантная функция.
5. Нейронные сети. Виды активационных функций. Алгоритм обратного распространения ошибки.