

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Дисциплина: «Анализ данных»

Домашнее задание на тему:
«Лабораторная работа №14»

Выполнил: Осипов Лев,
студент группы 301ПИ (1).

Москва, 2015 г.

СОДЕРЖАНИЕ

Теоретическая часть.....	3
Задание 3	3
Задание 4	3
Список литературы	4

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

ЗАДАНИЕ 3

Можно попробовать в соответствии с весом объекта задать количество таких объектов в выборке. В случае нецелого веса – округлять. Если все веса будут располагаться, предположим, в интервале 0.1 – 0.4, можно преобразовать эти значения в целые (простейший способ – взять десятую или сотую часть числа). В общем, из заданных весов совершить как можно более пропорциональные преобразования к целым числам и задать такое количество этих объектов в выборке.

Также можно сформировать несколько выборок, в которых объекты будут входить с вероятностью, которая зависит от их весов (перед этим веса надо нормировать), а потом объединить эти выборки. Правда, для этого надо подбирать некое эвристическое значение, являющееся критерием (значением вероятности) для попадания объектов в выборки.

ЗАДАНИЕ 4

Для пост-редукции может просто не хватить объектов, так как она использует контрольную подвыборку, а если учесть, что каждый алгоритм композиции уже имеет в своем распоряжении подвыборку, количество объектов для пост-редукции может оказаться весьма мало.

Если смотреть глубже, по своей сути дерево решений – классификатор нестабильный, и поэтому малые изменения в выборках приводят к существенным изменениям в дереве. Пост-редукция – неплохой инструмент для снижения переобученности алгоритма, но ее использование в композиции, а если быть точнее, в бэггинге, где алгоритмы обучаются на разных подвыборках, приводит к тому, что уменьшается разница между этими алгоритмами (а цель бэггинга как раз в том, чтобы использовать различные алгоритмы).

СПИСОК ЛИТЕРАТУРЫ

- 1) **Анализ данных (Программная инженерия) –**
[http://wiki.cs.hse.ru/Анализ_данных_\(Программная_инженерия\)](http://wiki.cs.hse.ru/Анализ_данных_(Программная_инженерия))