

В классе мы обсуждали кластерный анализ, автоматический отбор предикторов, различные регуляризации и SVM.

Просьба присылать результаты в виде PDF- или HTML-отчетов. Кросс-валидация работает очень долго, я не смогу запустить код каждого из вас.

1 Домашнее задание

Задание 1.1 (возврат к истокам, незачетное). Осмыслить новые сведения применительно к старым задачам.

1. Использовать регуляризацию и best subset selection для старых задач регрессии (`Universities`, `Advertising`, `concrete`)
2. Использовать кластеризацию для решения `mnist` в качестве предобработки
3. Использовать SVM для старых задач классификации (`parkinsons`, `seeds`)

Задание 1.2 (SVM, простое). Используя SVM, построить классификатор для данных `GlaucosomaMVF`. Сравнить между собой разные ядра (линейное, полиномиальное, гауссовское), объяснить результат.

Данные аналогичны разобранным к классе, поэтому задание отмечено как “простое”

Задание 1.3 (SVM). Используя SVM, построить классификатор для данных `Diabetis`. Сравнить между собой разные ядра (линейное, полиномиальное, гауссовское), объяснить результат.

Задание 1.4 (SVM, сложнее). Используя SVM, построить классификатор для данных `hepatitis` (сегодня у нас только медицинские примеры=).

Будьте внимательны, в этих данных различное число больных и здоровых. Необходимо использовать веса классов (параметр `class.weights`), другую меру ошибки и механизм ROC-кривых для оценки классификатора (аналогично тому, как мы делали в LDA-QDA, можно доработать методы из конспекта). Чтобы извлечь из SVM непрерывную оценку вероятности, при обучении необходимо задать параметр `probability = TRUE` (для `svm`) или `prob.model = TRUE` (для `ksvm`). Пример построения ROC-кривой есть в конспекте.