

Линейные методы

Набор данных

- Выберите любой набор данных для задачи бинарной классификации. Желательно использовать набор данных из предыдущей лабораторной работы. Если он был не для бинарной классификации, то объедините несколько классов.
- Преобразуйте его в числовой вид и нормализуйте.
- Разбейте его на тренировочную и тестовую часть.
- Выберите целевую функцию ошибки или качества для задачи бинарной классификации.
- Если после преобразования в числа в наборе данных слишком много признаков, выберите часть признаков из набора данных любым способом. Например, при помощи абсолютной корреляции с целевым (с заменой на ± 1).

Алгоритмы

- Реализуйте алгоритм линейной регрессии в матричном виде с гребневой регуляризацией. Можно использовать библиотеки для работы с матрицами. Преобразуйте его в алгоритм линейной классификации через замену целевого признака на ± 1 .
- Реализуйте алгоритм линейной классификации на основе градиентного спуска. Алгоритм должен поддерживать не менее трёх эмпирических рисков, Elastic Net регуляризацию и настраиваемую скорость градиентного спуска. Эмпирические риски должны быть именно для задачи линейной классификации, вычисляться через отступ. Производные и градиент должны быть вычислены аналитически.

Задача

- Постарайтесь подобрать хорошие гиперпараметры для каждого алгоритма.
- Постройте кривую обучения со слаженным эмпирическим риском на тренировочном множестве для линейной классификации.
- Постройте кривую обучения с целевой функцией ошибки или качества на тестовом множестве для линейной классификации. Не обязательно замерять целевую функцию на каждой итерации, если их много. Переберите разные разбиения на тренировочную и тестовую часть как при перекрёстной проверке. Отметьте доверительный интервал на графике.
- Отметьте на предыдущем графике значение целевой функции на тестовом множестве для линейной регрессии. Например, горизонтальной прямой.
- Постройте график зависимости значений коэффициентов линейной модели от значения коэффициента L_1 регуляризации (коэффициент для L_2 должен быть равен нулю). Постройте второй график для зависимости от коэффициента L_2

регуляризации (для L_1 коэффициент равен нулю). Желательно перебирать коэффициенты регуляризации в логарифмированном пространстве.