

Задачи по линейным методам для коллоквиума

Задача 1. У Вас есть набор данных из 1000 объектов, описанных 10 признаками. Вы обучаете модель линейной регрессии с константным признаком. Вы рассматриваете 2 модификации модели: (а) включение Lasso-регуляризации и (б) добавление к исходным признакам квадратов каждого признака. Таким образом, у Вас может получиться 4 модели: () — без регуляризации с обычными признаками, (а), (б) и (аб). Для удобства запишем эти 4 варианта в виде таблицы:

	Без регуляризации	С регуляризацией
Исходные признаки	()	(а)
Исх. + квад. признаки	(б)	(аб)

Заполните 2 таблицы:

Сколько параметров нужно настроить (сколько чисел)?

	Без регуляризации	С регуляризацией
Исходные признаки		
Исх. + квад. признаки		

Сколько гиперпараметров нужно настроить (сколько чисел)?

	Без регуляризации	С регуляризацией
Исходные признаки		
Исх. + квад. признаки		

Задача 2. Вы решаете задачу бинарной линейной классификации в трехмерном признаковом пространстве. Соответственно, решающее правило имеет вид:

$$a(x) = \text{sign}(w_0 + w_1x_1 + w_2x_2 + w_3x_3)$$

У Вас есть следующие объекты:

x_1	x_2	x_3	y
0.2	0.4	0	+1
0.5	0.9	0	+1
0.3	0.3	0	+1
0.1	0.8	1	-1
0.5	0.7	1	-1
0.9	0.9	1	-1
0.1	0.3	1	-1

(а) Сколько существует наборов коэффициентов (w_0, w_1, w_2, w_3) , при которых задача будет решена идеально (с ассигасу=1)?

(б) Если ответ в п. (а) — 0, пропустите это задание. Если ответ в п. (а) — 1 или 2, запишите этот набор коэффициентов (или 2 набора). Иначе приведите хотя бы три таких набора коэффициентов.

Задача 3. Вы решаете задачу бинарной линейной регрессии в двумерном признаковом пространстве. Соответственно, решающее правило имеет вид:

$$a(x) = w_0 + w_1x_1 + w_2x_2$$

У Вас есть следующие объекты:

x_1	x_2	y
1	1	0
0	0	1
0.5	0.5	0.5

(а) Сколько существует наборов коэффициентов (w_0, w_1, w_2) , при которых задача будет решена идеально (с $\text{MSE}=0$)?

(б) Если ответ в п. (а) — 0, пропустите это задание. Если ответ в п. (а) — 1 или 2, запишите этот набор коэффициентов (или 2 набора). Иначе приведите хотя бы три таких набора коэффициентов.

Задача 4. Пусть в задаче бинарной классификации (отнести каждый объект к одному из двух классов) классификатор $a(x)$ предсказывает вероятность того, что объект x принадлежит классу $+1$. Как известно из лекций, в этом случае для настройки параметров можно использовать следующую функцию потерь:

$$L(a(x), y) = [y = +1] \log_2 a(x) + [y = -1] \log_2(1 - a(x)). \quad (0.1)$$

Теперь представим, что мы хотим решить задачу тегирования: для каждого объекта определить, к каким из двух классов он принадлежит. Таким образом, каждый объект может принадлежать к одному, двум классам или ни к одному. Пусть наш классификатор возвращает две вероятности — принадлежности первому и второму классу: $a(x) = (p_{-1}(x), p_{+1}(x))$. Предложите, как можно изменить функцию потерь (0.1), чтобы использовать ее для настройки параметров в новой задаче.

Задача 5. Для задачи линейной регрессии с L_2 -регуляризацией с одним признаком запишите формулу для определения веса при этом признаке с помощью оптимизации среднеквадратичной ошибки. Считайте, что свободный член $w_0 = 0$ (его не нужно настраивать). В качестве ответа приведите формулу, в которой участвуют следующие величины:

- вектор $x = (x_1^1, \dots, x_\ell^1)$ значений признака для всех объектов $1 \dots \ell$,
- вектор $y = (y_1, \dots, y_\ell)$ правильных ответов для всех объектов,
- C — коэффициент регуляризации.