



# Инструкция по рецензированию

## Файлы решения

Linear Models.ipynb (/solution\_file/19943/Linear Models.ipynb)

## Рецензии

<b>Вопрос 1. Как зависят параметры модели от константы регуляризации? А качество?</b>		
1.0 балл	Дан правильный ответ на вопрос. Качество на обучающей выборке падает с ростом константы регуляризации. Параметр $b$ растёт, параметр $w$ уменьшается	<div>Рецензия №1</div> <div>Рецензия №2</div> <div>Рецензия №3</div>
0.0 баллов	Дан неправильный ответ/ответа нет	---

<b>Вопрос 2. Почему при обучении линейных моделей, коэффициент <math>b</math> не регуляризуется?</b>		
1.0 балл	Дан правильный ответ на вопрос. Можем считать правильным любой ответ, близкий по смыслу идее, что нет смысла штрафовать константную составляющую прогноза (ибо она не переобучается), а вот если за неё штрафовать, то модель будет стремиться давать нулевые ответы с ростом константы регуляризации. Если есть упоминание несмещённости линейной модели, то вообще отлично.	<div>Рецензия №1</div> <div>Рецензия №2</div> <div>Рецензия №3</div>
0.0 баллов	Дан неправильный ответ на вопрос/ответа нет.	---

<b>Вопрос 3. Рассмотрите больше примеров с двумя признаками <math>x_1</math> и <math>x_2</math>, где <math>x_2</math> линейно зависит от <math>x_1</math>. Убедитесь, что линейная модель без регуляризации крайне неустойчива.</b>		
2.0 балла	Код написан верно. Примеры, где абсолютное значение весов больше миллиона есть. (Самые очевидные примеры, когда просто $x_2 = c * x_1$ , и перебираются разные значения $c$ )	<div>Рецензия №1</div> <div>Рецензия №2</div> <div>Рецензия №3</div>
0.0 баллов	Код написан неверно/примеров , где абсолютное значение весов больше миллиона - нет.	---

<b>Вопрос 4. Как думаете, почему отношение между весами постоянно?</b>		
--	--	--



2.0 балла

Дано строгое математическое обоснование (например, правильно посчитаны частные производные по весам, и через приравнивание к нулю выведено искомое отношение).

Рецензия №1

Рецензия №2

Рецензия №3

1.0 балл

Дано интуитивное обоснование, либо в математическом обосновании есть ошибки, не повлиявшие на результат

---

0.0 баллов

Обоснование ошибочно/отсутствует

---

**Вопрос 5. Почему в нашем примере l1 регуляризация приведёт к разреживанию весов?**

2.0 балла

Дано строгое математическое обоснование (например через приравнивание частных производных в ненулевых весах к нулю или через субградиенты).

Рецензия №1

Рецензия №2

Рецензия №3

1.0 балл

Дано интуитивное обоснование, либо в математическом обосновании есть ошибки, не повлиявшие на результат

---

0.0 баллов

Обоснование ошибочно/отсутствует

---

**Вопрос 6. Почему в итоге при  $\alpha=0.00001$  получились веса не равные нулю?**

2.0 балла

Дан правильный ответ (не хватило числа итераций). Приведён код, где показано, что число итераций совпадает с максимальным и/или модель обучена до конца.

Рецензия №1

Рецензия №3

1.0 балл

Дан правильный ответ, но без кода.

Рецензия №2

0.0 баллов

Неправильный ответ/ответа нет

---

**Вопрос 7. Приведите пример, когда линейная регрессия с l2 регуляризацией гарантированно занулит какой-нибудь признак?**

2.0 балла

Дан правильный ответ (например, константный признак)

Рецензия №1

Рецензия №2

Рецензия №3

0.0 баллов

Дан неправильный ответ/ответа нет

---



**Вопрос 8. Почему LBFGS является методом по умолчанию, а не Newton-CG?**

2.0 балла

Дан правильный ответ. Newton-CG требует намного больше вычислений на одну итерацию, из-за чего с увеличением числа признаков он становится очень медленным (достаточно назвать эту причину). Также обычно мы нормализуем данные, а L-BFGS хорошо сходится на нормализованных данных.

Рецензия №1

Рецензия №3

0.0 баллов

Дан неправильный ответ/ответа нет

Рецензия №2

**Вопрос 9. Что это за страна?**

1.0 балл

Дан правильный ответ (Австралия)

Рецензия №1

Рецензия №2

Рецензия №3

0.0 баллов

---

Дан неправильный ответ/ответа нет

**Вопрос 10. Попробуйте объяснить для каких-нибудь признаков получившиеся значения корреляции?**

1.0 балл

Дано адекватное пояснение для хотя бы двух признаков. Например, признак Sunshine означает количество солнечного света, и ясно, что чем солнечнее было сегодня, тем менее вероятен дождь завтра.

Рецензия №1

Рецензия №2

Рецензия №3

0.5 баллов

---

Дано адекватное пояснение только для одного признака

0.0 баллов

---

Все пояснения неправильные/пояснений нет совсем

**Вопрос 11. Сколько итераций понадобилось для обучения логистической регрессии?**

1.0 балл

Выведено число итераций (должно быть много, около 400-600). Число итераций доступно через `model.n_iter_`

Рецензия №1

Рецензия №2

Рецензия №3

0.0 баллов

---

Дан неправильный ответ/ответа нет

**Вопрос 12. А сколько итераций потребовалось теперь? Как это связано с модельным примером?**



2.0 балла

Выведено число итераций (должно быть около 30-35). Указано, что в модельном примере метод плохо сходилсся также из-за неотмасштабированных данных

Рецензия №1

Рецензия №2

Рецензия №3

1.0 балл

---

Дана половина ответа (либо число итераций, либо про модельный пример)

0.0 баллов

---

Дан неправильный ответ/ответа нет

**Вопрос 13. Сравните получившееся качество с качеством до нормализации. Объясните, почему оно изменилось, хотя мы применяли линейную модель к линейно преобразованным признакам? Могло ли качество измениться в другую сторону?**

2.0 балла

Дан правильный ответ. Качество выросло и изменилось из-за регуляризации, которая по разному влияет на веса из-за разного масштаба признаков. Качество изменилось из-за регуляризационного слагаемого в функции потерь. Качество могло и упасть.

Рецензия №1

1.0 балл

Правильный ответ дан на часть пунктов (хотя бы на один)

Рецензия №2

Рецензия №3

0.0 баллов

---

Дан неправильный ответ/ответа нет

**Вопрос 14. Почему даже после нормализации график не до конца похож на гистограмму корреляций?**

1.0 балл

Дан разумный ответ на вопрос (например, гистограмма корреляций не учитывает важность признаков в контексте других признаков, и может завышать важность тех или иных признаков)

Рецензия №1

Рецензия №3

0.0 баллов

Дан неправильный ответ/ответа нет

Рецензия №2

**Вопрос 15. Чем отличаются валидационная и тестовая выборки?**

1.0 балл

Дан правильный ответ. По валидационной выборке мы подбираем гиперпараметры. На тестовой выборке мы оцениваем финальное качество модели.

Рецензия №1

Рецензия №3

0.0 баллов

Дан неправильный ответ/ответа нет

Рецензия №2



**Вопрос 16. Почему с ростом  $C$  качество не падает, хотя модель должна недообучаться?**

1.0 балл

---

Дан правильный ответ (в sklearn константа регуляризации равна  $1 / C$ , а из-за большого числа объектов даже слабо регуляризованные модели хороши по качеству)

0.0 баллов

Дан неправильный ответ/ответа нет

Рецензия №1

Рецензия №2

Рецензия №3

**Вопрос 17. Почему мы не подбираем коэффициент регуляризации  $C$  по обучающей выборке? По тестовой выборке?**

2.0 балла

Рецензия №1

Дан правильный ответ. Потому что в первом случае мы рискуем переобучиться, а во втором получить нечестное качество (выше настоящего).

1.0 балл

---

Правильный ответ дан только про обучающую или тестовую выборки.

0.0 баллов

Рецензия №2

Рецензия №3

Дан неправильный ответ/ответа нет

**Вопрос 18. Почему качество не выросло?**

2.0 балла

Рецензия №3

Дан правильный ответ (нужно показать, что реально AUC увеличился только в 5-6 знаках, а мы округляем до 4-го)

0.0 баллов

Рецензия №1

Рецензия №2

Дан неправильный ответ/ответа нет

**Вопрос 19. Счётчики показывают худшее качество по AUC ROC, нежели OneHotEncoding. В чём же преимущества счётчиков?**

1.0 балл

Рецензия №1

Рецензия №3

Дан правильный ответ. В некоторых ситуациях one-hot кодирование приводит к появлению слишком большого числа признаков, что влечёт вычислительные проблемы.

0.0 баллов

Рецензия №2

Дан неправильный ответ/ответа нет

**Вопрос 20. Можно ли использовать  $l_1$  регуляризацию в логистической регрессии?**



1.0 балл  
Дан правильный ответ. Да можно, в классе LogisticRegression  
есть даже параметр "penalty".

Рецензия №1	Рецензия №2
Рецензия №3	

0.0 баллов  
Дан неправильный ответ/ответа нет

---

Оцените, насколько понятным для вас был код  
автора

0.0 баллов  
Все понятно

Рецензия №1	Рецензия №2
Рецензия №3	

0.0 баллов  
В целом понятно, были неясные места

---

0.0 баллов  
Есть, куда расти

---

Оцените, насколько понятным для вас было оформление ноутбука  
автора в целом

0.0 баллов  
Все понятно

Рецензия №1
Рецензия №3

0.0 баллов  
В целом понятно, были неясные места

Рецензия №2
-------------

0.0 баллов  
Есть, куда расти

---

Комментарии

Рецензия №1
Рецензия №2
Рецензия №3 ты молодец