



Линейные модели: статистический взгляд



7/7 баллов получено (100%)

Тест пройден!

Вернуться к неделе 3



Баллов: 1
/ 1

1.

При комплексном обследовании нескольких тысяч человек по измерявшимся показателям (включая пульс, давление, ЭКГ и т.д.) оценивался риск возникновения сердечного заболевания. Ста пациентам с самым высоким риском была предложена оздоровительная программа, включающая диету, упражнения и приём профилактических препаратов. Через несколько месяцев после окончания программы пациенты снова прошли диспансеризацию; средний оцениваемый риск возникновения сердечного заболевания существенно уменьшился.

Что можно сказать об эффективности оздоровительной программы?



Программа явно эффективна — риск уменьшился, значит, пациенты стали здоровее!



Данных недостаточно: поскольку были выбраны пациенты с наибольшим риском, измеренный эффект может объясняться регрессией к среднему. Для оценки эффективности программы нужно использовать контрольную группу пациентов с таким же высоким риском, для которых программа не проводилась, и сравнить изменения в двух группах.



Правильный ответ

Баллов: 1
/ 1

2.

Из 15 клиентов банка, которым менеджер предложил подключить автоплатёж, четверо согласились. Подключение услуги — бинарный признак, который можно описать распределением Бернулли. Запишите функцию правдоподобия $L(X^n, p)$ для такой выборки и посчитайте её значение, если истинная вероятность подключения услуги $p = 0.2$. Запишите ответ с пятью знаками после десятичной точки.

Правильный ответ

$$L(X^n, p) = \prod_{i=1}^n p[X_i = 1](1 - p)[X_i = 0] = p^4 \cdot (1 - p)^{11},$$

$$L(X^n, 0.2) = 0.2^4 0.8^{11} \approx 0.00014.$$

Баллов: 1
/ 1

3.

По выборке из предыдущей задачи найдите оценку максимального правдоподобия для параметра p . Запишите ответ с тремя знаками после десятичной точки.

Правильный ответ

В данном случае $\hat{p}_{\text{ОМП}} = \bar{X}_n = \frac{4}{15} \approx 0.267$.

Баллов: 1
/ 1

4.

Выберите верные утверждения о регрессии, получаемой методом наименьших квадратов.



Если шум описывается нормальным распределением с нулевым средним и постоянной дисперсией, метод наименьших квадратов даёт оценку максимального правдоподобия.

Правильный ответ



Получаемая оценка приближает условную по x медиану отклика y .



Правильный ответ



Если шум описывается лапласовским распределением с нулевым средним и постоянной дисперсией, метод наименьших квадратов даёт оценку максимального правдоподобия.



Правильный ответ



Получаемая оценка приближает условное по x матожидание отклика y .



Правильный ответ



Баллов: 1
/ 1

5.

Какие из приведённых ниже средств помогают от переобучения линейных моделей?



Регуляризаторы.



Правильный ответ



Упрощение модели.



Правильный ответ



Увеличение количества данных.



Правильный ответ



Использование средней абсолютной ошибки вместо среднеквадратичной.



Правильный ответ



Усложнение модели.

Правильный ответ



Баллов: 1
/ 1

6.

Выберите верные утверждения о регрессии, получаемой методом наименьших квадратов с L_1/L_2 регуляризаторами.



Оба типа регуляризаторов уменьшают по модулю веса признаков, но L_2 некоторые ещё и обнуляет.

Правильный ответ

Наоборот, обнуляет веса только L_1 -регуляризатор.



Из-за регуляризаторов оценки коэффициентов модели получаются смещённые.

Правильный ответ



Константное слагаемое не должно входить в регуляризатор.

Правильный ответ



Регуляризаторы увеличивают дисперсию оценок коэффициентов регрессионной модели.

Правильный ответ



Регуляризация подходит только для линейной регрессии, к сожалению, в задачах остальных типов её использовать нельзя.

Правильный ответ

Обратный пример — логистическая регрессия.



При использовании L_2 -регуляризатора в МНК решение можно найти аналитически, а с L_1 -регуляризатором — только численно.



Правильный ответ



Регуляризаторы позволяют бороться с переобучением модели.



Правильный ответ



Баллов: 1
/ 1

7.

Выберите верные утверждения о логистической регрессии.



Оценка параметров модели делается минимизацией кросс-энтропии.



Правильный ответ

Минимизация кросс-энтропии — то же самое, что максимизация правдоподобия.



Логистическая регрессия позволяет построить линейную модель вероятности $P(y = 1|X)$



Правильный ответ

Линейная модель строится для логита $\ln \frac{P(y=1|X)}{P(y=0|X)}$



При настройке моделей могут возникать проблемы, если классы линейно разделимы в пространстве признаков.



Правильный ответ

В таком случае $\|w\| \rightarrow \infty$



Функция, минимизируемая при настройке параметров модели, невыпуклая, у неё много локальных экстремумов, поэтому задача всегда имеет много решений.



Правильный ответ

Такая проблема возникает только в случае, если матрица X вырождена.

