Закрыть

Линейные модели: статистический взгляд

7 вопросов

1  
point

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.

1  
point

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*(*Xn*,*p*) для такой выборки и посчитайте её значение, если истинная вероятность подключения услуги *p*=0.2. Запишите ответ с пятью знаками после десятичной точки.



1  
point

3.

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



1  
point

4.

Выберите верные утверждения о регрессии, получаемой методом наименьших квадратов.



Получаемая оценка приближает условное по *x* матожидание отклика *y*.



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



Получаемая оценка приближает условную по *x* медиану отклика *y*.



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

1  
point

5.

Какие из приведённых ниже средств помогают от переобучения линейных моделей?



Использование средней абсолютной ошибки вместо среднеквадратичной.



Регуляризаторы.



Усложнение модели.



Упрощение модели.



Увеличение количества данных.

1  
point

6.

Выберите верные утверждения о регрессии, получаемой методом наименьших квадратов с *L*1/*L*2 регуляризаторами.



Регуляризаторы позволяют бороться с переобучением модели.



Из-за регуляризаторов оценки коэффициентов модели получаются смещённые.



При использовании *L*2-регуляризатора в МНК решение можно найти аналитически, а с *L*1-регуляризатором — только численно.



Регуляризаторы увеличивают дисперсию оценок коэффициентов регрессионной модели.



Оба типа регуляризаторов уменьшают по модулю веса признаков, но *L*2некоторые ещё и обнуляет.



Константное слагаемое не должно входить в регуляризатор.



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

1  
point

7.

Выберите верные утверждения о логистической регрессии.



Оценка параметров модели делается минимизацией кросс-энтропии.



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



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



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