



В тестовых заданиях первая галочка — правильный ответ, вторая галочка — выбранный ответ. Цвет обозначает, правильно ли в данном пункте поставлена галочка. Если все пункты верные (галочки совпадают / все пункты зеленые), то за задание ставится полный балл, в противном случае ставится 0 баллов.

1. Что из нижеперечисленного приводит к увеличению дисперсии и уменьшению смещения получаемой модели:

- уменьшение числа главных компонент в PLS регрессии
- такое изменение параметра L1 или L2 регуляризации, что отношение нормы вектора коэффициентов полученной регрессии к норме вектора коэффициентов регрессии МНК стремится к 0
- такое изменение параметра L1 или L2 регуляризации, что отношение нормы вектора коэффициентов полученной регрессии к норме вектора коэффициентов регрессии МНК стремится к 1
- увеличение числа главных компонент в PLS регрессии
- удаление переменных пошаговым методом
- добавление переменных пошаговым методом

**Балл: 0**

**Комментарий к правильному ответу:**

2. В случае небольшой выборки для оценки параметров логистической регрессии используется метод наименьших квадратов с итеративным пересчетом весов (IRLS), который:

- является итерационным ньютоновским методом оптимизации
- является итерационным градиентным методом оптимизации
- является обычным (не итерационным) методом наименьших квадратов, только с разными весами для разных наблюдений

**Балл: 0**

**Комментарий к правильному ответу:**

3. Предположим, что все признаки линейной регрессии независимы. Верно ли, что если знак коэффициента  $b(i)$  при признаке  $x(i)$  линейной регрессии положителен, то увеличение значения признака всегда приводит к увеличению отклика  $y$ ?

  да  нет**Балл: 1.0****Комментарий к правильному ответу:**

4. Зачем используют матричные разложения (например, SVD) при поиске коэффициентов линейной регрессии МНК:

  Для повышения устойчивости вычислений (для уменьшения влияния накопленных погрешностей вычислений)  Для повышения устойчивости модели (для уменьшения влияния выбросов)  Для уменьшения сложности вычислений  Для борьбы с мультиколлинеарностью  Для борьбы с переобучением  Для уменьшения сложности модели**Балл: 0****Комментарий к правильному ответу:**

5. Могут ли меняться прогнозы линейного многоклассового классификатора без регуляризации при изменении масштаба отдельных признаков и последующего повторного обучения модели? (пренебрегаем ограниченной точностью вычислений компьютера)

  нет  да**Балл: 0****Комментарий к правильному ответу:**

6. Если у вас есть набор данных с небольшим числом наблюдений и большим числом признаков, то какой метод построения не переобученной линейной регрессии вы выберете, если хотите сохранить в финальной модели все исходные предикторы:



регрессию главных компонент

PLS регрессию

прямой пошаговый метод

обратный пошаговый метод

регуляризацию L2

регуляризацию L1

**Балл: 0**

**Комментарий к правильному ответу:**

7. Если построена логистическая регрессия с линейным уравнением зависимости логита от предикторов, то:

для разных точек пространства признаков в рамках одного класса вероятности положительного отклика могут быть разными

для разных точек пространства признаков направление роста вероятности положительного отклика может быть разным

для всех точек пространства признаков направление роста вероятности положительного отклика всегда одинаково

для всех точек пространства признаков в рамках одного класса значение вероятности положительного отклика всегда одинаковое

**Балл: 0**

**Комментарий к правильному ответу:**

8. Если вы построили модель бинарной классификации с некоторой дискриминантной функцией и задали два значения порога  $t_1$  и  $t_2$  для отделения целевого класса от нецелевого, как могут измениться оценки качества классификации, если  $t_1 < t_2$ :

Чувствительность для  $t_2$  может уменьшиться по сравнению с  $t_1$

Специфичность для  $t_2$  может увеличиться по сравнению с  $t_1$

Специфичность для  $t_2$  может уменьшиться по сравнению с  $t_1$

ROC AUC (площадь под кривой Detection rate по вертикали, False positive rate по горизонтали) для  $t_2$  может увеличиться по сравнению с  $t_1$



- ROC AUC (площадь под кривой Detection rate по вертикали, False positive rate по горизонтали) для t2 может уменьшиться по сравнению с t1
- Чувствительность для t2 может увеличиться по сравнению с t1

**Балл: 0**

**Комментарий к правильному ответу:**

9. Пусть аналитическая оценка метода наименьших квадратов для линейной регрессии не определена. Выберите, что может сделать оценку определяемой:

- добавить регуляризацию в ф-цию потерь
- добавить новые признаки
- удалить часть признаков

**Балл: 0**

**Комментарий к правильному ответу:**

10. Логистическая регрессия с линейным уравнением зависимости логита от предикторов позволяет найти:

- нелинейную зависимость отклика (вероятности целевого класса) от предикторов
- нелинейную границу между классами в пространстве признаков
- линейную границу между классами в пространстве признаков
- линейную зависимость отклика (вероятности целевого класса) от предикторов

**Балл: 0**

**Комментарий к правильному ответу:**