Todo list

1 Разное

1. Гипотеза H_0 по-английски читается как «H naught»
2. При проверке гипотезы об адекватности регрессии НЕЛЬЗЯ писать $H_0 : R^2 = 0$.
   Гипотезы имеет смысл проверять о ненаблюдаемых неизвестных константах. Проверить гипотезу о том, что $R^2 = 0$ легко. Для этого не нужно знать ничего из теории вероятностей, достаточно просто сравнить посчитанное значение R^2 с нулём.
   Более того, даже корректировка $\mathbb{E}(R^2) = 0$ неверна. Случайная величина R^2 всегда неотрицательна, поэтому при любых разумных предпосылках на ε окажется, что $\mathbb{P}(R^2 > 0) > 0$. А это приведёт к тому, что $\mathbb{E}(R^2) > 0$ даже если Y никак не зависит от X.
   Единственный правильный вариант — $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$ и $H_a : \exists i \geqslant 2 : \beta_i \neq 0$.

2 Ridge/Lasso regression

LASSO — Least Absolute Shrinkage and Selection Operator. Метод построения регрессии, предложенный Robert Tibshirani в 1995 году.
Вспомним обычный МНК:

$$\min_{\beta} (y - X\beta)'(y - X\beta) \tag{1}$$

LASSO вместо исходной задачи решает задачу условного экстремума:

$$\min_{\beta} (y - X\beta)'(y - X\beta) \tag{2}$$

при ограничении $\sum_{j=1}^{k} |\beta_j| \leqslant c$.

Проверить! Нет ли у β_1 особого положения?

Естественно, при больших значениях c результат LASSO совпадает с МНК. Что происходит при малых c?
Для наглядности рассмотрим задачу с двумя коэффициентами β: β_1 и β_2. Линии уровня целевой функции — эллипсы. Допустимое множество имеет форму ромба с центром в начала координат.

на картинке три c: очень большое — дающиее мнк решение, меньше — ненулевые β, маленькое — одна из β равна 0

То есть при малых c LASSO обратит ровно в ноль некоторые коэффициенты β.
Применим метод множителей Лагранжа для случая, когда ограничение $\sum_{j=1}^{k} |\beta_j| \leqslant c$ активно, то есть выполнено как равенство.

$$L(\beta, \lambda) = (y - X\beta)'(y - X\beta) + \lambda \left(\sum_{j=1}^{k} |\beta_j| - c \right) \tag{3}$$

Необходимым условием первого порядка является $\partial L / \partial \beta = 0$. Это условие первого порядка не изменится, если мы зачеркнём c в выражении. Таким образом мы получили альтернативную формулировку метода LASSO:

$$\min_{\beta} (y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^{k} |\beta_j| \tag{4}$$

LASSO пытается минимизировать взвешенную сумму $RSS = (y - X\beta)'(y - X\beta)$ и «размера» коэффициентов $\sum_{j=1}^{k} |\beta_j|$.
Мы не будем вдаваться в численные алгоритмы, которые используются при решении этой задачи.
Ridge regression отличается от LASSO ограничением $\sum \beta_j^2 \leqslant c$. Также как и LASSO Ridge regression допускает альтернативную формулировку:

$$\min_{\beta} (y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^{k} \beta_j^2 \tag{5}$$

Также как и LASSO Ridge regression тоже приближает значения коэффициентов β_j к нулю. Принципиальное отличие LASSO и RR. В LASSO краевое решение с несколькими коэффициентами равными нулю является типичной ситуацией. В RR коэффициент β_j может оказаться точно равным нулю только по чистой случайности.
LASSO допускает байесовскую интерпретацию...
Предположим, что априорное распределение параметров следующее:
...
Тогда мода апостериорного распределения будут приходится в точности (?) на оценки LASSO.

Может ли появиться мультимодальность? В точности ли на моду или только примерно?