

Прорешиваем обучающую кр

29 декабря 2023 г.

1 Задача 2. Линейные регрессионные модели.

- Свести задачу к виду линейной регрессионной модели. Т.е. конвертировать условие задачи в формулу вида

$$X = Z\theta + \varepsilon;$$

В которой:

- X - это вектор результатов экспериментов/измерений/взвешиваний;
 - Z - это фиксированная матрица измерений (что и по сколько штук измеряли);
 - θ - это вектор неизвестных параметров - измеряемые величины/веса объектов;
 - ε - это вектор случайного шума, погрешностей;
- Привести задачу к такому виду, что:
 - $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T : \forall i \varepsilon_i \sim N(0, \sigma^2)$ [Обычно из условия задачи сразу следует несмещенность];
 - Стобцы матрицы Z - ЛНЗ [Обычно сразу есть из условия задачи];
 - Элементы вектора ε - попарно некоррелированы [Обычно есть попарная независимость из условия задачи, откуда сразу же есть и некоррелированность];

В таком случае наша модель будет называться **линейной гауссовской** моделью.

- Получив линейную гауссовскую модель можно воспользоваться двумя известными оценками, обладающими полезными свойствами:

$$\begin{aligned} - \hat{\theta} &= (Z^T Z)^{-1} Z^T X; \\ - \hat{\sigma}^2 &= \frac{1}{n-k} \|X - Z\hat{\theta}\|^2. \end{aligned}$$

n - размерность X , k - размерность θ ;

Полученные оценки будут удовлетворять всем требованиям в задаче, так как они являются оптимальными в линейной гауссовской модели.

2 Задача 4. Проверка линейных гипотез.

В гауссовской линейной модели можно рассматривать различные гипотезы вида:

$$H_0 : T\theta = \tau \text{ V.S. } H_1 : T\theta \neq \tau.$$

Для проверки таких гипотез есть хороший критерий - F-критерий.

- Собственно, начнем с того, чтобы свести условие к такому виду, а эксперимент - к виду гауссовской линейной модели. Обозначим $m = rk \ T$.
- Далее, научимся решать систему из условия $T\theta = \tau$. Если $k > m$, то система имеет бесконечно много решений, тогда их можно представить как линейную оболочку какого-то набора векторов. В задачах зачастую $k = m + 1$ и множество решений представляется как $v \cdot \beta$, где v - фиксированный вектор, а β - любое число.
- Теперь, решаем задачу оптимизации: $\min_{\theta: T\theta=\tau} \|X - Z\theta\|^2$. Если $\theta = c\beta$, то уравнение будет такого вида: $\min_{\beta} \|X - Zv\beta\|^2$, что превращается в поиск минимума параболы.
- **Альтернативная формула F-критерия.** Остается применить такую формулу для F-критерия:

$$R = \left\{ \frac{\min_{\theta: T\theta=\tau} \|X - Z\theta\|^2 - \|X - Z\hat{\theta}\|^2}{\|X - Z\hat{\theta}\|^2} \cdot \frac{n - k}{m} > f_{1-\alpha} \right\},$$

где f_p - p -квантиль распределения Фишера со степенями свободы $m, n - k$, т.е. $F_{m, n-k}$.