Семинар №10.
Линейная регрессионная модель.

Постановка задачи.

Пусть наблюдение X — случайный вектор из $\mathbb{R}^n$, причём

$$X = l + \varepsilon,$$

где l — фиксированный неизвестный вектор (который мы и хотим оценить), а ε — случайный вектор, $E\varepsilon = 0, D\varepsilon = \sigma^2 I_n$, I_n — единичная диагональная матрица размера $n \times n$.

Линейность модели: про l известно, что $l \in L$ — линейное подпространство в $\mathbb{R}^n$, $\dim L = k < n$.
Что неизвестно? l и σ^2.
Дано: L задано в виде базиса $Z_1, \ldots, Z_k$ (вектор-столбцы). Введём матрицу $Z = (Z_1, \ldots, Z_k)$ (просто столбцы рядом поставили). Тогда

$$l = Z_1\theta_1 + \ldots + Z_k\theta_k = Z\theta,$$

где $\theta = (\theta_1, \ldots, \theta_k)^T$ - неизвестные координаты l в базисе Z. Таким образом, задача оценки l сведена к задаче оценки $\theta \in \mathbb{R}^k$.

Метод наименьших квадратов.

$$\widehat{\theta} := \arg\min_{\theta} \|X - Z\theta\|^2,$$

т.е. мы ищем такое $Z\theta$ из линейного подпространства L, которое минимизирует евклидово расстояние до X. Нетрудно догадаться, что $Z\widehat{\theta} = proj_L X$ (проекция X на L).

Определение. $\widehat{\theta}$ называется оценкой θ по методу наименьших квадратов (далее – о.н.к.).

Утверждение 1. $\widehat{\theta} = (Z^TZ)^{-1}Z^TX$.

Доказательство.

$$\|X - Z\theta\|^2 = (X - Z\theta)^T(X - Z\theta) = X^TX - X^TZ\theta - \theta^TZ^TX + \theta^TZ^TZ\theta =$$

$$= X^TX - 2X^TZ\theta + \theta^T(Z^TZ)\theta.$$

Стоит заметить, что $\|X - Z\theta\|^2$ — это число, поэтому $X^TZ\theta$ и θ^TZ^TX — тоже числа. Но $\forall a \in \mathbb{R}$ $a = a^T$, поэтому $X^TZ\theta = (X^TZ\theta)^T = \theta^TZ^TX$. Дифференцируем по θ_i и приравниваем к нулю, чтобы найти минимум:

$$-2(X^TZ)_i + 2(\theta^TZ^TZ)_i = 0 \Rightarrow X^TZ - \theta^TZ^TZ = 0 \Rightarrow \widehat{\theta} = (Z^TZ)^{-1}Z^TX. \ \square$$

Что взять в качестве оценки σ^2?

Утверждение 2. $E\left(\frac{1}{n-k}\|X - Z\widehat{\theta}\|^2\right) = \sigma^2$.

Задача. Имеется 2 объекта с весами a и b. Мы взвесили (с ошибками) первый, второй и потом оба вместе на одних и тех же весах. Найти оценки наименьших квадратов для a и b.
Решение. Пусть взвешивания показали результат X_1, X_2 и X_3. $X = (X_1, X_2, X_3)^T$, тогда $X = l + \varepsilon$, где

$l = (a, b, a+b)^T$ и $\theta = (a, b)$. Далее, $l = (1,0,1)^T a + (0,1,1)^T b$, т.е. базис линейного подпространства L можно выбрать таким: $\{(1,0,1)^T, (0,1,1)^T\}$ и матрица

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

Тогда

$$\widehat{\theta} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = (Z^T Z)^{-1} Z^T X = \begin{pmatrix} \frac{2}{3}X_1 - \frac{1}{3}X_2 + \frac{1}{3}X_3 \\ -\frac{1}{3}X_1 + \frac{2}{3}X_2 + \frac{1}{3}X_3 \end{pmatrix} \square$$

Гауссовская линейная модель.

Это модель линейной регрессии $X = l + \varepsilon$, в которой $\varepsilon \sim N(0, \sigma^2 I_n)$.

Утверждение 3. *В гауссовской линейной модели* $(proj_L X, \|proj_{L^\perp}\|^2)$ *– достаточная статистика для* (l, σ^2).

Доказательство. Распишем плотность вектора наблюдений $X \sim N(l, \sigma^2 I_n)$.

$$p(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - l_i)^2\right) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(\frac{1}{2\sigma^2}\|X - l\|^2\right).$$

Но по теореме Пифагора (sic!) $\|X - l\|^2 = \|proj_L X - proj_L l\|^2 + \|proj_{L^\perp} X - proj_{L^\perp} l\|^2$. Вектор l лежит в линейном подпространстве L, поэтому $proj_L l = l$, а $proj_{L^\perp} l = 0$, следовательно, плотность вектора X представляется в виде

$$p(x) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\|proj_L X - l\|^2 + \|proj_{L^\perp} X\|^2}{2\sigma^2}\right).$$

отсюда по критерию факторизации получаем, что $(proj_L X, \|proj_{L^\perp} X\|^2)$ – достаточная статистика. $\square$
К тому же, она является полной по теореме об экспоненциальных семействах.

Следствие 1. $\widehat{\theta}$ *– оптимальная оценка для* θ, $\frac{1}{n-k}\|X - Z\widehat{\theta}\|^2$ *– оптимальная оценка для* σ^2.

Доказательство. Теорема Лемана-Шефаре говорит, что если у нас есть несмещённая оценка, которая является функцией от полной достаточной статистики, то эта оценка является оптимальной. Ранее мы устанавливали, что $Z\widehat{\theta} = proj_L X$, поэтому $X - Z\widehat{\theta} = X - proj_L X = proj_{L^\perp} X \Longrightarrow \frac{1}{n-k}\|X - Z\widehat{\theta}\|^2$ — оптимальная оценка для σ^2. Ну а $\widehat{\theta}$ — несмещённая оценка θ и, к тому же, $\widehat{\theta} = (Z^T Z)^{-1} Z^T \cdot Z\widehat{\theta}$, т.е. $\widehat{\theta}$ — тоже оптимальная. $\square$

Теорема 1. *(Об ортогональном разложении гауссовского вектора)*
Пусть $X \sim N(a, \sigma^2 I_n)$, $L_1 \oplus \ldots \oplus L_r$ *— разложение* $\mathbb{R}^n$ *в прямую сумму ортогональных подпространств. Положим* $Y_j = proj_{L_j} X$. *Тогда* $Y_1, \ldots, Y_r$ *— независимые в совокупности, причём* $\frac{1}{\sigma^2}\|Y_j - EY_j\|^2 \sim \chi^2_{\dim L_j}$ *(где* $\chi^2_{\dim L_j}$ *– хи-квадрат распределение с* $\dim L_j$ *степенями свободы).*