

Семинар 10.

Мультиколлинеарность. Метод главных компонент (МГК, PCA).

1. Метод главных компонент (Principal Component Analysis, PCA).

- (a) Метод главных компонент. Теория.
- (b) Разберите геометрическую интерпретацию на примере двух нормальных признаков $x_1 \sim N(a_1, \sigma_1^2)$ и $x_2 \sim N(a_2, \sigma_2^2)$.

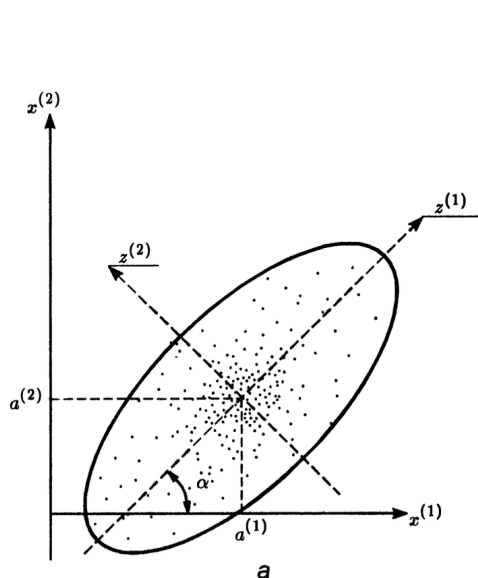


Рис. 1. Умеренный разброс точек.

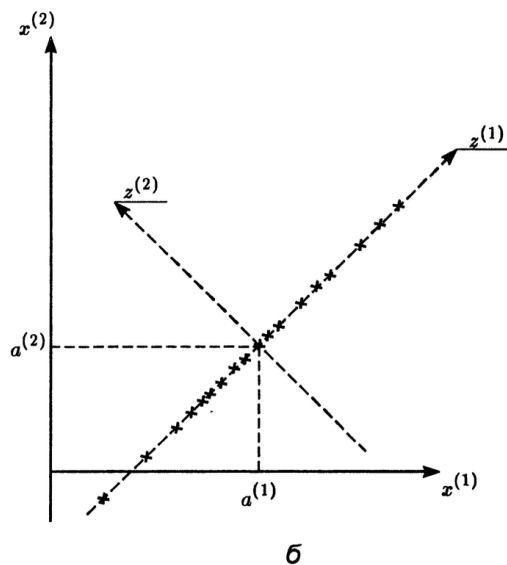


Рис. 2. Отсутствие разброса точек в направлении второй главной компоненты (вырожденный случай).

- (c) Определим матрицу факторной нагрузки как $A = C^T \Sigma_Z^{1/2}$, где Σ_Z — ковариационная матрица некоррелированных главных компонент. Покажите, что элемент (a_{ij}) матрицы A является коэффициентом корреляции между стандартизированной главной компонентой j на стандартизированном признаком x_i
2. Теоретическая регрессионная зависимость и выборочная корреляционная матрица стандартизированных регрессоров X имеют вид:

$$y_i = \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i,$$

$$\hat{R}(X) = \begin{pmatrix} 1 & 0.95 & 0 \\ 0.95 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- (a) Вычислите все главные компоненты. Сколько главных компонент надо выбрать, чтобы они объясняли не менее 70% общей дисперсии?

- (b) Вычислите матрицу факторной нагрузки. Проинтерпретируйте полученные результаты.