

1 Матричная алгебра

1. Пусть $\mathbf{C} \in \mathbb{R}^{p \times p}$ — симметричная неотрицательно-определенная матрица, $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ — ее собственные числа.

Тогда $\text{trace}(\mathbf{C}) = \sum_{i=1}^p \lambda_i$, $|\mathbf{C}| = \prod_{i=1}^p \lambda_i$, $|\mathbf{I} + \mathbf{C}| = \prod_{i=1}^p (1 + \lambda_i)$, $1/|\mathbf{I} + \mathbf{C}| = \prod_{i=1}^p 1/(1 + \lambda_i)$.

2. Пусть $\mathbf{A} \in \mathbb{R}^{p \times p}$ — симметричная неотрицательно-определенная матрица ранга r , Пусть $\mathbf{B} \in \mathbb{R}^{p \times p}$ — симметричная положительно-определенная матрица. Тогда матрица $\mathbf{B}^{-1}\mathbf{A}$ (несимметричная, вообще говоря) имеет p неотрицательных собственных чисел $\lambda_1 \geq \dots \geq \lambda_p \geq 0$. Число ненулевых собственных чисел равно рангу r матрицы \mathbf{A} . Соответствующие собственные вектора образуют базис, который можно выбрать так, чтобы $\mathbf{U}^T \mathbf{B} \mathbf{U} = \mathbf{I}$, где \mathbf{U} — матрица, составленная из собственных векторов (говорят, что собственные вектора ортонормированы относительно матрицы \mathbf{B}).

Так как $\mathbf{B}^{-1}\mathbf{A}\mathbf{U} = \lambda\mathbf{U}$ равносильно $\mathbf{A}\mathbf{U} = \lambda\mathbf{B}\mathbf{U}$, то эту задачу называют обобщенной задачей на собственные числа и собственные вектора (вместо единичной матрицы стоит симметричная положительно определенная матрица \mathbf{B}).

Для обобщенной задачи на собственные значения справедливы те же свойства оптимальности, что и для обычной, только с ортогональностью относительно матрицы \mathbf{B} . А именно, $\sup_Z Z^T \mathbf{A} Z / Z^T \mathbf{B} Z = \sup_{Z: Z^T \mathbf{B} Z = 1} Z^T \mathbf{A} Z$ равен максимальному собственному числу λ_1 и достигается на $Z = U_1$; $\sup_{Z: Z^T \mathbf{B} U_1 = 0} Z^T \mathbf{A} Z / Z^T \mathbf{B} Z$ равен λ_2 и достигается на $Z = U_2$. И т.д.

3. В условиях предыдущего пункта, $|\mathbf{B}|/|\mathbf{B} + \mathbf{A}| = 1/|\mathbf{I} + \mathbf{B}^{-1}\mathbf{A}| = \prod_{i=1}^r 1/(1 + \lambda_i)$.

Следствие. Пусть \mathbf{A} имеет распределение Уишарта $W_p(\mathbf{I}, \nu_A)$, \mathbf{B} имеет распределение Уишарта $W_p(\mathbf{I}, \nu_B)$, $\nu_B \geq p$. $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ — собственные числа $\mathbf{B}^{-1}\mathbf{A}$. Тогда $\prod_{i=1}^s 1/(1 + \lambda_i)$, где $s = \min(p, \nu_A)$, имеет распределение Лямбда Уилкса $\Lambda_p(\nu_A, \nu_B)$.

2 Множественная регрессия и one-way ANOVA

Ниже будем считать, что все случайные величины имеют нормальное распределение.

1. Multiple regression: $\eta, \zeta = (\zeta_1, \dots, \zeta_k)^T$ — количественные признаки. Пусть $\hat{\eta}$ — наилучшее линейное приближение по МНК. В случае нормальной модели, это $\hat{\eta} = \mathbb{E}(\eta|\zeta_1, \dots, \zeta_k)$.

Тогда получаем разложение дисперсии

$$\mathbb{E}(\eta - \mathbb{E}\eta)^2 = \mathbb{E}(\hat{\eta} - \mathbb{E}\eta)^2 + \mathbb{E}(\eta - \hat{\eta})^2.$$

Разложение суммы квадратов имеет вид

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y}_i)^2.$$

Получаем: $SS_{total} = SS_{regression} + SS_{error}$ и соотношение для степеней свободы $(n - 1) = (k) + (n - k - 1)$.

Гипотеза $H_0 : \mathbb{E}(\hat{\eta} - \mathbb{E}\eta)^2 = 0$

Значимость регрессии: $t = SS_{regression}/(k)/(SS_{error}/(n - k - 1)) \sim F(k, n - k - 1)$.

2. One-way ANOVA: η, ξ — одномерный качественный признак, принимающий значения A_1, \dots, A_k .

Тогда получаем разложение дисперсии

$$\mathbb{E}(\eta - \mathbb{E}\eta)^2 = \mathbb{E}(\mathbb{E}(\eta|\xi) - \mathbb{E}\eta)^2 + \mathbb{E}(\eta - \mathbb{E}(\eta|\xi))^2.$$

Разложение суммы квадратов имеет вид

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2.$$

Получаем: $SS_{total} = SS_{between} + SS_{within}$ и соотношение для степеней свободы $(n - 1) = (k - 1) + (n - k)$.

Гипотеза $H_0 : \mathbb{E}(\mathbb{E}(\eta|\xi) - \mathbb{E}\eta)^2 = 0$

Значимость ANOVA: $t = SS_{between}/(k - 1)/(SS_{within}/(n - k)) \sim F(k - 1, n - k)$.

3. Введем фиктивные (dummy) переменные $\zeta_1, \dots, \zeta_{k-1}$ на основе качественного признака ξ : $\zeta_i = 1$, если $\xi = A_i$; иначе $\zeta_i = 0$.

Тогда разложения дисперсии для ANOVA превращается в разложение дисперсии для линейной регрессии η на $\zeta_1, \dots, \zeta_{k-1}$.

Корреляционное отношение η к ξ равно множественному коэффициенту корреляции между η и $\zeta_1, \dots, \zeta_{k-1}$ (построение наилучшего функционального предсказания от ξ эквивалентно построению наилучшего линейного предсказания от $\zeta_1, \dots, \zeta_{k-1}$).

4. Введем единые обозначения, $SST = SSH + SSE$ (H — hypothesis, E — error), ν_H и ν_E соответствующее число степеней свободы.

Тогда критерий выглядит как $F = (SSH/\nu_H)/(SSE/\nu_E) \sim F(\nu_H, \nu_E)$. При этом $R^2 = SSH/SST$ — это коэффициент детерминации, или оценка множественного коэффициента корреляции, или оценка корреляционного отношения.

5. Обозначим $\lambda_1 = SSH/SSE$.

Тогда корреляция $r^2 = \lambda_1/(1 + \lambda_1)$. Чем больше r^2 , тем больше значимость ANOVA/регрессия.

$F = \lambda_1 \cdot (\nu_E/\nu_H) \sim F(\nu_H, \nu_E)$. Чем больше λ (и F), тем больше значимость ANOVA/регрессия.

$\Lambda = 1/(1 + \lambda_1) = 1 - r^2 \sim \Lambda_1(\nu_H, \nu_E)$. Чем **меньше** Λ , тем больше значимость ANOVA/регрессия.