

ЗАДАНИЕ ПО КУРСУ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ

Автор: Хоружий Кирилл

От: 23 октября 2021 г.

Содержание

1	A2	2
2	A3	3
3	A4	5

1 A2

№5

Рассмотрим многомерное нормальное распределение для \tilde{y}_i с симметричной матрицей ковариации Σ :

$$\tilde{y} = \frac{1}{(2\pi)^{l/2} \sqrt{\det \Sigma}} \exp \left(-\frac{1}{2} (\tilde{y} - y) \Sigma^{-1} (\tilde{y} - y) \right).$$

Нормировка. В силу симметричности Σ существует S такая, что $S^T \Sigma^{-1} S = E$, тогда $\det S = \sqrt{\det \Sigma}$. Тогда, в силу знания о линейной замене переменных в кратном интеграле, при замене $\tilde{y} - y = Sz$, верно:

$$\int \tilde{y} d^l \tilde{y} = \frac{\det S}{\sqrt{\det \Sigma}} \int \frac{1}{(\sqrt{2\pi})^l} \exp \left(-\frac{1}{2} z^T S^T \Sigma^{-1} S z \right) d^l \tilde{y},$$

что приводит к факторизации, и, по теореме Фубини, можем записать

$$\int \tilde{y} d^l \tilde{y} = \int \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} z_1^2 \right) dz_1 \cdot \dots \cdot \int \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} z_l^2 \right) dz_l = 1,$$

следовательно, указанное распределение нормировано.

Парные корреляторы. Вообще, говорят, что набор случайных величин ξ имеет многомерное нормальное распределение, если найдётся вектор a , невырожденная матрица C и набор *независимых* стандартных нормальных величин η такие, что

$$\xi = a + C\eta.$$

Так гораздо удобнее найти $\text{cov}(\xi_i, \xi_k)$:

$$\langle\langle \xi_i, \xi_j \rangle\rangle = \langle\langle (a + C\eta)_i, (a + C\eta)_j \rangle\rangle = \sum_{\alpha=1}^l \sum_{\beta=1}^l c_{i\alpha} c_{j\beta} \underbrace{\langle\langle \eta_\alpha, \eta_\beta \rangle\rangle}_{\delta_{\alpha\beta}} = \sum_{\alpha=1}^l c_{i\alpha} c_{j\alpha} = (CC^T)_{ij} = \Sigma_{ij},$$

где последнее равенство следует из факторизации распределения для η .

Погрешности параметров. Оценим погрешности параметров, аналогично расчёту с лекции:

$$\begin{aligned} w_\alpha = Q_{\alpha i} y_i & \Rightarrow \langle\langle \tilde{w}_\alpha, \tilde{w}_\beta \rangle\rangle = \dots = Q_{\alpha i} Q_{\beta j} \langle\langle \tilde{y}_i, \tilde{y}_j \rangle\rangle = Q_{\alpha i} Q_{\beta j} \Sigma_{ij} = (Q \Sigma Q^T)_{\alpha\beta}, \\ \tilde{w}_\alpha = Q_{\alpha i} \tilde{y}_i & \end{aligned}$$

что похоже на правду, по крайней мере формы совпадают.

Погрешности в линейной регрессии. Считая $A = \text{diag}(A_1, \dots, A_l)$, оценим погрешности $\text{var } w_\alpha$. Рассмотрим, видимо, линейную регрессию, тогда, как и раньше

$$X = \begin{pmatrix} x_1 & 1 \\ \dots & \dots \\ x_l & 1 \end{pmatrix}, \quad X^T X = l \begin{pmatrix} \bar{x}^2 & \bar{x} \\ \bar{x} & 1 \end{pmatrix}, \quad (X^T X)^{-1} = \frac{1}{l \text{var } x} \begin{pmatrix} 1 & -\bar{x} \\ -\bar{x} & \bar{x}^2 \end{pmatrix}.$$

Здесь, наверное, будет удобнее сразу найти

$$Q = (X^T X)^{-1} X^T = \frac{1}{l \text{var } x} \begin{pmatrix} x_1 - \bar{x} & \dots & x_l - \bar{x} \\ \bar{x}^2 - \bar{x}x_1 & \dots & \bar{x}^2 - \bar{x}x_l \end{pmatrix}.$$

Тогда искомые погрешности могут быть найдены, как

$$\begin{aligned} \text{var } w_1 &= (Q \Sigma Q^T)_{11} = \frac{1}{l(\text{var } x)^2} \left(\langle x_i^2 \sigma_i^2 \rangle - 2\bar{x} \langle x_i \sigma_i^2 \rangle + \bar{x}^2 \langle \sigma_i^2 \rangle \right), \\ \text{var } w_0 &= (Q \Sigma Q^T)_{22} = \frac{1}{l(\text{var } x)^2} \left((\bar{x}^2)^2 \langle \sigma_i^2 \rangle - 2\bar{x} \cdot \bar{x}^2 \langle x_i \sigma_i^2 \rangle + \bar{x}^2 \langle x_i^2 \sigma_i^2 \rangle \right). \end{aligned}$$

где $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_l)$, и $\text{var } \tilde{y}_i \sim \sigma_i^2$. Действительно, при $\sigma_i^2 = s^2 = \text{const}$ всё сходится.

2 А3

№1

Запишем теорему Байеса

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}.$$

По условиям, количество срабатываний счетчика Гейгера за минуту n подчиняется распределению Пуассона

$$P_\lambda(n) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

В ходе эксперимента счётчик сработал m раз за минуту. Будем считать априорную вероятность $p(\lambda)$ константной, тогда

$$P(\lambda|m) = p(\lambda) \frac{P(m|\lambda)}{\int_0^\infty p(\alpha) P(m|\alpha) d\alpha} = \frac{P(m|\lambda)}{\int_0^\infty \frac{\alpha^m}{m!} e^{-\alpha} d\alpha} = P(m|\lambda) = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Наша априорная вероятность обновилась, теперь априорной вероятностью считаем $p(\lambda) = P(\lambda|m)$. Тогда после второго эксперимента, опять же по теореме Байеса

$$P(\lambda|m') = \frac{\lambda^m}{m!} e^{-\lambda} \frac{\frac{\lambda^{m'}}{m'!} e^{-\lambda}}{\int_0^\infty \frac{\alpha^m}{m!} e^{-\alpha} \cdot \frac{\alpha^{m'}}{m'!} e^{-\alpha} d\alpha} = \frac{(2\lambda)^{m+m'}}{m'!m!} e^{-2\lambda} \cdot \frac{2m!m'!}{(m+m')!} = 2 \frac{(2\lambda)^{m+m'}}{(m+m')!} e^{-2\lambda},$$

с максимум $\lambda = \frac{1}{2} (m + m')$, что вполне логично.

№2

Снова запишем теорему Байеса, считая априорную вероятность быть заболевшим $p = 10^{-5}$:

$$P(\text{болен}|+) = \frac{P(+|\text{болен}) \cdot p}{P(+|\text{болен}) \cdot p + P(+|\text{здоров}) \cdot (1-p)} = \frac{0.99 \cdot 10^{-5}}{0.99 \cdot 10^{-5} + 0.01 \cdot (1 - 10^{-5})} = 10^{-3},$$

то есть получив положительный тест (+) на вирус, Петя действительно заболевший с вероятностью 0.1%.

№3

см. блокнот

№4

см. блокнот

№5

Покажем, что задача минимизации квадратичной функции потерь с дополнительным ограничением (лассо Тибширани):

$$\mathcal{L} = \|Xw - y\|^2 \rightarrow \min_w, \quad \sum_{\alpha} |w_{\alpha}| < C,$$

эквивалентна L_1 -регуляризации.

Действительно, введем

$$w_{\alpha}^{+} = \begin{cases} 0, & w_{\alpha} \leq 0, \\ w_{\alpha}, & w_{\alpha} \geq 0, \end{cases}, \quad w_{\alpha}^{-} = \begin{cases} w_{\alpha}, & w_{\alpha} \leq 0, \\ 0, & w_{\alpha} \geq 0, \end{cases}, \quad w_{\alpha} = w_{\alpha}^{+} + w_{\alpha}^{-}, \quad |w_{\alpha}| = w_{\alpha}^{+} - w_{\alpha}^{-}.$$

По условию Каруша-Куна-Таккера, задача о поиске экстремума на компакте, эквивалентна

$$\mathcal{L} = \text{RSS} + \lambda_i g_i \rightarrow \min, \quad \lambda_i \geq 0, \quad \lambda_i g_i = 0.$$

То есть задача действительно эквивалентна L_1 -регуляризации.

№6. Bias-Variance decomposition

Упростим задачу, уйдя от распределения, как введению некоторого шума ε_i , тогда

$$\tilde{y}_i = y(x_i) + \varepsilon_i,$$

где $\varepsilon_i \in \mathcal{N}(0, \sigma^2)$, соответственно $\mathbb{E} \varepsilon = 0$. Соответственно есть некоторая оценка \hat{y} . Также учтём, что ε_i и $\hat{y}_i(x_i)$ независимы. В общем,

$$\begin{aligned} \mathcal{L} &= \mathbb{E}[(\tilde{y} - \hat{y})^2] = \mathbb{E}[(y(x) + \varepsilon - \hat{y})^2] = \mathbb{E}[(y(x) + \varepsilon - \hat{h} + \mathbb{E} \hat{y} - \mathbb{E} \hat{y})^2] = \\ &= \mathbb{E}[(y - \mathbb{E} \hat{y})^2] + \mathbb{E} \varepsilon^2 + \mathbb{E}[(\mathbb{E} \hat{y} - \hat{y})^2] + 2 \underbrace{\mathbb{E}[(y - \mathbb{E} \hat{y})\varepsilon]}_{\mathbb{E}[(y - \mathbb{E} \hat{y})] \cdot \mathbb{E} \varepsilon = 0} + 2 \underbrace{\mathbb{E}[\varepsilon(\mathbb{E} \hat{y} - \hat{y})]}_{\mathbb{E}[(\hat{y} - \mathbb{E} \hat{y})] \cdot \mathbb{E} \varepsilon = 0} + 2 \underbrace{\mathbb{E}[(\mathbb{E} \hat{y} - \hat{y})(y - \mathbb{E} \hat{y})]}_{(\mathbb{E} \hat{y} - \mathbb{E} \hat{y}) \cdot \mathbb{E}[(y - \mathbb{E} \hat{y})] = 0} \\ &= \underbrace{\mathbb{E} \varepsilon^2}_{\text{noize}} + \underbrace{\mathbb{E}[(y - \mathbb{E} \hat{y})^2]}_{\text{bias}} + \underbrace{\mathbb{E}[(\mathbb{E} \hat{y} - \hat{y})^2]}_{\text{variance}}, \end{aligned}$$

что и требовалось доказать: \mathcal{L} факторизируется на шум, смещение и разброс.

3 A4

№2

Знаем, что верно разложение $X = V\sqrt{\Lambda}U^T$, также $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{\tilde{F}}, \dots, \lambda_F)$, и вводим $\tilde{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_{\tilde{F}}, 0, \dots, 0)$. Тогда можем найти норму $\tilde{X} = V\sqrt{\tilde{\Lambda}}U^T$:

$$\|\tilde{X} - X\|^2 = \|V \underbrace{(\sqrt{\tilde{\Lambda}} - \sqrt{\Lambda})}_{\Delta} U^T\|^2 = \text{tr}(U \Delta^T V^T V \Delta U^T) = \text{tr}(U^T U \Delta^T \Delta) = \text{tr}(\Delta^T \Delta) = \sum_{k=\tilde{F}+1}^F \lambda_i,$$

что и требовалось доказать.

№3

Покажем, что сингулярный вектор матрицы X , отвечающий наибольшему сингелярному числу, является решением задачи

$$\mathbf{u} = \text{argmax}_{\|\mathbf{u}\|=1} (X\mathbf{u})^2.$$

Само собой считаем $\lambda_1 > \lambda_2 > \dots > \lambda_F$, также знаем, что

$$\|X\mathbf{u}\|^2 = \langle X\mathbf{u}|X\mathbf{u} \rangle = \langle \mathbf{u}|X^T X|\mathbf{u} \rangle, \quad \|X^T X\| = \lambda_1,$$

а норма где-то достигается, и достигается как раз на первом сингулярном векторе.

Действительно, по определению $XU = V\sqrt{\Lambda}$, где $U = (\mathbf{u}_1, \dots, \mathbf{u}_F)$ – собственные вектора $X^T X$, и $X\mathbf{u}_1 = \sqrt{\lambda_1}v_1$.

№4

Сведем задачу к предыдущей (№3). Для этого запишем дисперсию вдоль вектора \mathbf{a} :

$$S^2 = \frac{1}{l} \sum_{i=1}^l \langle \mathbf{a} | \mathbf{x}_i \rangle = \frac{1}{l} \sum_{i=1}^l \sum_{j=1}^F (X_{ij} a_j)^2 \rightarrow \max,$$

где $X^T = (\mathbf{x}_1, \dots, \mathbf{x}_l)$. По теореме Пифагора верно, что

$$\|\mathbf{x}_i - \mathbf{a} \langle \mathbf{a} | \mathbf{x}_i \rangle\| = \|\mathbf{x}_i\| - \langle \mathbf{a} | \mathbf{x}_i \rangle^2,$$

тогда, суммируя по i , находим

$$\underbrace{\sum_i \|\mathbf{x}_i - \mathbf{a} \langle \mathbf{a} | \mathbf{x}_i \rangle\|}_{\rightarrow \min} = \underbrace{\sum_i \|\mathbf{x}_i\|}_{\equiv \text{const}} - \underbrace{\sum_i \langle \mathbf{a} | \mathbf{x}_i \rangle^2}_{\rightarrow \max}.$$

Осталось заметить, что

$$\sum_i \langle \mathbf{a} | \mathbf{x}_i \rangle^2 = (X\mathbf{a})^T (X\mathbf{a}) = \langle \mathbf{a} | X^T X | \mathbf{a} \rangle,$$

которая достигает максимума на \mathbf{a} – первом сингулярном векторе (см. №3), а значит поставленные задачи равносильны.

№5

Посмотрим на X :

$$X = \begin{pmatrix} x_1 & y_1 & z_1 \\ & \dots & \\ x_N & y_N & z_N \end{pmatrix}, \quad X^T X = \begin{pmatrix} x_i x_i & x_i y_i & x_i z_i \\ x_i y_i & y_i y_i & y_i z_i \\ x_i z_i & y_i z_i & z_i z_i \end{pmatrix},$$

где подразумевается суммирование при повторяющемся индексе.

Вспомним, что тензор инерции выглядит, как

$$I = \begin{pmatrix} x_i x_i + z_i z_i & -x_i y_i & -x_i z_i \\ -x_i z_i & x_i x_i + z_i z_i & -y_i z_i \\ -x_i z_i & -y_i z_i & x_i x_i + y_i y_i \end{pmatrix},$$

так что осталос только заметить, что I и $X^T X$ диагонализуются одновременно, а значит задачи равносильны.

P.S. задачи №1 и №6 см. в блокноте.