Задание по курсу вычислительной математики

Автор: Хоружий Кирилл

От: *4 декабря 2021 г.*

Содержание

1 A2

№5

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

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

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

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

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

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

следовательное, указанное распределение нормировано.

Парные корреляторы. Вообще, говорят, что набор случайных величин ξ имеет многомерноное 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 $\operatorname{cov}(\xi_i,\, \xi_k)$:

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

где последнее равенство следует из факторизации распределения для η.

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

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

что похоже на правду, по крайней мере формы совпадают.

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

$$X = \begin{pmatrix} x_1 & 1 \\ \ldots & \ldots \\ x_l & 1 \end{pmatrix}, \qquad X^{\mathrm{T}}X = l\begin{pmatrix} \bar{x^2} & \bar{x} \\ \bar{x} & 1 \end{pmatrix}, \qquad (X^{\mathrm{T}}X)^{-1} = \frac{1}{l\operatorname{var} x}\begin{pmatrix} 1 & -\bar{x} \\ -\bar{x} & \bar{x^2} \end{pmatrix}..$$

Здесь, наверное, будет удобнее сразу найти

$$Q = \left(X^{\mathrm{T}}X\right)^{-1} X^{\mathrm{T}} = \frac{1}{l\operatorname{var} x}\begin{pmatrix} x_1 - \bar{x} & \ldots & x_l - \bar{x} \\ \overline{x^2} - \bar{x}x_1 & \ldots & \overline{x^2} - \bar{x}x_l \end{pmatrix}.$$

Тогда искомые погрешности могут быть найдены, как

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

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

2 A3

№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):

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

эквивалентна L_1-регуляризации.

Действительно, введем

$$w_\alpha^+ = \begin{cases} 0, & w_\alpha \leqslant 0,\\ w_\alpha, & w_\alpha \geqslant 0,\end{cases}, \qquad w_\alpha^- = \begin{cases} w_\alpha, & w_\alpha \leqslant 0,\\ 0, & w_\alpha \geqslant 0,\end{cases} \qquad w_\alpha = w_\alpha^+ + w_\alpha^-, \qquad |w_\alpha| = w_\alpha^+ - w_\alpha^-.$$

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

$$\mathcal{L} = \mathrm{RSS} + \lambda_i g_i \to \min, \qquad \lambda_i \geqslant 0, \qquad \lambda_i g_i = 0.$$

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

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

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

3 А4

№2

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

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

что и требовалось доказать.

№3

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

$$\boldsymbol{u} = \mathrm{argmax}_{\|u\|=1}\left(X\boldsymbol{u}\right)^2.$$

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

$$\|X\boldsymbol{u}\|^2 = \langle X\boldsymbol{u}|X\boldsymbol{u}\rangle = \langle\boldsymbol{u}|X^{\mathrm{T}}X|\boldsymbol{u}\rangle, \qquad \|X^{\mathrm{T}}X\| = \lambda_1,$$

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

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

№4

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

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

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

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

тогда, суммируя по i, находим

$$\underbrace{\sum_i\|\boldsymbol{x}_i - \boldsymbol{a}\langle\boldsymbol{a}|\boldsymbol{x}_i\rangle\|}_{\to\min} = \underbrace{\sum_i\|\boldsymbol{x}_i\|}_{\equiv\mathrm{const}} - \underbrace{\sum_i\langle\boldsymbol{a}|\boldsymbol{x}_i\rangle^2}_{\to\max}.$$

Осталось заметить, что

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

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

№5

Посмотрим на X:

$$X = \begin{pmatrix} x_1 & y_1 & z_1 \\ & \ldots & \\ x_N & y_N & z_N \end{pmatrix}, \qquad X^{\mathrm{T}}X = \begin{pmatrix} x_ix_i & x_iy_i & x_iz_i \\ x_iy_i & y_iy_i & y_iz_i \\ x_iz_i & y_iz_i & z_iz_i \end{pmatrix},$$

где подразумевается суммирование при повторяющемся индексе.

Вспомним, что тензор инерции выглядит, как

$$I = \begin{pmatrix} x_ix_i + z_iz_i & -x_iy_i & -x_iz_i \\ -x_iz_i & x_ix_i + z_iz_i & -y_iz_i \\ -x_iz_i & -y_iz_i & x_ix_i + y_iy_i \end{pmatrix},$$

так что осталос только заметить, что I и $X^{\mathrm{T}}X$ диагонализируются одновременно, а значит задачи равносильны.

P.S. задачи №1 и №6 см. в блокноте.

4 B4

Thr 4.1. *Если* $\forall i$ $|a_{ii}| \geqslant \sum_{i,j} |a_{ij}| + \varepsilon$, *то метод Зейделя сходится.*

$\triangle$. Введем погрешность решения $\delta^{(k+1)} = x^{(k)} - x^{(*)}$, для которой верно, что $^{(}k \hat{+} 1) = R\delta^{(k)}$, а значит

$$\delta_i^{(k+1)} = \frac{1}{a_{ii}} \left(\sum_{j=1}^{i-1} a_{ij} \delta_j^{(k+1)} + \sum_{j=i+1}^{n} a_{ij} \delta_j^k \right),$$

откуда получаем оценку

$$|\delta_i^{(k+1)}| \leqslant \max_{1 \leqslant j < i} |\delta_j^{k+1}| \alpha_i + \max_{i < j \leqslant n} |\delta_j^{(k)}| \beta_i, \quad \Rightarrow \quad \|\delta^{(k+1)}\| \leqslant \alpha_i \|\delta^{(k+1)}\| + \beta_i \|\delta^{(k)}\|, \quad \alpha_i + \beta_i < 1,$$

в силу диагонального преобладания. Выражая норму, находим

$$\|\delta^{(k+1)}\| \leqslant \frac{\beta_i}{1 - \alpha_i} \|\delta^{(k)}\| = q_{\text{seid}} \|\delta^{(k)}\|, \quad \Rightarrow \quad q_{\text{seid}} = \frac{\beta_i}{1 - \alpha_i}.$$

$\square$

Осталось вспомнить, что $q_{\text{gauss}} = \alpha_i + \beta_i$, и сравнить $\beta_i/(1 - \alpha_i)$ и $\alpha_i + \beta_i$, для которых верно

$$\frac{\beta_i}{1 - \alpha_i} < \alpha_i + \beta_i, \quad \Leftrightarrow \quad \alpha_1(\alpha_i + \beta_i) > \alpha_i, \quad \Rightarrow \quad q_{\text{seid}} < q_{\text{gauss}},$$

а значит метод Зейделя сходится быстрее.