№1

Нужно минимизировать функцию потерь

$$L = \sum_{i=1}^{n} (y - \check{y})^2,$$

моделируя данные постоянной величиной $\check{y} = c$. Значит, $0 = \dfrac{dL}{d\check{y}} = -2\sum_{i=1}^{n}(y - c)$, то есть

$$c = \frac{1}{n}\sum_{i=1}^{n} y = \bar{y}.$$

Это и есть решение задачи.

№3

Нужно решить задачу на условный экстремум с набором функций связи $Xw = y$:

$$\begin{cases} L = w^T w + \lambda^T (Xw - y) \\ Xw = y \end{cases}$$

По необходимому условию:

$$\frac{\partial L}{\partial w} = 2w^T + \lambda^T X = 0 \to w = -\frac{1}{2}X^T\lambda$$

Гессиан функции Лагранжа оказался пропорционален единичной матрице:

$$\frac{\partial^2 L}{\partial w^i \partial w^j} = 2\mathbb{I}$$

Значит, будет минимум, как и положено.

$$XX^T\lambda = -2y \to \lambda = -2(XX^T)^{-1}y.$$

Наконец,

$$w = X^T(XX^T)^{-1}y.$$

И правда, получилось, что искомые коэффициенты выражаются через правую псевдообратную матрицу к X.

№5

Заметим, что распределение нормировано неправильно. Но мы сделаем правильно заменив A на A^{-1}:

$$\int_{\mathbb{R}^n} \frac{dy}{\sqrt{\det A(2\pi)^n}} \exp\left(-\frac{(\tilde{y} - y)A^{-1}(\tilde{y} - y)^T}{2}\right) = \int_{\mathbb{R}^n} \frac{dx}{\sqrt{\det A(2\pi)^n}} \exp\left(-\frac{x^T\Lambda^{-1}x}{2}\right) = \\ = \prod_{k=1}^{n} \int_{\mathbb{R}} \frac{dx}{\sqrt{(2\pi)^n \det A}} \exp -\frac{x_k^2\lambda_k}{2} = 1$$

Далее

$$\langle\langle \tilde{y}_i, \tilde{y}_j \rangle\rangle = C\int_{\mathbb{R}^n} \exp\left(-\frac{(\tilde{y} - y)^T A^{-1}(\tilde{y} - y)}{2}\right)(\tilde{y}_i - y_i)(\tilde{y}_j - y_j)dy$$

Воспользуемся примечанием после замены $Y = \tilde{y} - y$:

$$
\begin{aligned}
C\int dY \exp\left(-\frac{Y^TA^{-1}Y}{2}+J^TY\right) = \int dY \exp\left(-\frac{Y^TA^{-1}Y}{2}+J^TY\right) = \\
= C\det(S^{-1})\int dY \exp\left(-\frac{Z^TZ-2K^TS^TSZ+(S^TSK)^T(S^TSK)}{2}+\frac{(S^TSK)^TS^TSK}{2}\right) = \\
= C\det(S^{-1})\int dZ \exp\left(-\frac{(Z-S^TSK)^T(Z-S^TSK)}{2}\right)\exp\left(\frac{(S^TK)^TS^TK}{2}\right) = \\
= \exp\left(\frac{J^TS^TSJ}{2}\right) = \exp\left(\frac{J^TAJ}{2}\right)
\end{aligned}
$$

Учитывая явное выражение для интеграла и дифференцируя по компонентам i, j вектора J в нуле:

$$\langle\langle \tilde{y}_i, \tilde{y}_j \rangle\rangle = A_{ij}$$

Пишем формулы: $w_a = ((X^TX)^{-1}X)_{ai}y_i = Q_{ai}y$, $\tilde{w}_a = Q_{ai}\tilde{y}_i$ Т.к. $\langle\langle w_a, w_b\rangle\rangle = \langle (w_a-\tilde{w}_a)(w_b-\tilde{w}_b)\rangle = Q_{ai}Q_{bj}A_{ij} = (QAQ^T)_{ab}$. В частности, $\langle\langle w_a, w_a\rangle\rangle = var(w_a) = (QAQ^T)_{aa}$ – дисперсия для коэффициентов. Далее в частном случае $A = diag(a_1, \ldots, a_n)$, $y = w_1x + w_0$. Тогда $X = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}$ и $X^TX = l\begin{pmatrix} \bar{x^2} & \bar{x} \\ \bar{x} & 1 \end{pmatrix}$, $Q = (X^TX)^{-1}X = \frac{1}{lvar(x)}\begin{pmatrix} x_1-\bar{x} & \ldots & x_l-\bar{x} \\ \overline{x^2}-\bar{x}x_1 & \ldots & \overline{x^2}-\bar{x}x_l \end{pmatrix}$

$$var(w_0) = (QAQ^T)_{00} = \frac{1}{l(var(x))^2}\left(\overline{x_i^2a_i^2} - 2\overline{x}\overline{x_ia_i^2} + \overline{x}^2\overline{a_i^2}\right)$$

$$var(w_1) = (QAQ^T)_{11} = \frac{1}{l(var(x))^2}\left(\overline{x^2}^2\overline{a_i}^2 - 2\bar{x}\bar{x}^2 + \overline{x_ia_i^2}\right)$$

Видно, что просто напросто следует выбрать $a_i = s_i^2$.