№1

а) Выборка y_i Модель: $\tilde{y} = const$

$$L = \sum_{i=1}^{\ell}(y_i - \tilde{y})^2 \to \min_{\tilde{y}}$$

$$\frac{\partial L}{\partial \tilde{y}} = -2\sum_{i=1}^{\ell}(y_i - \tilde{y}) = 0 \Rightarrow \sum_{i=1}^{\ell}(y_i - \tilde{y}) = 0; \quad \sum_{i=1}^{\ell} y_i = \ell\tilde{y} \Rightarrow$$

$$\tilde{y} = \frac{1}{\ell}\sum_{i=1}^{\ell} y_i = \langle y_i \rangle$$

б) $\tilde{y} = kx + B$

$$L = \sum_{i=1}^{\ell}(y_i - kx_i - B)^2$$

$$\begin{cases} \frac{\partial L}{\partial k} = -2\sum_{i=1}^{\ell}(y_i - kx_i - B)x_i = 0 \\ \frac{\partial L}{\partial B} = -2\sum_{i=1}^{\ell}(y_i - kx_i - B) = 0 \end{cases} \qquad \begin{cases} \sum_{i=1}^{\ell}(x_iy_i - kx_i^2 - Bx_i) = 0 & \Big| \cdot \frac{1}{\ell} \\ \sum_{i=1}^{\ell}(y_i - kx_i - B) = 0 & \Big| \cdot \frac{1}{\ell} \end{cases}$$

$$(*) \begin{cases} \langle xy \rangle - k\langle x^2 \rangle - B\langle x \rangle = 0 \\ \langle y \rangle - k\langle x \rangle - B = 0 & \Big| \cdot \langle x \rangle \end{cases} \qquad \begin{cases} \langle xy \rangle - k\langle x^2 \rangle - B\langle x \rangle = 0 \\ \langle y \rangle\langle x \rangle - k\langle x \rangle^2 - B\langle x \rangle = 0 \end{cases}$$

$$\langle xy \rangle - \langle x \rangle\langle y \rangle - k(\langle x^2 \rangle - \langle x \rangle^2) = 0$$

$$k = \frac{\langle xy \rangle - \langle x \rangle\langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2}$$

$$\langle y \rangle - k\langle x \rangle = B \Rightarrow B = \langle y \rangle - \frac{\langle xy \rangle - \langle x \rangle\langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2}\langle x \rangle$$

Из (*) видно, что $\langle y \rangle = k\langle x \rangle + B \Rightarrow$ аппроксимирующая прямая проходит через точку $(\langle x \rangle, \langle y \rangle)$

Для многомерного случая:

$\tilde{y} = Xw \qquad \tilde{y}_i = \sum_\alpha x_{i\alpha} w_\alpha \qquad w$ - столбец весов признаков

$$L = \sum_{i=1}^{\ell}(\tilde{y}_i - y_i)^2 \to \min$$

$$L = \sum_{i=1}^{\ell}\Big(\sum_\alpha x_{i\alpha}w_\alpha - y_i\Big)^2 \qquad \frac{\partial L}{\partial w_{\alpha_1}} = 2\sum_{i=1}^{\ell}\Big(\sum_\alpha x_{i\alpha}w_\alpha - y_i\Big)x_{i\alpha_1} = 0$$

Для столбца свободных членов $x_{i\alpha_k} = 1 \ \forall i \Rightarrow \frac{\partial L}{\partial w_{\alpha_k}} = 2\sum_{i=1}^{\ell}\Big(\sum_\alpha x_{i\alpha}w_\alpha - y_i\Big) = 0$

$\langle \sum x_{i\alpha} w_\alpha \rangle = \langle y \rangle \Rightarrow \langle y \rangle = w_1\langle x_1 \rangle + w_2\langle x_2 \rangle + \ldots + w_{k-1}\langle x_{k-1} \rangle + w_k$

Проходит через „центр масс" выборки. $\langle x_i \rangle$ - среднее i-ого столбца мат. X

№2

$$y = x_1 w_1 + x_2 w_2 + \ldots + x_n w_n + w_0 = Xw + w_0$$

$$\langle y \rangle = \langle x_1 \rangle w_1 + \langle x_2 \rangle w_2 + \ldots + \langle x_n \rangle w_n + w_0 = \bar{X} w + w_0$$

$$\underbrace{y - \langle y \rangle}_{\tilde{y}} = \underbrace{(X - \bar{X})}_{\tilde{X}} w$$ провели централизацию признаков, или, грубо говоря, сместили координаты так, что гиперплоскость проходит через нуль.

Рассмотрим ту же задачу, только введём новый признак $f_0(x) = 1 = x_0$

$$y = x_1 w_1 + x_2 w_2 + \ldots + x_n w_n + x_0 w_0$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} & 1 \\ x_{21} & x_{22} & \ldots & x_{2n} & 1 \\ \ldots & \ldots & \ldots & \ldots & \vdots \\ x_{m1} & x_{m2} & \ldots & x_{mn} & 1 \end{pmatrix} \begin{pmatrix} \vec{w} \\ w_0 \end{pmatrix} = \hat{X} \hat{w}$$

Теперь имеем дело со смещённой в нуль задачей

Тогда $w = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y$ $\quad$ $\tilde{X} \in Mat_{m \times n}(\mathbb{R})$, $\tilde{X}^T \in Mat_{n \times m}(\mathbb{R})$ $\Rightarrow (\tilde{X}^T \tilde{X})^{-1} \in Mat_{n \times n}(\mathbb{R})$

$w' = (\hat{X}^T \hat{X})^{-1} \hat{X}^T y$ $\quad$ $\hat{X} \in Mat_{m \times (n+1)}(\mathbb{R})$, $\hat{X}^T \in Mat_{(n+1) \times n}(\mathbb{R})$ $\Rightarrow (\hat{X}^T \hat{X})^{-1} \in Mat_{(n+1) \times (n+1)}(\mathbb{R})$

Если вычеркнуть у $\hat{X}$ последний столбец, а у $(\hat{X}^T \hat{X})^{-1}$ последние столбец и строку, то получим w' без последней строки, т.е. w

Таким образом, мы получили два эквивалентных подхода по поиску весов, не стоящих перед свободным членом: либо смещаем начало координат, либо добавляем новый признак (то же самое, что добавить новую координатную ось и по МНК искать гиперплоскость на 1 порядок выше исходной задачи)

№3

$y = Xw \quad X \in Mat_{m \times n}(\mathbb{R}) \quad m < n, \quad rg X = m$

Будем искать такую псевдообратную матрицу Q, что

$L = \| QX - E_n \|^2 \to \min\limits_{Q \in Mat_{n \times m}(\mathbb{R})}$, где $\|A\|^2 = \sum\limits_{i,j} |a_{ij}|^2$

$\frac{\partial L}{\partial Q} = 2(QX - E_n)X^T \Rightarrow QXX^T = X^T$

$\Rightarrow \boxed{Q = X^T(XX^T)^{-1}}$ — правая псевдообратная матрица

№4

а) $X = V\sqrt{\Lambda}U^T$ $\quad \Lambda = \mathrm{diag}\{\lambda_1 \geq \dots \geq \lambda_F\}$ $\quad$ Здесь предполагается норма Фробениуса

$\tilde{\Lambda} = \mathrm{diag}\{\lambda_1 \geq \dots \geq \lambda_{\tilde{F}}\}$

$\tilde{X} = V\sqrt{\tilde{\Lambda}}U^T \neq X$ $\qquad \|X - \tilde{X}\|^2 = \|V\sqrt{\Lambda}U^T - V\sqrt{\tilde{\Lambda}}U^T\|^2 =$

$= tr(V\sqrt{\Lambda}U^T - V\sqrt{\tilde{\Lambda}}U^T)^T(V\sqrt{\Lambda}U^T - V\sqrt{\tilde{\Lambda}}U^T) =$

$= tr(U\sqrt{\Lambda}V^T - U\sqrt{\tilde{\Lambda}}V^T)(V\sqrt{\Lambda}U^T - V\sqrt{\tilde{\Lambda}}U^T) =$

$= tr(U\sqrt{\Lambda}V^TV\sqrt{\Lambda}U^T - U\sqrt{\Lambda}V^TV\sqrt{\tilde{\Lambda}}U^T - U\sqrt{\tilde{\Lambda}}V^TV\sqrt{\Lambda}U^T + U\sqrt{\tilde{\Lambda}}V^TV\sqrt{\tilde{\Lambda}}U^T) =$

$= tr(U\Lambda U^T - 2U\sqrt{\Lambda}\sqrt{\tilde{\Lambda}}U^T + U\tilde{\Lambda}U^T) = tr(U\Lambda U^T - 2U\tilde{\Lambda}U^T + U\tilde{\Lambda}U^T) =$

$= \underset{\underset{tr(\Lambda - \tilde{\Lambda})}{\parallel}}{tr(U\Lambda U^T - U\tilde{\Lambda}U^T)} = \sum\limits_{i=\tilde{F}+1}^{F} \lambda_i$ ч.т.д.

б)
$$X = V\sqrt{\Lambda}\,U^T$$

$$X u_k = \sqrt{\lambda_k}\, v_k\, u_k^T u_k = \sqrt{\lambda_k}\, v_k$$

$$(X u_k)^2 = (X u_k)^T X u_k = v_k^T \sqrt{\lambda_k}\sqrt{\lambda_k}\, v_k = v_k^T \lambda_k v_k$$

$$u_k = \arg\max v_k^T \lambda_k v_k = \arg\max v_k^T \sigma_k^2 v_k = \arg\max \sigma_k^2$$

$\Rightarrow$ u_k отвечает наибольшему сингулярному числу.