Q_1:

To prove Gauss-Markov Theorem, we need to prove unbiasedness and linearity

First of all, $E(\hat{\beta}) = E((X^TX)^{-1}X^Ty) = (X^TX)^{-1}X^TE(y)$

$= (X^TX)^{-1}X^TE(\beta X + \varepsilon)$

$E(\varepsilon) = 0 \Rightarrow = (X^TX)^{-1}X^TE(X\beta)$

$= (X^TX)^{-1}X^TX\beta$

since $(X^TX)^{-1} \cdot (X^TX) = 1$

$\therefore (X^TX)^{-1}X^TXB = 1 \cdot \beta = \beta$

$\therefore E(\hat{\beta}) = \beta$

In linear regression $y = XB + \varepsilon$

suppose $(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n)$

we have $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \cdots \beta_k x_{ik} + \varepsilon_i$

from the form of linear model, we can see that OLS estimate is a linear function of the given data. The dependent variable y of each data can be expressed as a linear combination between $\beta_0, \beta_1, \beta_2 \cdots \beta_k$ and corresponding variable x_i

$\therefore$ linearity proved

$$\begin{cases} a^T\hat{\beta} = a^T\beta + \varepsilon_1 \\ c^Ty = a^T\beta + \varepsilon_2 \end{cases} \Rightarrow \begin{cases} Var(a^T\beta + \varepsilon_1) = Var(a^T\hat{\beta}) \\ Var(a^T\beta + \varepsilon_2) = Var(c^Ty) \end{cases}$$

$$\Rightarrow \begin{cases} Var(a^T\hat{\beta}) = Var(\varepsilon_1) \\ Var(c^Ty) = Var(\varepsilon_2) \end{cases}$$

Since c^Ty is an unbiased estimator for $a^T\beta$, $E(c^Ty) = a^T\beta$

$\therefore E(\varepsilon_2) = 0$

$Var(\varepsilon_1) \leq Var(\varepsilon_1 + \varepsilon_2)$

$\Rightarrow Var(a^T\hat{\beta}) \leq Var(c^Ty)$

$\therefore$ the least squares estimate in linear regression is BLUE.

Q_2:

Since columns $x_0, \dots x_p$ of X are orthogonal, they are pairwise orthogonal to each other. The dot product of different columns are '0'.

linear regression: $y = \beta_0 x_0 + \beta_1 x_1 \dots + \beta_p x_p + \varepsilon$

$$\therefore \hat{\beta}_j = \frac{x_j' \cdot y}{x_j' x_j} = \frac{x_j' \cdot y}{\|x_j\|^2} \quad (j \text{ from } 1 \text{ to } p)$$

Q_3: from the question: we assume the SVD of X is $U\Sigma V^T$.
$U \in \mathbb{R}_r^{n\times(p+1)}$ satisfies $U^T U = I_r$, $V \in \mathbb{R}^{(p+1)\times r}$ satisfies $V^T V = I_r$ and $\Sigma = diag(\sigma_1, \dots \sigma_r)$ is the diagonal matrix of positive singular values.

(a)

SVD of X: $U\Sigma V^T$

SVD of X^T: $(U\Sigma V^T)^T$

$$\Rightarrow (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T$$
$$= V\Sigma^T \Sigma V^T \cdot I_r$$
$$= V\Sigma^2 \cdot V^T$$

consider $X^T X\beta = X^T y$

$$V\Sigma^2 V^T \beta = U\Sigma V^T y$$
$$V^T V\Sigma^2 V^T \beta = V^T U\Sigma V^T y$$

$I_r \cdot \Sigma^2 V^T \beta = V^T U \Sigma V^T y$

$\Sigma^{-2} \cdot \Sigma^2 V^T \beta = \Sigma^{-2} V^T U \Sigma V^T y$

$I_r \cdot V^T \beta = \Sigma^{-2} V^T U \Sigma V^T y$

$V \cdot V^T \beta = V \Sigma^{-2} V^T U \Sigma V^T y$

$\Rightarrow \beta = V \Sigma^{-2} V^T U \Sigma V^T y$

$\beta = V \Sigma^{-1} \cdot \Sigma^{-1} V^T U \cdot \Sigma V^T \cdot y = V \cdot \Sigma^{-1} \cdot (\Sigma^{-1} \cdot V^T \cdot U \Sigma V^T y)$

since $\Sigma = \text{diag}(\sigma_1 \dots \sigma_n) \Rightarrow \Sigma^{-1} = \Sigma^T$

since $V^T V = I_r \Rightarrow V^T = V^{-1}$

$\therefore \Sigma^{-1} V^{-1}$ is the inverse of $U \Sigma V^T$

$\therefore \beta = V \cdot \Sigma^{-1} \cdot (U^T y)$ which $\Sigma^{-1} \cdot V^T \cdot U \Sigma V^T y = U^T y)$

$\therefore \beta_{mns} = V \Sigma^{-1} U^T y$

(b) hint $\beta = \beta_{mns} + b$

$X^T X \beta = X^T y \Rightarrow X^T X (\beta_{mns} + b) = X^T y$

$X^T X \beta_{mns} + X^T X b = X^T y$

since β_{mns} is a solution to the normal equations:

$X^T X \beta_{mns} = X^T y$

$\therefore X^T y + X^T X b = X^T y$

$\therefore X^T X b = 0$

X^TX is not invertible, b lies in the null space of X^TX and can be expressed as a linear combination of the columns of V corresponding to 0 singular values.

since $V^TV = I_r$, columns of V are orthogonal

$\therefore$ b can be expressed as the form of $b = b_1 + b_2 \cdots b_n$, where b_i is a component in the null space of X^TX

as for $\|\beta\| = \|\beta_{mns} + b\|$

$\therefore \|\beta\| \leq \|\beta_{mns}\| + \|b\|$

$\|b\| = \|b_1 + b_2 \cdots b_n\| \leq \|b_1\| + \|b_2\| \cdots \|b_n\|$

However, $X^TXb = 0$ and $\|b_i\|$ is positive

$\therefore \|\beta\| \geq \|\beta_{mns}\|$

(c) the pseudo-inverse of X has properties of:

$$\begin{cases} X^{\dagger}X = I \\ XX^{\dagger}X = X \\ (X^{\dagger}X)^T = X^{\dagger}X \\ (XX^{\dagger})^T = XX^{\dagger} \end{cases}$$

When it comes to $V\Sigma^{-1}U^T$

$V\Sigma^{-1}U^T X = V\Sigma^{-1}U^T \cdot U\Sigma V^T = I$

$X(V\Sigma^{-1}U^T)X = U\Sigma V^T$

$(V\Sigma^{-1}U^T X)^T = V\Sigma^{-1}U^T U\Sigma V^T$

$(X V\Sigma^{-1}U^T)^T = U\Sigma V^T \cdot V\Sigma^{-1}U^T$

$\therefore$ Yes. $V\Sigma^{-1}X^T$ is the pseudo-inverse of X.