More questions and answers:

Consider the classical linear regression model $\underset{(n\times 1)}{Y} = \underset{(n\times k)}{X}\underset{(k\times 1)}{\beta} + \underset{(n\times 1)}{u}$. Prove that the least-squares $\underset{(k\times 1)}{b}$ vector is $b = (X'X)^{-1}X'Y$.

If the unknown vector $\beta$ is replaced by some guess or estimate $b$, this defines a vector of residuals $e$,

$$\underset{(n\times 1)}{e} = \underset{(n\times 1)}{Y} - \underset{(n\times k)}{X}\underset{(k\times 1)}{b},$$

The least-squares principle is to choose $b$ to minimize the residual sum of squares, $\sum_{i=1}^{n} e_i^2 = e'e$; $e' = [e_1 \; e_2 \; \cdots e_n]$:

$$\begin{aligned} e'e &= (Y - Xb)'(Y - Xb) \\ &= (Y' - b'X')(Y - Xb) \\ &= Y'Y - b'X'Y - Y'Xb + b'X'Xb. \end{aligned}$$

Next, note that $\underset{(1\times n)}{Y'}\underset{(n\times k)}{X}\underset{(k\times 1)}{b}$ is a scalar and hence its transpose $(Y'Xb)' = b'X'Y$ is the same scalar $(Y'Xb)$. Thus, $b'X'Y = Y'Xb$.

From the above analysis it follows that

$$e'e = Y'Y - 2b'X'Y + b'X'Xb.$$

The first order conditions are

$$\frac{\partial(e'e)}{\partial b} = -2X'Y + 2X'Xb = 0,$$

giving the least-squares $b$ vector as a function of the data:

$$\begin{aligned} X'Xb &= X'Y, \text{ or} \\ b &= (X'X)^{-1}X'Y. \end{aligned}$$

(b) Let $x_1, x_2, \ldots, x_n$ be a random sample from the Bernoulli distribution: $f(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, $0 \le \theta \le 1$. Derive the maximum likelihood estimator of $\theta$.

The likelihood function is given by

$$\begin{aligned} L(\theta; x) &= \prod_{i=1}^{n} f(x_i) = \\ &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_i^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}. \end{aligned}$$

Therefore, the log-likelihood function is

$$l = \ln L = \sum_i^n x_i \ln\theta + (n - \sum_{i=1}^n x_i)\ln(1-\theta).$$

The first order condition is

$$\frac{\partial l}{\partial \theta} = \frac{\sum_i^n x_i}{\theta} - \frac{(n - \sum_{i=1}^n x_i)}{(1-\theta)} = 0,$$

which gives the MLE

$$\widehat{\theta} = \frac{\sum_i^n x_i}{n} = \overline{x}.$$