1. (Scatter Plot Basic) Below is a scatter diagram for certain data set. Fill in the blanks.



2. As done in the first picture below, sketch a line that fits the data given in the following graphs:



3. (Finding the best line fit from Scatter Plot) Consider the following data:

x	у
2	6
4	8.5
1	2.5
7	15
5	11

- (a) Make a scatter plot of (x, y) in the graph below:
- (b) It is believed that y = ax + b is the true relationship. Below will help us find the best a, b.
 - i. Using the scatter plot, the predicted value of y for x_i is defined as $\hat{y}_i = ax_i + b$. Write down $\hat{y}_i = ax_i + b$ for i = 1, 2, 3, 4, 5.
 - ii. The prediction errors are defined by $d_i = y_i \hat{y}_i$. Write down d_i for i = 1, 2, 3, 4, 5
 - iii. In the above scatter plot draw a line for some a > 0 and b > 0 and mark the d_i on the graph for the line that you drew.
 - iv. Find a suitable notion of error and try to minimise it to the values of a and b.

4. Let $\{w_i : 1 \le i \le n\}$ be a set of real numbers $f(x) = \sum_{i=1}^n (w_i - x)^2$. Find the global minimum of f.

5. Let $\{x_i, y_i : 1 \le i \le n\}$ be a set of real numbers $f(a, b) = \sum_{i=1}^n (ax_i + b - y_i)^2$. Find the critical points of f and see if you can identify global minimum.

6. Let $\{(x_i, y_i) : 1 \le i \le n\}$ be a set of points on the plane. Let $a, b \in \mathbb{R}, \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ and $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$

$$\sum_{i=1}^{n} (ax_i + b - y_i)^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 + a^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 - 2a \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) + n(\bar{y} - a\bar{x} + b)^2.$$

- (a) Identify b that minimizes $\sum_{i=1}^{n} (ax_i + b y_i)^2$
- (b) Proceed to find a that minimizes $\sum_{i=1}^{n} (ax_i + b y_i)^2$.