Consider regression in one dimension, with a data set $(x_i, y_i)_{i=1,\ldots,m}$.

Problem 1

To find the weight and bias, we will partially differentiate the cost function given w.r.t weight and bias.

$$\begin{aligned}
c.f &= \sum_{i=1}^{m} (\hat{w}x_i + b - y_i)^2 \\
\frac{\partial(c.f)}{\partial \hat{w}} &= 2\sum_{i=1}^{m} (\hat{w}x_i + b - y_i)x_i \\
\frac{\partial(c.f)}{\partial \hat{b}} &= 2\sum_{i=1}^{m} (\hat{w}x_i + b - y_i)
\end{aligned}$$

Equating the above equations to zero,

$$\begin{aligned}
2\sum_{i=1}^{m} (\hat{w}x_i + b - y_i)x_i &= 0 \\
\hat{w}\sum_{i=1}^{m} x_i^2 + \hat{b}\sum_{i=1}^{m} x_i - \sum_{i=1}^{m} y_i x_i &= 0 \\
\hat{w}\,\mathbb{E}[x^2] + \hat{b}\,\mathbb{E}[x] &= \mathbb{E}[xy]
\end{aligned}$$

$$\begin{aligned}
2\sum_{i=1}^{m} (\hat{w}x_i + b - y_i) &= 0 \\
\hat{w}\sum_{i=1}^{m} x_i + \hat{b}\frac{m}{m} &= \sum_{i=1}^{m} y_i \\
\hat{b} &= \mathbb{E}[y] + \hat{w}\,\mathbb{E}[x]
\end{aligned}$$

Putting the value of $\hat{b}$ in $\hat{w}$, we get

$$\begin{aligned}
\hat{w} &= \frac{\mathbb{E}[xy] - \mathbb{E}[x]\,\mathbb{E}[y]}{var(x)} \\
\hat{w} &= \frac{cov(xy)}{var(x)}
\end{aligned} \tag{1}$$

$$\hat{b} = \mathbb{E}[y] - \frac{cov(xy)}{var(x)}\,\mathbb{E}[x] \tag{2}$$

$\hat{w}$ can also be called as the slope of the equation, since slope also has the same equation.

Problem 2

The true linear model is represented by $y_i = wx_i + b + \epsilon_i$, where ϵ_i $N(0, \sigma^2)$. Since, y_i corresponds to some random variable (let us assume Y_i, we can think of $\hat{w} and \hat{b}$ to be random variables too, denoted by the following equations:

$$\hat{w} = \frac{\sum_{i=1}^{m} (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$
$$\hat{b} = \bar{Y} - \hat{w}x$$

Let $t_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$ Let us first see if $\hat{w}$ is unbiased or not.

$$\begin{aligned}
\sum_{i}^{m} t_i &= \sum_{i}^{m} \frac{x_i - \bar{x}}{\sum_{i=1}^{m} (x_i - \bar{x})^2} \\
&= \frac{\sum_{i}^{m} (x_i - \bar{x})}{\sum_{i=1}^{m} (x_i - \bar{x})^2} \\
&= 0 \\
\sum_{i}^{m} t_i (x_i - \bar{x}) &= \sum_{i}^{m} \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{m} (x_i - \bar{x})^2} \\
&= 1
\end{aligned}$$

Using the above equations,

$$\begin{aligned}
\mathbb{E}[\hat{w}] &= \mathbb{E}[\sum_{i}^{m} t_i Y_i] \\
&= \sum_{i}^{m} t_i \, \mathbb{E}[Y_i] \\
&= \sum_{i}^{m} t_i (wx_i + b + \mathbb{E}[\epsilon_i)] \\
&= w \sum_{i}^{m} t_i x_i + b \sum_{i}^{m} t_i \\
&= w
\end{aligned}$$

This shows that $\hat{w}$ is indeed unbiased. Let us look into $\hat{b}$ now,

$$\begin{aligned}
\mathbb{E}[\hat{b}] &= \mathbb{E}[\bar{Y}] - \mathbb{E}[\hat{w}\bar{x}] \\
&= b
\end{aligned}$$

This is also unbiased.

Problem 3

In this problem, we have to calculate the variance of $\hat{w}$ and $\hat{b}$. Let us first see for weight:

$$\begin{aligned}\hat{w} &= \sum_{i}^{m} t_i(wx_i + b + \epsilon_i)\\ &= constant + \sum_{i}^{m} t_i\epsilon_i\end{aligned}$$

$$\begin{aligned}var(\hat{w}) &= \sum_{i}^{m} t_i^2 var(\epsilon_i)\\ &= \sigma^2 \sum_{i}^{m} t_i^2\\ &= \sigma^2 \sum_{i}^{m} \left(\frac{x_i - \bar{x}}{\sum_{i=1}^{m}(x_i - \bar{x})^2}\right)^2\\ &= \frac{\sigma^2}{m^2 * var(x)^2} \sum_{i=1}^{m} (x_i - \bar{x})^2\\ &= \frac{\sigma^2}{m^{\cancel{2}} * var(x)^2} \cancel{m} * var(x)\\ &= \frac{\sigma^2}{m * var(x)}\end{aligned}$$

Now for bias,

$$\begin{aligned}\hat{b} &= \frac{1}{m}\sum_{i}^{m} Y_i - \hat{x}\sum_{i}^{m} w_i Y_i\\ &= \sum_{i=1}^{m}\left(\frac{1}{m} - \hat{x}w\right)Y_i\\ &= constant + \sum_{i}^{m}\left(\frac{1}{m} - \hat{x}w\right)\epsilon_i\end{aligned}$$

$$\begin{aligned}var(\hat{b}) &= \sum_{i}^{m}\left(\frac{1}{m} - \hat{x}w\right)^2 var(\epsilon_i)\\ &= \sigma^2\sum_{i}^{m}\left(\frac{1}{m} - \hat{x}w\right)^2\\ &= \sigma^2\left(\frac{1}{m^2} - \frac{2}{m}\hat{x}\sum_{i}^{m} w_i + \hat{x}^2\sum_{i}^{m} w_i^2\right)\\ &= \sigma^2\frac{E[x^2]}{m * var(x)}\end{aligned}$$

Problem 4

If we shift/recenter the data i.e. $x_i' = x_i - \mu$, the error produced by the data on $\hat{w}$ is the same, but for $\hat{b}$ the error is minimized. This can be shown as below:

New values w:

$$\begin{aligned}
var(\hat{w'}) &= \frac{\sigma^2}{m * var(x')} \\
&= \frac{\sigma^2}{\sum_{i=1}^{m}(x_i' - \bar{x}')^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^{m}(x_i - \mu - (\bar{x} - \mu))^2} \\
&= \frac{\sigma^2}{m * var(x)}
\end{aligned}$$

New value for b (using the knowledge of $\mu = E[x] = \bar{x}$):

$$\begin{aligned}
var(\hat{b'}) &= \sigma^2 \frac{E[(x-\mu)^2]}{m * var(x-\mu)} \\
&= \frac{\sigma^2}{m^2 * var(x)} \sum_{i=1}^{m}(x_i - \mu)^2 \\
&= \frac{\sigma^2}{m^2 * var(x)} \sum_{i=1}^{m}(x_i - E[x])^2 \\
&= \frac{\sigma^2}{m * var(x)} \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m} \\
&= \frac{\sigma^2 * var(x)}{m * var(x)} \\
&= \frac{\sigma^2}{m}
\end{aligned}$$

Now let us check if $\hat{b}' < \hat{b}$:

$$\begin{aligned}
\frac{\sigma^2}{m} &< \frac{\sigma^2 E[x^2]}{m * var(x)} \\
\frac{\sigma^2}{m} - \frac{\sigma^2 E[x^2]}{m * var(x)} &< 0 \\
\frac{\sigma^2}{m}(1 - \frac{E[x^2]}{m * var(x)} &< 0 \\
\frac{\sigma^2}{m}(\frac{\cancel{E[x^2]} - E[x]^2 - \cancel{E[x^2]}}{var(x)}) &< 0 \\
-\frac{\sigma^2}{m}(\frac{E[x]^2}{var(x)}) &< 0
\end{aligned}$$

Since the estimator is unbiased, the only component that will contribute to the error is variance. And as seen above, when the data points are shifted about the mean value, the error is reduced.

Problem 5

Solution attached as a Jupyter notebook pdf.

Problem 6

When the data is shifted by a constant value (μ), all the points in the data just change their positions individually. As a group of points, they maintain the same orientation as earlier. In other ways, one can say that only the scaling of those data points change. Hence, the slope remains the same. An example to support the answer can be: Suppose you have a two data points at (1, 1) and (2, 2). The slope of the line passing through these points will be 1. Now, if I subtract a constant value from these data points and convert them to (0, 0) and (1, 1) respectively, we see that the slope is still 1.

Problem 7

According to the question, $\underline{x}$ is defined as:

$$\underline{x} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_m \end{bmatrix}$$

Since, $\Sigma = X^T X$, we can expand Σ in the following way:

$$\begin{aligned}
\Sigma &= \begin{bmatrix} 1 & 1 & . & . & . & 1 \\ x_1 & x_2 & . & . & . & x_m \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_m \end{bmatrix} \\
&= \begin{bmatrix} 1 + 1 + 1 .. + 1 & x_1 + x_2 + ... + x_m \\ x_1 + x_2 + ... + x_m & x_1^2 + x_2^2 + ... + x_m^2 \end{bmatrix} \\
&= \begin{bmatrix} m & m * E[x] \\ m * E[x] & m * E[x^2] \end{bmatrix} \\
&= m * \begin{bmatrix} 1 & E[x] \\ E[x] & E[x^2] \end{bmatrix}
\end{aligned}$$

Now, to find the eigenvalues for this 2x2 matrix, we need to equate its characteristic equation to zero (can remove m since equating to zero).

$$\begin{aligned}
|\Sigma - \lambda.I| &= 0 \\
\begin{bmatrix} 1 & E[x] \\ E[x] & E[x^2] \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} &= 0 \\
\begin{bmatrix} 1-\lambda & E[x] \\ E[x] & E[x^2]-\lambda \end{bmatrix} &= 0 \\
\lambda^2 - \lambda(1+E[x^2]) + Var(x) &= 0
\end{aligned}$$

Now, we can find the two values for λ, and divide the larger value by smaller value to get condition number ($\kappa(\Sigma)$).

$$\begin{aligned}
\lambda_1 &= \frac{(1+E[x^2]) + \sqrt{(1+E[x^2])^2 - 4 * var(x)}}{2} \\
&= \frac{(1+E[x^2])}{2}(1 + \sqrt{1 - \frac{4 * var(x)}{(1+E[x^2])^2}}) \\
\lambda_2 &= \frac{(1+E[x^2])}{2}(1 - \sqrt{1 - \frac{4 * var(x)}{(1+E[x^2])^2}}) \\
\kappa(\Sigma) &= \frac{\lambda_1}{\lambda_2} \\
&= \frac{\cancel{\frac{(1+E[x^2])}{2}}(1 + \sqrt{1 - \frac{4*var(x)}{(1+E[x^2])^2}})}{\cancel{\frac{(1+E[x^2])}{2}}(1 - \sqrt{1 - \frac{4*var(x)}{(1+E[x^2])^2}})} \\
&= \frac{1 + \sqrt{1 - \frac{4*var(x)}{(1+E[x^2])^2}}}{1 - \sqrt{1 - \frac{4*var(x)}{(1+E[x^2])^2}}}
\end{aligned}$$

P.T.O.

Now, if the data is centered about the mean, new $\kappa(\Sigma)$ will be:

$$\begin{aligned}
\kappa(\Sigma') &= \frac{1 + \sqrt{1 - \frac{4*var(x')}{(1+E[x'^2])^2}}}{1 - \sqrt{1 - \frac{4*var(x')}{(1+E[x'^2])^2}}} \\
&= \frac{1 + \sqrt{1 - \frac{4*var(x)}{(1+var(x))^2}}}{1 - \sqrt{1 - \frac{4*var(x')}{(1+var(x))^2}}} \\
&= \frac{1 + \sqrt{\frac{(1-var(x))^2}{(1+var(x))^2}}}{1 - \sqrt{\frac{(1-var(x))^2}{(1+var(x))^2}}} \\
&= \frac{1 + \frac{(1-var(x))}{(1+var(x))}}{1 - \frac{(1-var(x))}{(1+var(x))}} \\
&= \frac{1}{var(x)}
\end{aligned}$$