1. A)

let $X = \left[x^{(1)}, x^{(2)}, \dots, x^{(m)}\right]^T \rightarrow X\theta = \left[\theta^T x^{(1)}, \theta^T x^{(2)}, \dots, \theta^T x^{(m)}\right]^T$

let $W = \frac{1}{2}\mathrm{diag}(w_i) \rightarrow W(X\theta - y) = \frac{1}{2}\left[w_1(\theta^T x^{(1)} - y_1), \dots, w_m(\theta^T x^{(m)} - y_m)\right]^T$

$(X\theta - y)^T W (X\theta - y) = \frac{1}{2}\left(w_1\left(\theta^T x^{(1)} - y_1\right)^2 + \dots + w_m\left(\theta^T x^{(m)} - y_m\right)^2\right) = J(\theta)$

1. B)

$J(\theta) = \theta^T X^T W X\theta - 2\theta^T X^T W y + y^T W y$

$\nabla_\theta J(\theta) = 2X^T W X\theta - 2X^T W y = 0 \rightarrow \theta_{optimal} = (X^T W X)^{-1} X^T W y$

1. C)

$$P(y|X;\theta) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{\left(y_i - \theta^T x^{(i)}\right)^2}{2\sigma_i^2}\right\}$$

$$\log\left(P(y|X;\theta)\right) = \sum_{i=1}^{m} \log\left(\frac{1}{\sqrt{2\pi}\sigma_i}\right) - \sum_{i=1}^{m} \frac{\left(y_i - \theta^T x^{(i)}\right)^2}{2\sigma_i^2}$$

$$= f(\sigma) - (X\theta - y)^T W (X\theta - y) \text{ where } W = \frac{1}{2}\mathrm{diag}\left(\frac{1}{\sigma_i^2}\right)$$

in ML estimation, $\underset{\theta}{\mathrm{argmax}}\, P(y|X;\theta)$ is found. In regression, $\underset{\theta}{\mathrm{argmin}}\, J(\theta)$ is found.

As shown above, $\underset{\theta}{\mathrm{argmax}}\, P(y|X;\theta)$ can be found by taking the logarithm of $P(y|X;\theta)$.

The term in $\log\left(P(y|X;\theta)\right)$ which is dependent on θ has the form of weighted regression.

the weights are like this: $w_i = \frac{1}{\sigma_i^2}$

2)

The claim is that w^* is a linear combination of the feature vectors.
more precisely, because $X = \left[x^{(1)}, x^{(2)}, \dots, x^{(m)}\right]^T$, it is claimed that w^* is in the column space of X^T.
being in the column space of X^T is equivalent to being in the row space of X.
We know the row space and null space of a linear map are orthogonal
$\rightarrow$ it suffices to show that w^* is not in the null space of X

lemma 1: $X^T X$ is symmetric $\rightarrow (X^T X)^T = X^T X$

lemma 2: $X^T X$ is nonsingular, therefor it has an inverse.
proof 2: let y be in the null space of $X^T X \rightarrow X^T X z = 0, z^T X^T X z = \left||Xz|\right|_2^2 = 0 \rightarrow Xz = 0$
with the assumption that the features are independent, we know $\mathrm{Null}(X) = \{0\} \rightarrow z = 0$
then the null space of $X^T X = \{0\}$, therefor it is nonsingular and has an inverse.

lemma 3: The inverse of a symmetric matrix is also symmetric.
proof: let A be a nonsingular symmetric matrix. we know $A^{-1}A = AA^{-1} = I$
$I^T = I \rightarrow (AA^{-1})^T = A^{-1}A \rightarrow (A^{-1})^T A^T = A^{-1}A \rightarrow (A^{-1})^T A = A^{-1}A \rightarrow (A^{-1})^T AA^{-1} = A^{-1}AA^{-1}$
$\rightarrow (A^{-1})^T = A^{-1} \rightarrow A^{-1}$ is symmetric

putting the 3 lemmas together, we find that $(X^T X)^{-1}$ exists and is also symmetric.

we know that w^* in classical regression when l_2 loss is minimized is equal to $(X^T X)^{-1} X^T y$.
because $(X^T X)^{-1}$ is symmetric, it has eigendecomposition VDV^T.

therefor we have: $w^* = VDV^T X^T y$
let's assume that w^* is in the null space of X
$\rightarrow XVDV^T X^T y = 0 \rightarrow y^T VDV^T X^T y = 0 \rightarrow \left||V^T X^T y|\right|_2^2 = 0 \rightarrow V^T X^T y = 0$
if $V^T X^T y = 0$, then $w^* = 0$ and all α_i would be equal to 0
if $V^T X^T y \neq 0$, then we have reached a contradiction $\rightarrow w^* \notin \mathrm{Null}(X) \rightarrow w^* \in \mathrm{Column}(X^T)$
$\rightarrow X\alpha = w^*$ is a consistent system of equations $\rightarrow$ at least one α is found.

3.1)

We want to minimize the loss function $\left||\Phi\theta - y|\right|_2^2$
$\Phi = \left[\phi(x^{(1)}), \phi(x^{(2)}), \dots, \phi(x^{(m)})\right]^T$

$\nabla_\theta \left||\Phi\theta - y|\right|_2^2 = 2\Phi^T\Phi\theta - 2\Phi^T y$

we want to prove that for every t, θ_t is a linear combination of $\{\phi(x^{(i)})\}$, in other words, it should be in the column space ofΦ^T

Induction:
Base case: $\theta_0 = 0 \rightarrow \beta_0 = 0$
Inductive step: $\theta_{t+1} = \theta_t - \eta\, \nabla_\theta L(\theta_t) = (I - 2\eta\Phi^T\Phi)\theta_t + 2\eta\Phi^T y$
$\rightarrow$ θ_t is in the column space of Φ^T
$\rightarrow$ $2\eta\Phi^T y$ is also equal to $\sum 2\eta y_i\, \phi(x^{(i)}) \rightarrow$ is in the column space of Φ^T

$\rightarrow$ Proof is complete

3.2)

$\beta_0 = 0$
$\theta_t = \Phi^T\beta_t$
$\theta_{t+1} = \theta_t - \eta\, \nabla_\theta L(\theta_t) = \Phi^T\beta_t - 2\eta\Phi^T\Phi\Phi^T\beta_t + 2\eta\Phi^T y = \Phi^T(\beta_t - 2\eta\Phi\Phi^T\beta_t + 2\eta y)$
$\Phi\Phi^T_{i,j} = \langle\phi(x^{(i)}), \phi(x^{(j)})\rangle = \phi_i^T\phi_j$

$\theta_{t+1} = \Phi^T\beta_{t+1} \rightarrow$ $\beta_{t+1} = \beta_t - 2\eta\Phi\Phi^T\beta_t + 2\eta y$

3.2) already in vector form!

4)

Jacobi Method:
$Ax = b$
$A = D + R$ where D is a diagonal matrix
$x^{(0)}$ = initial guess
$x^{(k+1)} = D^{-1}\left(b - Rx^{(k)}\right)$

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j\neq i} a_{ij}x_j^{(k)}\right)$$

Use the python code below to solve the problem:

```
import numpy as np

A = np.array([[4, -1, 0],
              [-1, 4, -1],
              [0, -1, 4]], dtype=float)
b = np.array([3, 2, 3], dtype=float)

x = np.zeros_like(b)

max_iterations = 5

for iteration in range(max_iterations):
    print(x)
    x_new = np.zeros_like(x)
    for i in range(A.shape[0]):
        s = sum(A[i, j] * x[j] for j in range(A.shape[1]) if j != i)
        x_new[i] = (b[i] - s) / A[i, i]

    x = x_new

print(f'Solution after {iteration + 1} iterations: {x}')
```

Result:

```
[0. 0. 0.]
[0.75 0.5  0.75]
[0.875 0.875 0.875]
[0.96875 0.9375  0.96875]
[0.984375 0.984375 0.984375]
Solution after 5 iterations: [0.99609375 0.9921875  0.99609375]
```

5)
Gauss – Siedel Method:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j>i} a_{ij}x_j^{(k+1)} - \sum_{i>j} a_{ij}x_j^{(k)}\right)$$

the equations:

$$Ax = b \rightarrow \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

Use the python code below to solve the problem:

```python
import numpy as np

def gauss_seidel(A, b, x0, tolerance=1e-5, max_iterations=1000):
    n = len(b)
    x = x0.copy()

    for iteration in range(max_iterations):
        x_new = x.copy()
        for i in range(n):
            sum1 = sum(A[i][j] * x_new[j] for j in range(i))
            sum2 = sum(A[i][j] * x[j] for j in range(i + 1, n))
            x_new[i] = (b[i] - sum1 - sum2) / A[i][i]

        if np.allclose(x, x_new, atol=tolerance):
            break

        x = x_new

    return x, iteration + 1

t1 = 5
t2 = 8
t3 = 12

A = np.array([[1, t1, t1**2],
              [1, t2, t2**2],
              [1, t3, t3**2]], dtype=float)
b = np.array([8.106, 2.177, 2.279], dtype=float)

x0 = np.array([1, 2, 5], dtype=float)

solution, iterations = gauss_seidel(A, b, x0)

print(f'Solution after {iterations} iterations: {solution}')
```

Result:

```
Solution after 99 iterations: [29.42920311 -5.69463872  0.28601015]
```

دقت کنید که در زمان حل این سوال، متوجه تغییر آن نشده بودم و با توجه به عدم اطلاع رسانی به موقع، به هنگام تحویل متوجه شدم که مقادیر جداول درون سوال عوض شده‌است و با توجه به فشردگی آزمون‌ها، فرصتی برای درست کردن آن نداشتم. ممنون از درک شما.