Q2.2. Apply trace properties of matrix, we know:

① If $A \in \mathbb{R}^{n\times n}$, $trA = \sum_{i=1}^{n} A_{ii}$.

② $trA \cdot B = trB \cdot A$.

③ If $a \in \mathbb{R}$, $tra = a$.

④ $\nabla_A trABA^TC = CAB + C^TAB^T$

⑤ $trABC = trCAB = trBCA$.

⑥ If $f(A) = trAB$, $\nabla_A trAB = B^T$

⑦ $trA = trA^T$

Because $J(\theta)$ is a scalar, according to ③

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

$$= \frac{1}{2}\nabla_\theta tr (X\theta - y)^T (WX\theta - Wy)$$

$$= \frac{1}{2}\nabla_\theta tr (WX\theta - Wy)(X\theta - y)^T \quad ②$$

$$= \frac{1}{2}\nabla_\theta tr(WX\theta \cdot \theta^T X^T - WX\theta y^T - Wy\theta^T X^T + Wyy^T)$$

$$= \frac{1}{2}(\nabla_\theta tr\theta^T X^T WX\theta - \nabla_\theta tr \theta y^T WX - \nabla_\theta tr \theta^T X^T Wy) \quad ⑤$$

$$= \frac{1}{2}((\theta^T X^T WX + \theta^T X^T WX)^T - X^T W^T y - (y^T W^T X)^T) \quad ④, ⑥, ⑦$$

Because $W = W^T$

$$= \frac{1}{2}(2X^T WX\theta - 2X^T Wy) \stackrel{Set}{=} 0$$

$$\Rightarrow X^T WX\theta = X^T Wy \quad \text{(normal equation)}$$

$$\Rightarrow \theta = (X^T WX)^{-1} X^T Wy$$