$\therefore\ z^T z = \sum_i z_i^2$

$$\frac{1}{2}(X\theta - \vec{y})^T (X\theta - \vec{y}) = \frac{1}{2}\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$
$$= J(\theta)$$

$\therefore\ \nabla_{A^T} \operatorname{tr} ABA^T C = B^T A^T C^T + BA^T C$

$$\nabla_\theta J(\theta) = \nabla_\theta \frac{1}{2}(X\theta - \vec{y})^T (X\theta - y)$$
$$= \frac{1}{2}\nabla_\theta (\theta^T X^T - \vec{y}^T)(X\theta - y)$$
$$= \frac{1}{2}\nabla_\theta (\theta^T X^T X\theta - \theta^T X^T y - y^T X\theta + y^T y)$$
$$= \frac{1}{2}\nabla_\theta \operatorname{tr}(\theta^T X^T X\theta - \theta^T X^T y - y^T X\theta + y^T y)$$
$$= \frac{1}{2}\nabla_\theta (\operatorname{tr}\theta^T X^T X\theta - 2\operatorname{tr}\vec{y}^T X\theta) \quad ???$$
$$= \frac{1}{2}(X^T X\theta + X^T X\theta - 2X^T\vec{y})$$
$$= X^T X\theta - X^T y$$

$\nabla_\theta J(\theta) \overset{\text{set}}{=} \vec{0}$ for θ st $J(\theta)$ in min^n

$$\nabla_\theta J(\theta) = X^T X\theta - X^T y = 0$$
$$X^T X\theta - X^T y = 0$$
$$\boxed{\theta = X^T y (X^T X)^{-1}}$$