$$J(\boldsymbol{\theta}) = (\underbrace{\mathbf{y} - \mathbf{X}\boldsymbol{\theta}}_{\Delta y})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\boldsymbol{\theta})^2$$

$\boxed{\text{Prove}}$

$$x^T\theta = \underline{\hat{y}}$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}^T \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix}$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix}$$

$$\begin{bmatrix} \theta_1 x_{11} + \theta_2 x_{12} + \theta_3 x_{13} \\ \theta_1 x_{21} + \theta_2 x_{22} + \theta_3 x_{23} \\ \theta_1 x_{31} + \theta_2 x_{32} + \theta_3 x_{33} \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \Delta y_1 \\ \Delta y_2 \\ \Delta y_3 \end{bmatrix}^T \begin{bmatrix} \Delta y_1 \\ \Delta y_2 \\ \Delta y_3 \end{bmatrix}$$

$$= \begin{pmatrix} \Delta y_1 & \Delta y_2 & \Delta y_3 \end{pmatrix} \begin{bmatrix} \Delta y_1 \\ \Delta y_2 \\ \Delta y_3 \end{bmatrix}$$

$$= \Delta y_1^2 + \Delta y_2^2 + \Delta y_3^2 = \sum_{i=1}^{n} (y_i - x_i^T\theta)^2$$

$$J(\theta) = (\overset{n\times d}{y - X} \overset{d\times 1}{\theta})^T (y - X\theta)$$

$$= y^T y - 2(\underbrace{X\theta}_{})^T y + (\overset{n\times 1}{X\theta})^T \overset{n\times 1}{X\theta}$$

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

$$\boxed{J(\theta) = y^T y - 2 y^T \hat{y} + \hat{y}^T \cdot \hat{y}}$$

Transpose of a product reverses the order of the factors and switches the transposition

switching order of factors causes switching transpose

$$\frac{dJ(\theta)}{d\theta} = 0 - 2\overset{1\times n}{y^T}\overset{n\times d}{X} + 2X^T X\theta$$

$$0 \overset{!}{=} -2y^T X + 2X^T X\hat{\theta}$$

$$\cancel{2} y^T X = \cancel{2} X^T X \hat{\theta}$$

$$\hat{\theta} = \frac{y^T X}{X^T X}$$

$$\hat{\theta} = (X^T X)^{-1} y^T X$$

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

$$\hat{y} = X_* \hat{\theta} = \underbrace{X (X^T X)^{-1} X^T}_{} y$$

$$= Hy$$