

$$\vec{y} = \underline{X} \vec{w} + \vec{\epsilon} \Rightarrow \vec{\epsilon} = \vec{y} - \underline{X} \vec{w}$$

$$y_i = \sum_j x_{ij} w_j + \epsilon_i$$

$$g = \sum_i \epsilon_i^2$$

$$\min_w g = \sum_i \epsilon_i^2$$

$$\sum_i \epsilon_i^2 = \vec{\epsilon}^T \vec{\epsilon}$$

$$g = \vec{\epsilon}^T \vec{\epsilon}$$

$$g = (\vec{y} - \underline{X} \vec{w})^T (\vec{y} - \underline{X} \vec{w})$$

$$g = (\vec{y}^T - \vec{w}^T \underline{X}^T) (\vec{y} - \underline{X} \vec{w})$$

$$g = \vec{y}^T \vec{y} - \underbrace{\vec{y}^T \underline{X} \vec{w}}_{= \vec{w}^T \underline{X}^T \vec{y}} - \vec{w}^T \underline{X}^T \vec{y} + \vec{w}^T \underline{X}^T \underline{X} \vec{w}$$

$$g(\vec{w}) = \vec{y}^T \vec{y} - 2 \vec{y}^T \underline{X} \vec{w} + \vec{w}^T \underline{X}^T \underline{X} \vec{w}$$

$$\frac{\partial g}{\partial \vec{w}} = 0 = \frac{\partial}{\partial \vec{w}} \vec{y}^T \vec{y} - \frac{\partial}{\partial \vec{w}} (2 \vec{y}^T \underline{X} \vec{w}) + \frac{\partial}{\partial \vec{w}} (\vec{w}^T \underline{X}^T \underline{X} \vec{w})$$

$$\frac{\partial (\underline{A} \vec{x})}{\partial \vec{x}} = \underline{A}^T$$

$$\frac{\partial (\vec{x}^T \underline{A} \vec{x})}{\partial \vec{x}} = (\underline{A}^T + \underline{A}) \vec{x}$$

$$\begin{aligned} \frac{\partial}{\partial \vec{w}} (2 \vec{y}^T \underline{X} \vec{w}) &= 2 \frac{\partial}{\partial \vec{w}} (\vec{y}^T \underline{X} \vec{w}) \\ &= 2 (\vec{y}^T \underline{X})^T = 2 \underline{X}^T \vec{y} \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \vec{w}} (\vec{w}^T \underline{X}^T \underline{X} \vec{w}) &= (\underline{X}^T \underline{X} + \underline{X} \underline{X}^T) \vec{w} \\ &= 2 \underline{X}^T \underline{X} \vec{w} \end{aligned}$$

$$\frac{\partial g}{\partial \vec{w}} = -2 \underline{X}^T \vec{y} + 2 \underline{X}^T \underline{X} \vec{w}$$

$$\frac{\partial g}{\partial \vec{w}} = 0 = -\cancel{2} \underline{\underline{X}}^T \vec{y} + \cancel{2} \underline{\underline{X}}^T \underline{\underline{X}} \vec{w}$$

$$\underbrace{\underline{\underline{X}}^T \underline{\underline{X}}}_{\underline{\underline{A}}} \vec{w} = \underbrace{\underline{\underline{X}}^T \vec{y}}_{\vec{b}}$$