3) a. Proof : Sum of Squares Error Function = Convex Function of Weight Parameter

Method : Show Hessian is a Symmetric Positive Semi-Definite Matrix.

- Sum of Squares Error

$$err(\hat{y}, y) = \sum_{i=1}^{t} (\hat{y}_i - y_i)^2$$

$$w^* = \underset{w}{argmin} \sum_{i=1}^{t} |X_{i:} w - y_i|^2$$

$$J(w) = (Xw - y)^T (Xw - y)$$

$$= (w^T X^T - y^T)(Xw - y)$$

$$= X^T X w w^T - \underline{X w y^T} - \underline{X^T w^T y} + y y^T$$

$$= X^T X w w^T - 2X^T w^T y + y y^T$$

$$\nabla_w J(w) = 2XX^T w - 2X^T y$$

$$H_w = 2(X^T X)$$

Show $X^T X$ (the Hessian) is symmetric positive semi-definite

- $\forall z, \; (\underbrace{z^T X^T}_{u^T})(\underbrace{X z}_{u}) \geq 0$

$u^T \cdot u = \sum_{i=1}^{t} u_i^2$ will always be <u>positive</u>

and since $u^T = u$, the matrix satisfies conditions for <u>symmetry</u>.

3) b. Compute gradient $\nabla f(x)$ & $\nabla^2 f(x)$

$$f(x) = 100(x_2 - x_1^2)^2 + (1-x_1)^2 \quad // \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- $\nabla_{x_1} f(x) = 2 \cdot 100(x_2 - x_1^2) \cdot -2x_1 + 2 \cdot (1-x_1)$
  $= -400x_1(x_2 - x_1^2) + 2(1-x_1)$
- $\nabla_{x_2} f(x) = 2 \cdot 100(x_2 - x_1^2) + 0$

$$\therefore \quad \nabla f = \begin{bmatrix} \frac{df}{dx_1} \\ \frac{df}{dx_2} \end{bmatrix} = \begin{bmatrix} -400x_1(x_2 - x_1^2) + 2(1-x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

- $H = \begin{bmatrix} \frac{dh_1}{dx_1} & \frac{dh_1}{dx_2} \\ \frac{dh_2}{dx_1} & \frac{dh_2}{dx_2} \end{bmatrix} \qquad h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$

$$\therefore \quad H = \begin{bmatrix} -400x_2 + 1200x_1 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

- Show $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is only minimizer:

$$H\left(\begin{bmatrix} 1 ; 1 \end{bmatrix}\right) = \begin{bmatrix} -400 + 1200 + 2 & -400 \\ -400 & 200 \end{bmatrix} = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

$$\begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}^T = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} \geq 0$$

$\therefore$ symmetric matrix & PSD.