(1) PallawK2 Pallaw Kumar

(1) $softmax(z) = softmax(z - c1)$

$$softmax(z - c1)_i = \frac{e^{z_i - c1}}{\sum_j e^{z_j - c1}}$$

$$= \frac{e^{z_i}}{\cancel{e^{c1}} \frac{\sum e^{z_j}}{\cancel{e^{c1}}}} = \frac{e^{z_i}}{\sum e^{z_j}} \quad \text{— (1)}$$

$$softmax(z)_i = \frac{e^{z_i}}{\sum e^{z_j}} \quad \text{— (2)}$$

from (1) and (2)

$softmax(z) = softmax(z - c1)$

(2) when we try to find the derivative of the softmax function, we talk about a Jacobian matrix, which is the matrix of all first order partial derivatives of vector-valued function.

$$y_i = softmax(z)_i = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}}$$

$$\frac{\partial y_i}{\partial z_j} = \frac{\partial \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}}}{\partial z_j}$$

$$h_i = \sum_{k=1}^{K} e^{z_j}$$

No matter which $z_j$ when we compute the derivative of $h_i$ with respect to $z_j$, it will be $e^{z_j}$.

when $i = j$ (derive the gradient of the diagonal of the Jacobian matrix)

$$\frac{\partial \frac{e^{z_i}}{\sum_{k=1} e^{z_k}}}{\partial z_j} = \frac{e^{z_i} \sum_{k=1}^{M} e^{z_k} - e^{z_j} e^{z_i}}{\left( \sum_{k=1}^{K} e^{z_k} \right)^2}$$

$$= \frac{e^{z_i} \sum_{k=1}^{K} e^{z_k} - e^{z_j} e^{z_i}}{\left( \sum_{k=1} e^{z_k} \right)^2}$$

$$= \frac{e^{z_i} \left( \sum_{k=1}^{K} e^{z_k} - e^{z_j} \right)}{\left( \sum_{k=1}^{K} e^{z_k} \right)^2}$$

$$= y_i (1 - y_j)$$

similarly, deriving the off diagonal entries of the Jacobian matrix will yield:

If $i \neq j$

$$\frac{\partial \frac{e^{z_i}}{\sum_{k=1}^{M} e^{z_k}}}{\partial z_j} = \frac{0 - e^{z_j} e^{z_i}}{\left( \sum_{k=1}^{M} e^{z_k} \right)^2}$$

$$= \frac{- e^{z_j} e^{z_i}}{\left( \sum_{k=1}^{M} e^{z_k} \right)^2}$$

$$= - y_i y_j$$

$$\frac{\partial y_i}{\partial z_j} = \begin{cases} y_i(1 - y_j) & i = j \\ - y_i y_j & i \neq j \end{cases}$$

$$\frac{\partial y_i}{\partial z_j} = y_i(\delta_{ij} - y_j)$$ where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise

(3) $z = w^T x - u$ ; $y_i = \dfrac{e^{w^T x_i - u}}{\sum_{j=1}^{K} e^{w^T x_j - u}}$

$$\frac{dz}{dx} = w^T \quad ; \quad \frac{\partial z}{\partial w} = x \quad ; \quad \frac{\partial z}{\partial u} = -1$$

$$\frac{\partial y_i}{\partial x} = \frac{\partial y_i}{\partial z} \frac{dz}{dx} = y_i(\delta_{ij} - y_j) * w^T$$

$$\frac{\partial y_i}{\partial w} = y_i(\delta_{ij} - y_j)x \quad \text{from } \frac{\partial y_j}{\partial w} = \frac{\partial y_i}{\partial z}\frac{\partial z}{\partial w}$$

$$\frac{\partial y_i}{\partial x} = y_i(\delta_{ij} - y_j)w^T \quad \text{from } \frac{\partial y_i}{\partial x} = \frac{\partial y_i}{\partial z}\frac{\partial z}{\partial x}$$

$$\frac{\partial y_i}{\partial u} = y_i(\delta_{ij} - y_j) * -1 \quad \text{from } \frac{\partial y_i}{\partial u} = \frac{\partial y_i}{\partial z}\frac{\partial z}{\partial u}$$

② ①

$$v = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

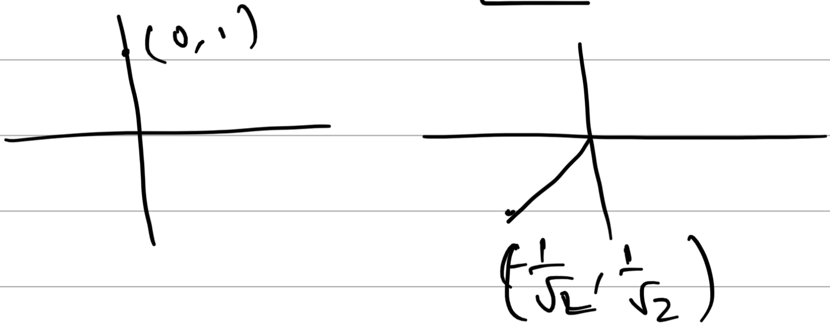
$$vx = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$


Matrix v rotates the input vector by 45 degrees in the counter-clockwise direction, scaling it by a factor of $1/\sqrt{2}$.

$$Vx = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$




Matrix v rotates the input vector by 135 degrees in the counter-clockwise direction, scaling it by a factor of $1/\sqrt{2}$.

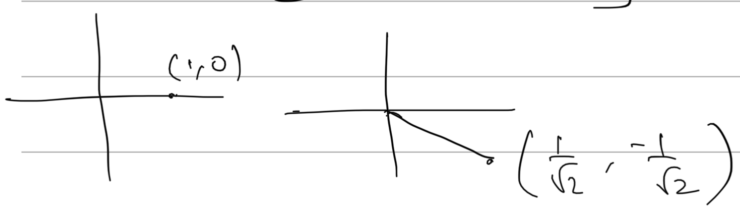
$$② \quad V^{-1} = \frac{1}{\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} - \left(\frac{-1}{\sqrt{2}}\right) \cdot \frac{1}{\sqrt{2}}} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

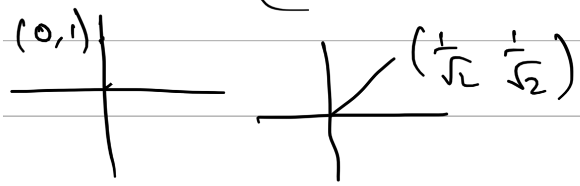
$$V^{-1} = V^T$$

$$V^T x = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$




Matrix V rotates the input vector by 45 degrees in the clockwise direction, scaling it by a factor of $1/\sqrt{2}$.

$$V^T x = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$




Matrix V rotates the input vector by 45 degees in the clockwise direction, scaling it by a factor of $\frac{1}{\sqrt{2}}$

③ $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$

$$\Sigma V^T = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -\frac{1}{\sqrt{2}} & 1/\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -2\sqrt{2} & 2\sqrt{2} \end{bmatrix}$$

(a) $\Sigma V^T x = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -2\sqrt{2} & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

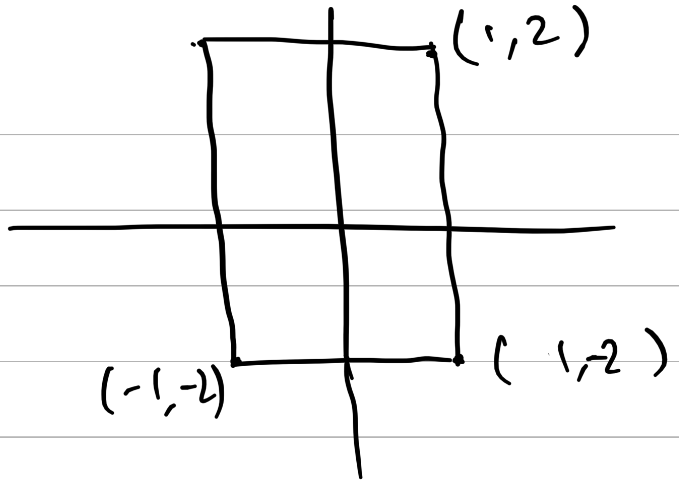
(b) $\Sigma V^T x = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -2\sqrt{2} & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

(c) $\Sigma V^T x = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -2\sqrt{2} & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

(d) $\Sigma V^T x = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -2\sqrt{2} & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$

shape of the result point is rectangle.

$(-1, 2)$



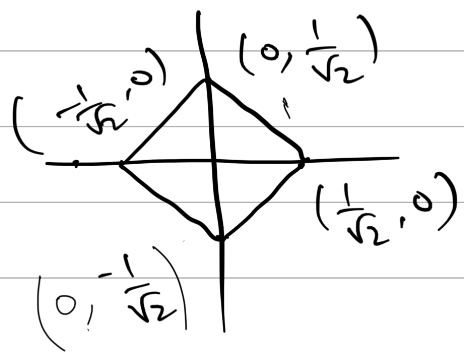

(4) $U = \begin{bmatrix} -\sqrt{3}/2 & 1/2 \\ -1/2 & -\sqrt{3}/2 \end{bmatrix}$

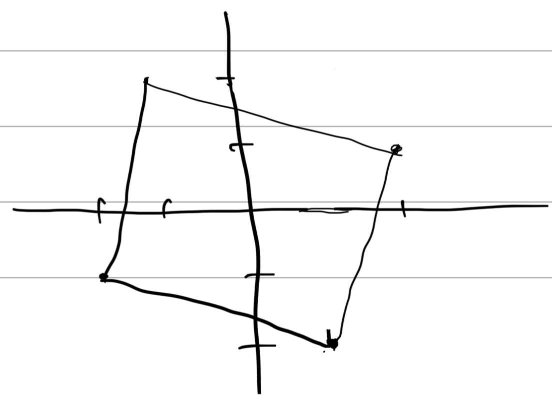
$$Ux = \begin{bmatrix} -\sqrt{3}/2 & 1/2 \\ -1/2 & -\sqrt{3}/2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{3}}{2\sqrt{2}} \\ -\frac{1}{2\sqrt{2}} \end{bmatrix}$$

$$Ux = \begin{bmatrix} -\sqrt{3}/2 & 1/2 \\ -1/2 & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sqrt{2}} \\ \frac{-\sqrt{3}}{2\sqrt{2}} \end{bmatrix}$$

$$Ux = \begin{bmatrix} -\frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} \end{bmatrix}$$

$$Ux = \begin{bmatrix} -\frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2\sqrt{2}} \\ \frac{\sqrt{3}}{2\sqrt{2}} \end{bmatrix}$$



U is rescaling all the points of $x$ and rotating it in the clock-wise direction.

(5) $A = U \Sigma V^T$

$$\Sigma V^T = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -2\sqrt{2} & 2\sqrt{2} \end{bmatrix}$$

$$U \Sigma V^T = \begin{bmatrix} -\frac{\sqrt{3}}{2} & 1/2 \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -2\sqrt{2} & 2\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} -\sqrt{3/2} - \frac{1}{2} & -\sqrt{3/2} + \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} + \sqrt{6} & -\frac{1}{\sqrt{2}} - \sqrt{6} \end{bmatrix}$$

Singular value decomposition (SVD) can be used on matrix $B$ to produce $B = U \Sigma V^T$. where $U$ and $V$ are orthogonal matrices and is a diagonal matrix with singular values on the diagonal. This results in a similar geometric interpretation for $Bx$. Then $Bx$ can be understood as $x$ rotated with $V$ first then rescaled along the coordinate axes with and then rotated with $U$ again. This makes it possible to conceptualize the matrix $B$ as applying rotation, scaling and rotation to the vector $x$.