1) a) In the autoencoder, we have a encoding layer that outputs the low dimensionality rep^{n} of x.

$\Rightarrow$ if $x \in R^n$, $Wx \in R^m$ is the low dimensional rep^{n} of x such that $W \in R^{m \times n}$ & $m < n$.

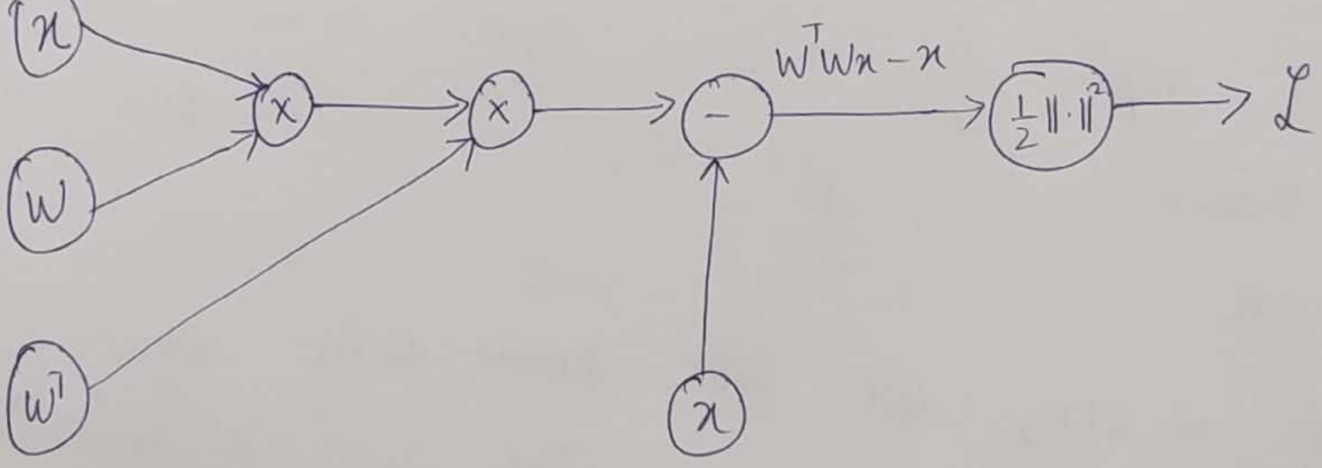
This rep^{n} of x is then passed to a decoder to get the reconstructed signal back.

As $m < n$, we lose some data when we perform Wx. ∴ The aim of encoding is to minimize this information loss.

To decode the rep^{n} back to original signal $\rightarrow$ $\underline{\underline{W^T W x}} \in R^n$

∴ by minimizing the Euclidean distance ($L2$-norm) b/n Wx & W^TWx we tune W, thus preserving most of the information.

(b) $$L = \frac{1}{2} \| W^T W x - x \|^2$$



c) From the computational graph in b, we can see that there are two paths to WL One corresponding to W. The other corresponding to W^T. ∴ The o/p's can be added to get the final cost function.

L = cost of path W + cost of path W^T

L_1 $\qquad$ L_2

$$\frac{\partial L}{\partial W} = \frac{\partial L_1}{\partial W} + \frac{\partial L_2}{\partial W}$$

$$L_1 = L_2 = L$$

$$\frac{\partial L_2}{\partial W} = \left(\frac{\partial L_2}{\partial W^T}\right)^T = \left(\frac{\partial L}{\partial W^T}\right)^T$$

$\therefore$ total gradient =

$$\nabla_W L = \frac{\partial L}{\partial W} = \frac{\partial L}{\partial W} + \underline{\underline{\left(\frac{\partial L}{\partial W^T}\right)^T}}$$

Q) Consider $z \in R^n$

$$f(z) = \|z\|_2^2$$

$$= \left(\left(\sum_{k=1}^{n} z_k^2\right)^{1/2}\right)^2 = \sum_{k=1}^{n} z_k^2$$

$$\frac{\partial f(z)}{\partial z_j} = \frac{\partial}{\partial z_j}\left(\sum_{k=1}^{n} z_k^2\right)$$

$$\boxed{\nabla_z f(z) = 2z_j}$$

Using the above derivative,

$$\mathcal{L} = \frac{1}{2}\| W^T W x - x \|^2$$

→ Back prop thro' $\frac{1}{2}\|\cdot\|^2$

$$\frac{\partial L}{\partial (W^T W x - x)} = \frac{1}{2} \cdot 2\,(W^T W x - x)$$

$$= W^T \underline{\underline{Wx - x}}$$

→ Back prop thro subtraction, derivative is passed

$$\frac{\partial L}{\partial W^T W x} = \frac{\partial L}{\partial (W^T W x - x)} \quad \text{—— (1)}$$

$$\rightarrow \quad \frac{\partial L}{\partial W x} = (W^T)^T \left(\frac{\partial L}{\partial W^T W x}\right) \quad \langle \text{using hint} \rangle$$

Scanned by CamScanner

$$\therefore \frac{\partial L}{\partial Wx} = W(W^T Wx - x)$$

→ Back prop. to W,

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial Wx} x^T$$

$$\frac{\partial L}{\partial W} = W(W^T Wx - x)x^T$$

From (1), back prop. to W^T

$$\frac{\partial L}{\partial W^T} = \frac{\partial L}{\partial W^T Wx} (Wx)^T$$

$$= (W^T Wx - x)(Wx)^T$$

$$\therefore \nabla_W L = \frac{\partial L}{\partial W} + \left(\frac{\partial L}{\partial W^T}\right)^T$$

$$\therefore \nabla_W L = W(W^T W x - x)x^T + \left((W^T W x - x)(Wx)^T\right)^T$$

$$\therefore \nabla_W L = W W^T W x x^T - W x x^T + W x (x^T W^T W - x^T)$$

$$\nabla_W L = W W^T W x x^T - W x x^T + W x x^T W^T W - W x x^T$$

$$\therefore \boxed{\nabla_W L = W(W^T W x x^T + x x^T W^T W - 2 x x^T)}$$