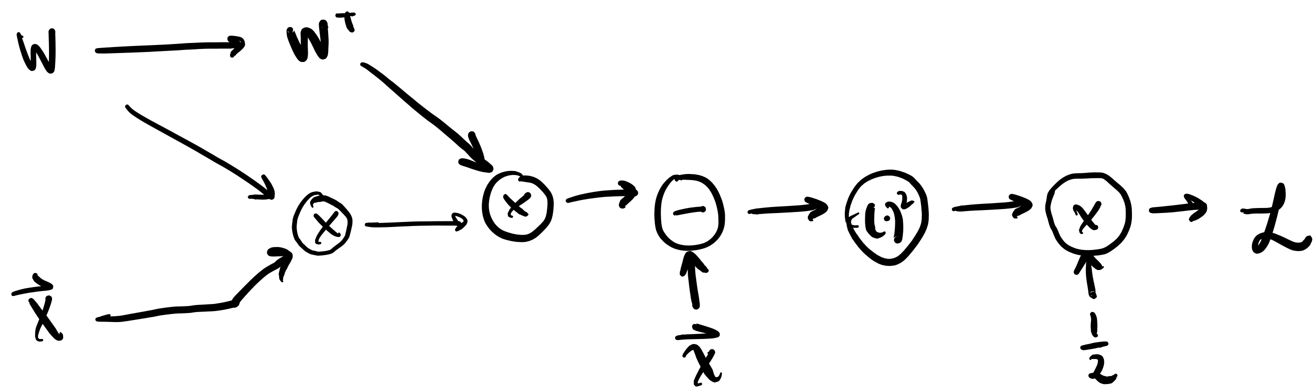
1. Backpropagation for auto encoders

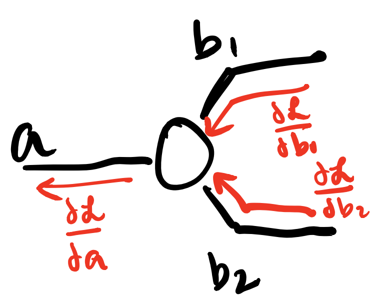
a) The dimensionality reduction key to an autoencoder comes from $\overset{m\times n}{W}\overset{n\times 1}{\vec{x}}$ which is $m\times 1$ and has lower dimensionality than $\vec{x}$ $(n\times 1)$ since $m<n$. However, multiplication by W^T is the decoding step that reconstructs data from the encodings since W^T is $n\times m$ so $\overset{n\times m}{W^T}\overset{m\times n}{W}\overset{n\times 1}{\vec{x}}$ is $n\times 1$, the same size as input $\vec{x}$, therefore preserving dimensionality.

Autoencoders, especially linear ones like in this example, are similar to PCA in which we find the eigenvectors and eigenvalues of the covariance matrix that minimizes the reconstruction error. Minimizing the given loss function achieves the same objective of finding W that saves the important information about $\vec{x}$.

b)



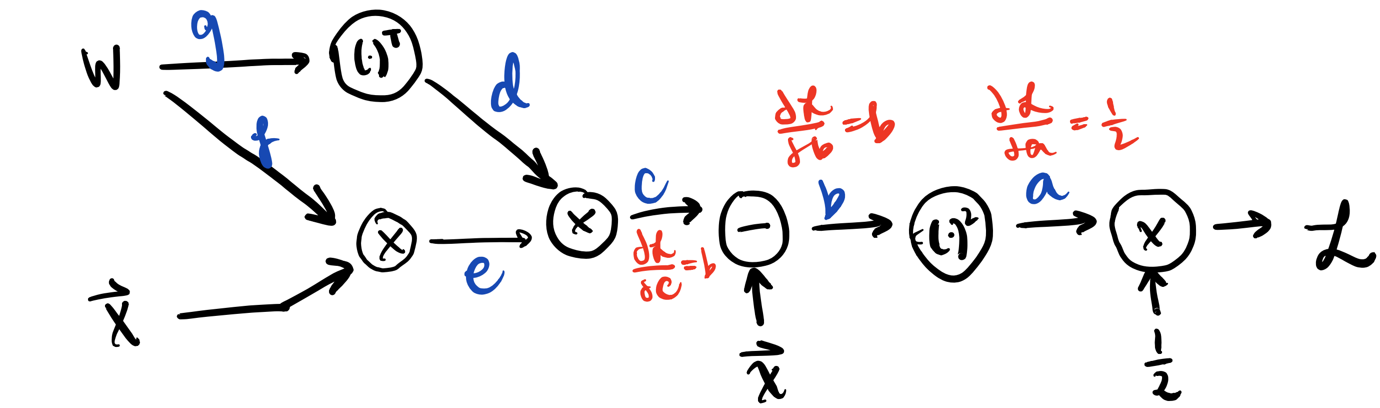
c) As per the law of total derivatives, we sum over the partial derivatives of each of the two paths in order to count each path's contribution to $\nabla_W \mathcal{L}$. For example, in the shown computational graph, $\nabla_a \mathcal{L}$ is:



$$\frac{\partial \mathcal{L}}{\partial a} = \sum_{i=1}^{n} \frac{\partial \mathcal{L}}{\partial b_i} \cdot \frac{\partial b_i}{\partial a} = \frac{\partial b_1}{\partial a}\frac{\partial \mathcal{L}}{\partial b_1} + \frac{\partial b_2}{\partial a}\frac{\partial \mathcal{L}}{\partial b_2}$$

so both $\frac{\partial \mathcal{L}}{\partial b_1}$ and $\frac{\partial \mathcal{L}}{\partial b_2}$ influence the value of $\frac{\partial \mathcal{L}}{\partial a}$.

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



$$\frac{\partial \mathcal{L}}{\partial a} = \frac{\partial}{\partial a}\left(\frac{1}{2}a\right) = \frac{1}{2}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \frac{\partial a}{\partial b}\frac{\partial \mathcal{L}}{\partial a} = \frac{\partial}{\partial b}b^2 \cdot \frac{1}{2} = 2b \cdot \frac{1}{2} = b$$

$$\frac{\partial \mathcal{L}}{\partial c} = \frac{\partial b}{\partial c}\frac{\partial \mathcal{L}}{\partial b} = \frac{\partial}{\partial c}(c - \vec{x}) \cdot b = 1 \cdot b = b$$

$$\frac{\partial \mathcal{L}}{\partial d} = \frac{\partial c}{\partial d}\frac{\partial \mathcal{L}}{\partial c} = \frac{\partial(W^T W x)}{\partial(W^T)} \cdot b = \underset{1\times m}{(Wx)^T}\ \underset{n\times 1}{b} = b(Wx)^T$$

$$\frac{\partial \mathcal{L}}{\partial e} = \frac{\partial c}{\partial e}\frac{\partial \mathcal{L}}{\partial c} = \frac{\partial(W^T W x)}{\partial(Wx)} \cdot b = Wb$$

$$\frac{\partial \mathcal{L}}{\partial f} = \frac{\partial e}{\partial f}\frac{\partial \mathcal{L}}{\partial e} = \frac{\partial(Wx)}{\partial W} \cdot Wb = \underset{1\times n}{x^T}\ \underset{m\times 1}{Wb} \to Wbx^T$$

$$\frac{\partial \mathcal{L}}{\partial g} = \left(\frac{\partial \mathcal{L}}{\partial d}\right)^T = Wxb^T$$

$$\frac{\partial \mathcal{L}}{\partial W} = \frac{\cancel{\partial f}'}{\cancel{\partial W}} \cdot \frac{\partial \mathcal{L}}{\partial f} + \frac{\cancel{\partial g}'}{\cancel{\partial W}}\frac{\partial \mathcal{L}}{\partial g}$$

$$= Wbx^T + Wxb^T$$

$$= \boxed{W(W^TWx - x)x^T + Wx(W^TWx - x)^T}$$