ASSIGNMENT - 6 PRIYANSH TRIVEDI, NILESH CHAKRABORTY

Let us consider an autoencoder network that consists of an encoder $h(x) = g(a(x))$ and a decoder, $f(x) = o(\hat{a}(x))$. In our cases $g(\cdot)$ and $o(\cdot)$ are linear functions. The weights of $a(x)$ and $\hat{a}(x)$ are tied. Thus, our network looks like:

$$h(x) = Wx + b_1$$
$$f(x) = W^T(Wx + b_1) + b_2$$

Ignoring the biases, our error function for reconstruction minimization becomes:

$$\cancel{\underset{W}{\arg\min} \sum_i \| W^T W x_i}$$

$$\underset{W}{\arg\min} \sum_{i=1}^{n} \| W^T W x_i - x_i \|_F^2$$

where x_i is the i'th row of matrix $X \in \mathbb{R}^{n \times d}$. ~~with~~ Each row of X is a data example of d dimensions. $W \in \mathbb{R}^{k \times d}$

$$\begin{aligned}
\mathcal{L} &= \sum_i \| W^T W x_i - x_i \|_F^2 \\
&= \sum_i (W^T W x_i - x_i)^T (W^T W x_i - x_i) \\
&= \sum_i (W^T W - I)^T x_i^T x_i (W^T W - I)
\end{aligned}$$

[multiplying by I]

$$= Tr[(W^T W - I) \sum_i x_i x_i^T (W^T W - I)^T]$$

$[y^T z = Tr(y z^T)]$

$\sum_i (x_i x_i^T)$ is a sum of outer products of x_i and thus an unnormalized covariance matrix. If we assume

$$x_i \leftarrow \frac{1}{\sqrt{n}}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)$$

then the Singular Value Decomposition of X corresponds to the eigendecomposition of X, which is used for Principal Component Analysis. Let $\sum_i (x_i x_i^T) = V \Lambda V^T$ where V is the ~~righ~~ eigenvector matrix and Λ is the diagonal eigenvalue ~~diagon~~ matrix.

From (*), we have,

$$\mathcal{L} = \mathrm{Tr}\left[(W^T W - I) V \Lambda V^T (W^T W - I)\right]$$

Since $\sum_i (x_i x_i^T)$ is positive semi-definite, the eigenvalues Λ have square roots.

$$\mathcal{L} = \|(W^T W - I) V \Lambda^{1/2}\|_F^2$$

Let W be decomposed into:

$$W = Q S U^T$$

$$\therefore W^T W = (Q S U^T)^T Q S U^T$$

$$= U S^T Q^T Q S U^T$$

$$= U \hat{S}^2 U^T$$

$\hat{S}$ is made square diagonal, with excess zeroes chopped off.

$$\mathcal{L} = \| (U\hat{S}^2 U^T - I) V \Lambda^{1/2} \|_F^2$$

$$= \| (U(I - \hat{S}^2) U^T) V \Lambda^{1/2} \|_F^2$$

$$= \| (U \sum_i ((1 - s_i^2) I_i) U^T) V \Lambda^{1/2} \|_F^2$$

I_i is a digonal matrix with $I_i^{(i,i)} = 1$.

$$\mathcal{L} = \sum_i (1 - s_i^2) \| U I_i U^T V \Lambda^{1/2} \|_F^2$$

We need to minimize $\mathcal{L}$. So, we set ~~$U=V$~~, $V = U$.

$$\mathcal{L} = \sum_i (1 - s_i^2) \| V I_i V^T V \Lambda^{1/2} \|_F^2$$

$$= \sum_i (1 - s_i^2)^2 (V^T V \Lambda)$$

$$= \sum_i (1 - s_i^2)^2 \lambda_i \quad [V \text{ is orthonormal}, \; V^T V = I, \; \lambda_i = \Lambda^{(i,i)}]$$

The loss is minimized by a certain configuration of the eigen values s_i and λ_i, of matrices W and X respectively.

Thus, for optimal pair of encoder and decoder,

$$\cancel{h(x) = W^T x = \hat{S}^T U^T x}$$

$$\cancel{h(x) = W}$$

$$h(x) = Wx = S V^T x$$

$$o(x) = W^T W x = V S^T h(x)$$

A linear autoencoder, with tied weights and K hidden states, learns to pick the K best eigenvectors, exactly like PCA.