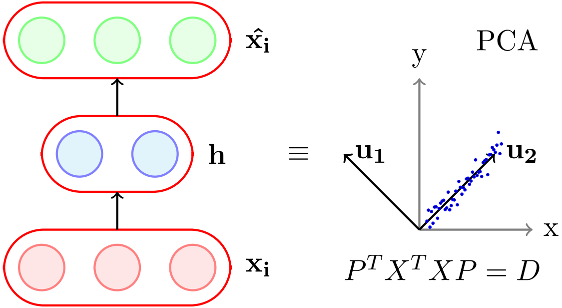
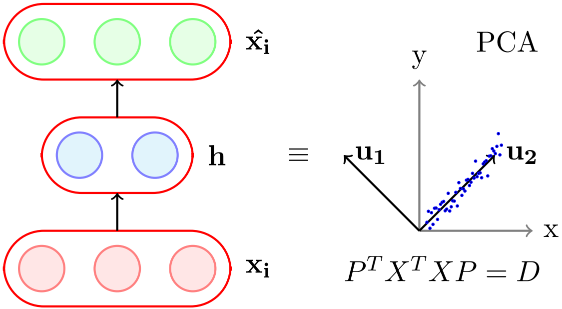
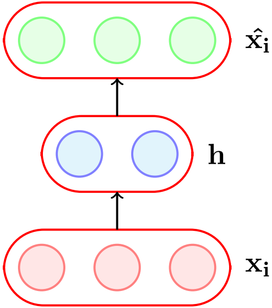
Module 7.2: Link between PCA and Autoencoders



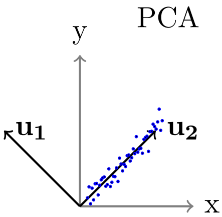
- We will now see that the encoder part of an autoencoder is equivalent to PCA if we



- We will now see that the encoder part of an autoencoder is equivalent to PCA if we
  - use a linear encoder

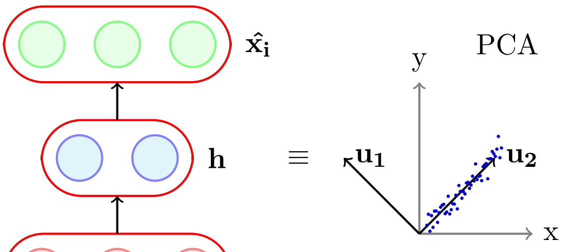


$\equiv$



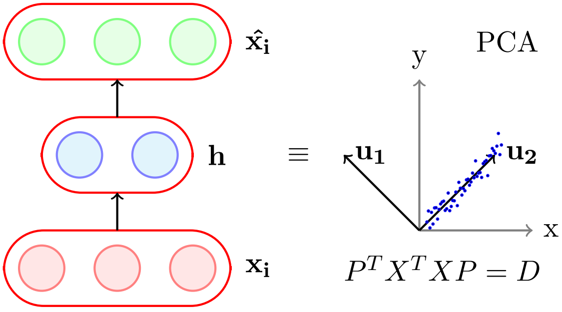
$$P^T X^T X P = D$$

- We will now see that the encoder part of an autoencoder is equivalent to PCA if we
  - use a linear encoder
  - use a linear decoder



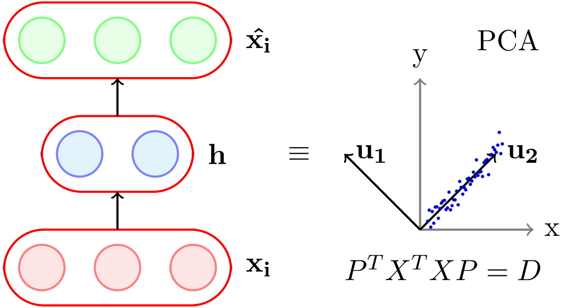
$$P^T X^T X P = D$$

- We will now see that the encoder part of an autoencoder is equivalent to PCA if we
  - use a linear encoder
  - use a linear decoder
  - use squared error loss function



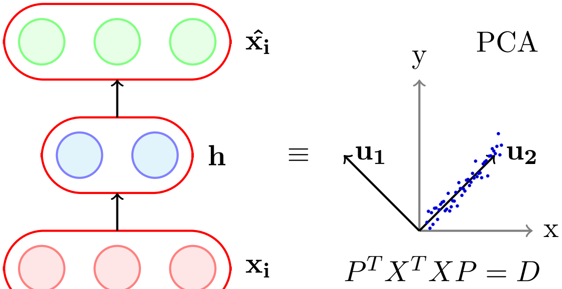
- We will now see that the encoder part of an autoencoder is equivalent to PCA if we
  - use a linear encoder
  - use a linear decoder
  - use squared error loss function
  - normalize the inputs to

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right)$$



- First let us consider the implication of normalizing the inputs to

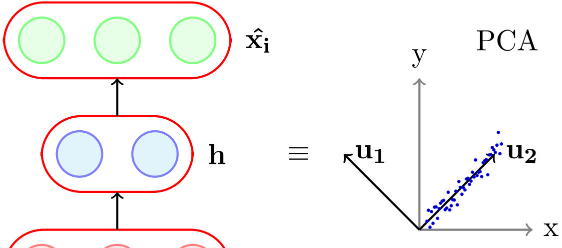
$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right)$$



- First let us consider the implication of normalizing the inputs to

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right)$$

- The operation in the bracket ensures that the data now has 0 mean along each dimension j (we are subtracting the mean)

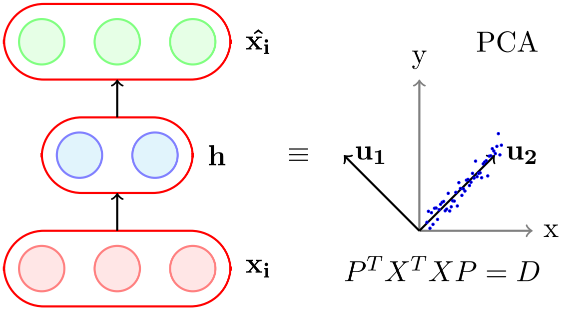


$P^TX^TXP = D$

- First let us consider the implication of normalizing the inputs to

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right)$$

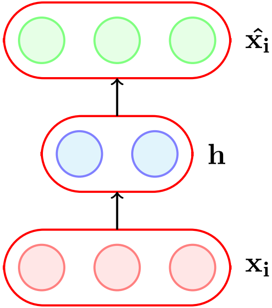
- The operation in the bracket ensures that the data now has 0 mean along each dimension j (we are subtracting the mean)
- Let X' be this zero mean data matrix then what the above normalization gives us is $X = \frac{1}{\sqrt{m}}X'$



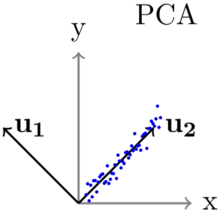
- First let us consider the implication of normalizing the inputs to

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right)$$

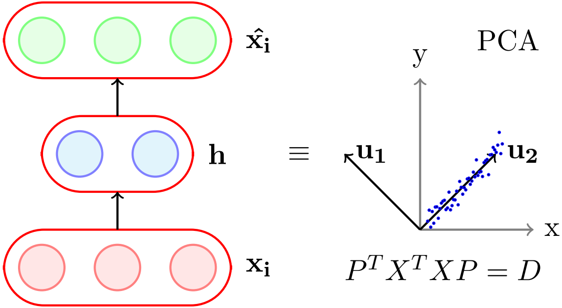
- The operation in the bracket ensures that the data now has 0 mean along each dimension j (we are subtracting the mean)
- Let X' be this zero mean data matrix then what the above normalization gives us is $X = \frac{1}{\sqrt{m}}X'$
- Now $(X)^TX = \frac{1}{m}(X')^TX'$ is the covariance matrix (recall that covariance matrix plays an important role in PCA)



$\equiv$

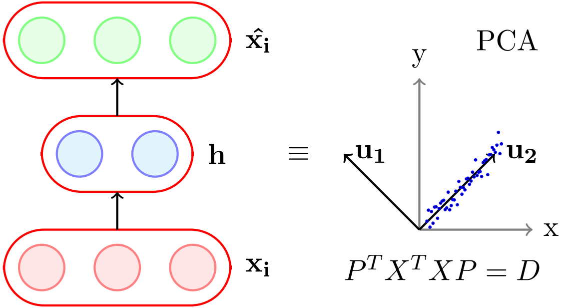


$$P^T X^T X P = D$$

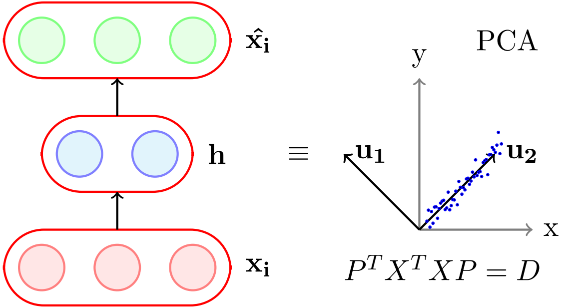


$$P^T X^T X P = D$$

- First we will show that if we use linear decoder and a squared error loss function then

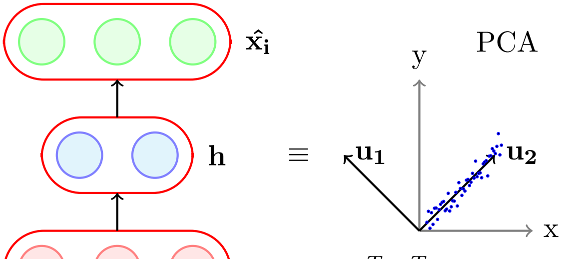


- First we will show that if we use linear decoder and a squared error loss function then
- The optimal solution to the following objective function



- First we will show that if we use linear decoder and a squared error loss function then
- The optimal solution to the following objective function

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \hat{x}_{ij})^2$$



$$P^T X^T X P = D$$

- First we will show that if we use linear decoder and a squared error loss function then
- The optimal solution to the following objective function

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \hat{x}_{ij})^2$$

is obtained when we use a linear encoder.

$$\min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2 \tag{1}$$

$$\min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2 \tag{1}$$

- This is equivalent to

$$\min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2 \tag{1}$$

- This is equivalent to

$$\min_{W^* H} (\|X - HW^*\|_F)^2$$

$$\min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2 \tag{1}$$

- This is equivalent to

$$\min_{W^* H} (\|X - HW^*\|_F)^2 \qquad \|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

$$\min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2 \tag{1}$$

- This is equivalent to

$$\min_{W^* H} (\|X - HW^*\|_F)^2 \qquad \|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

(just writing the expression (1) in matrix form and using the definition of $||A||_F$) (we are ignoring the biases)

$$\min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2 \quad (1)$$

- This is equivalent to

$$\min_{W^*H} (\|X - HW^*\|_F)^2 \qquad \|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

(just writing the expression (1) in matrix form and using the definition of $||A||_F$) (we are ignoring the biases)

- From SVD we know that optimal solution to the above problem is given by

$$HW^* = U_{.,\leq k} \Sigma_{k,k} V_{.,\leq k}^T$$

$$\min_{\theta} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2 \tag{1}$$

- This is equivalent to

$$\min_{W^*H} (\|X - HW^*\|_F)^2 \qquad \|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

  (just writing the expression (1) in matrix form and using the definition of $||A||_F$) (we are ignoring the biases)

- From SVD we know that optimal solution to the above problem is given by

$$HW^* = U_{.,\leq k} \Sigma_{k,k} V^T_{.,\leq k}$$

- By matching variables one possible solution is

$$H = U_{.,\leq k} \Sigma_{k,k}$$
$$W^* = V^T_{.,\leq k}$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$H = U_{\cdot, \leq k} \Sigma_{k,k}$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$\begin{aligned} H &= U_{.,\leq k}\Sigma_{k,k} \\ &= (XX^T)(XX^T)^{-1}U_{.,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I\textit{)} \end{aligned}$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$
\begin{aligned}
H &= U_{\cdot,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{\cdot,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I\textit{)} \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T\textit{)}
\end{aligned}
$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$
\begin{aligned}
H &= U_{\cdot,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{\cdot,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I\textit{)} \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T\textit{)} \\
&= XV\Sigma^T U^T(U\Sigma\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && (V^T V = I)
\end{aligned}
$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$\begin{aligned}
H &= U_{\cdot,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{\cdot,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I\textit{)} \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T\textit{)} \\
&= XV\Sigma^T U^T(U\Sigma\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && (V^T V = I) \\
&= XV\Sigma^T U^T U(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((ABC)^{-1} = C^{-1}B^{-1}A^{-1})
\end{aligned}$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$
\begin{aligned}
H &= U_{\cdot,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{\cdot,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I\textit{)} \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T\textit{)} \\
&= XV\Sigma^T U^T(U\Sigma\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && (V^TV = I) \\
&= XV\Sigma^T U^T U(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((ABC)^{-1} = C^{-1}B^{-1}A^{-1}) \\
&= XV\Sigma^T(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && (U^TU = I)
\end{aligned}
$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$
\begin{aligned}
H &= U_{\cdot,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{\cdot,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I\textit{)} \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T\textit{)} \\
&= XV\Sigma^T U^T(U\Sigma\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && (V^T V = I) \\
&= XV\Sigma^T U^T U(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((ABC)^{-1} = C^{-1}B^{-1}A^{-1}) \\
&= XV\Sigma^T(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && (U^T U = I) \\
&= XV\Sigma^T\Sigma^{T^{-1}}\Sigma^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((AB)^{-1} = B^{-1}A^{-1})
\end{aligned}
$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$
\begin{aligned}
H &= U_{\cdot,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{\cdot,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I\textit{)} \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T\textit{)} \\
&= XV\Sigma^T U^T (U\Sigma\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && (V^T V = I) \\
&= XV\Sigma^T U^T U(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((ABC)^{-1} = C^{-1}B^{-1}A^{-1}) \\
&= XV\Sigma^T(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && (U^T U = I) \\
&= XV\Sigma^T\Sigma^{T^{-1}}\Sigma^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((AB)^{-1} = B^{-1}A^{-1}) \\
&= XV\Sigma^{-1}I_{\cdot,\leq k}\Sigma_{k,k} && (U^T U_{\cdot,\leq k} = I_{\cdot,\leq k})
\end{aligned}
$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$
\begin{aligned}
H &= U_{\cdot,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{\cdot,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I\textit{)} \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T\textit{)} \\
&= XV\Sigma^T U^T (U\Sigma\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && (V^T V = I) \\
&= XV\Sigma^T U^T U(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((ABC)^{-1} = C^{-1}B^{-1}A^{-1}) \\
&= XV\Sigma^T(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && (U^T U = I) \\
&= XV\Sigma^T \Sigma^{T^{-1}}\Sigma^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((AB)^{-1} = B^{-1}A^{-1}) \\
&= XV\Sigma^{-1}I_{\cdot,\leq k}\Sigma_{k,k} && (U^T U_{\cdot,\leq k} = I_{\cdot,\leq k}) \\
&= XVI_{\cdot,\leq k} && (\Sigma^{-1}I_{\cdot,\leq k} = \Sigma^{-1}_{k,k})
\end{aligned}
$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$
\begin{aligned}
H &= U_{\cdot,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{\cdot,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I) \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T) \\
&= XV\Sigma^T U^T(U\Sigma\Sigma^T U^T)^{-1}U_{\cdot,\leq k}\Sigma_{k,k} && (V^T V = I) \\
&= XV\Sigma^T U^T U(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((ABC)^{-1} = C^{-1}B^{-1}A^{-1}) \\
&= XV\Sigma^T(\Sigma\Sigma^T)^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && (U^T U = I) \\
&= XV\Sigma^T\Sigma^{T^{-1}}\Sigma^{-1}U^T U_{\cdot,\leq k}\Sigma_{k,k} && ((AB)^{-1} = B^{-1}A^{-1}) \\
&= XV\Sigma^{-1}I_{\cdot,\leq k}\Sigma_{k,k} && (U^T U_{\cdot,\leq k} = I_{\cdot,\leq k}) \\
&= XVI_{\cdot,\leq k} && (\Sigma^{-1}I_{\cdot,\leq k} = \Sigma^{-1}_{k,k}) \\
H &= XV_{\cdot,\leq k}
\end{aligned}
$$

We will now show that H is a linear encoding and find an expression for the encoder weights W

$$
\begin{aligned}
H &= U_{.,\leq k}\Sigma_{k,k} \\
&= (XX^T)(XX^T)^{-1}U_{.,\leq K}\Sigma_{k,k} && \textit{(pre-multiplying } (XX^T)(XX^T)^{-1} = I) \\
&= (XV\Sigma^T U^T)(U\Sigma V^T V\Sigma^T U^T)^{-1}U_{.,\leq k}\Sigma_{k,k} && \textit{(using } X = U\Sigma V^T) \\
&= XV\Sigma^T U^T(U\Sigma\Sigma^T U^T)^{-1}U_{.,\leq k}\Sigma_{k,k} && (V^T V = I) \\
&= XV\Sigma^T U^T U(\Sigma\Sigma^T)^{-1}U^T U_{.,\leq k}\Sigma_{k,k} && ((ABC)^{-1} = C^{-1}B^{-1}A^{-1}) \\
&= XV\Sigma^T(\Sigma\Sigma^T)^{-1}U^T U_{.,\leq k}\Sigma_{k,k} && (U^T U = I) \\
&= XV\Sigma^T\Sigma^{T^{-1}}\Sigma^{-1}U^T U_{.,\leq k}\Sigma_{k,k} && ((AB)^{-1} = B^{-1}A^{-1}) \\
&= XV\Sigma^{-1}I_{.,\leq k}\Sigma_{k,k} && (U^T U_{.,\leq k} = I_{.,\leq k}) \\
&= XVI_{.,\leq k} && (\Sigma^{-1}I_{.,\leq k} = \Sigma^{-1}_{k,k}) \\
H &= XV_{.,\leq k}
\end{aligned}
$$

Thus H is a linear transformation of X and $W = V_{.,\leq k}$

- We have encoder $W = V_{\cdot,\leq k}$

- We have encoder $W = V_{\cdot,\leq k}$
- From SVD, we know that V is the matrix of eigen vectors of $X^T X$

- We have encoder $W = V_{.,\leq k}$
- From SVD, we know that V is the matrix of eigen vectors of $X^T X$
- From PCA, we know that P is the matrix of the eigen vectors of the covariance matrix

- We have encoder $W = V_{\cdot,\leq k}$
- From SVD, we know that V is the matrix of eigen vectors of $X^T X$
- From PCA, we know that P is the matrix of the eigen vectors of the covariance matrix
- We saw earlier that, if entries of X are normalized by

- We have encoder $W = V_{\cdot,\leq k}$
- From SVD, we know that V is the matrix of eigen vectors of $X^T X$
- From PCA, we know that P is the matrix of the eigen vectors of the covariance matrix
- We saw earlier that, if entries of X are normalized by

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right)$$

- We have encoder $W = V_{.,\leq k}$
- From SVD, we know that V is the matrix of eigen vectors of $X^T X$
- From PCA, we know that P is the matrix of the eigen vectors of the covariance matrix
- We saw earlier that, if entries of X are normalized by

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right)$$

then $X^T X$ is indeed the covariance matrix

- We have encoder $W = V_{\cdot,\leq k}$
- From SVD, we know that V is the matrix of eigen vectors of $X^T X$
- From PCA, we know that P is the matrix of the eigen vectors of the covariance matrix
- We saw earlier that, if entries of X are normalized by

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}} \left(x_{ij} - \frac{1}{m} \sum_{k=1}^{m} x_{kj} \right)$$

then $X^T X$ is indeed the covariance matrix

- Thus, the encoder matrix for linear autoencoder(W) and the projection matrix(P) for PCA could indeed be the same. Hence proved

Remember

The encoder of a linear autoencoder is equivalent to PCA if we

Remember

The encoder of a linear autoencoder is equivalent to PCA if we

- use a linear encoder

Remember

The encoder of a linear autoencoder is equivalent to PCA if we

- use a linear encoder
- use a linear decoder

Remember

The encoder of a linear autoencoder is equivalent to PCA if we

- use a linear encoder
- use a linear decoder
- use a squared error loss function

Remember

The encoder of a linear autoencoder is equivalent to PCA if we

- use a linear encoder
- use a linear decoder
- use a squared error loss function
- and normalize the inputs to

Remember

The encoder of a linear autoencoder is equivalent to PCA if we

- use a linear encoder
- use a linear decoder
- use a squared error loss function
- and normalize the inputs to

$$\hat{x}_{ij} = \frac{1}{\sqrt{m}}\left(x_{ij} - \frac{1}{m}\sum_{k=1}^{m} x_{kj}\right)$$