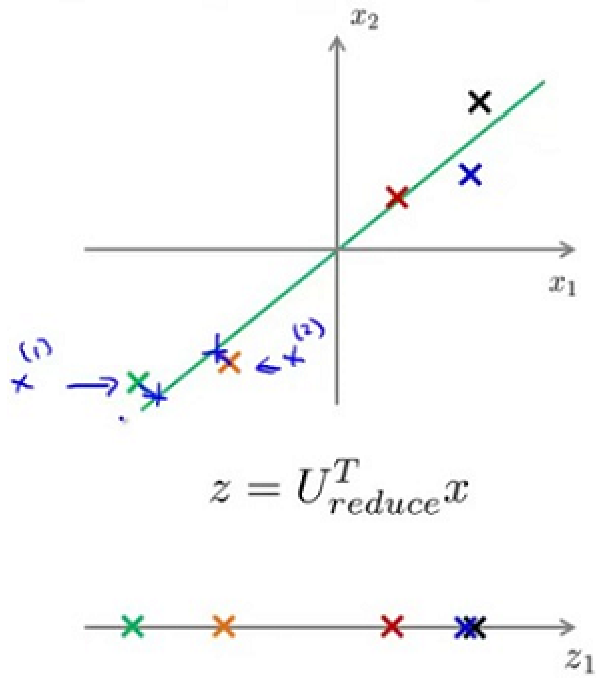
Reconstruction from Compressed Representation

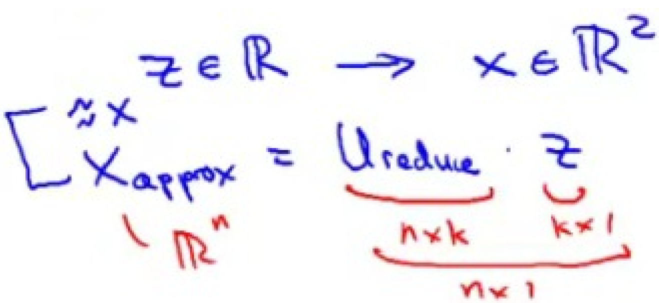
Previously we was talking about PCA as a compression algorithm where we may have say, 1000-dimensional data and compress it to 100-dimensional feature vector. So, if this is a compression algorithm, there should be a way to go back from this compressed representation back to an approximation of our original high-dimensional data.

Reconstruction of our high-dimensional data

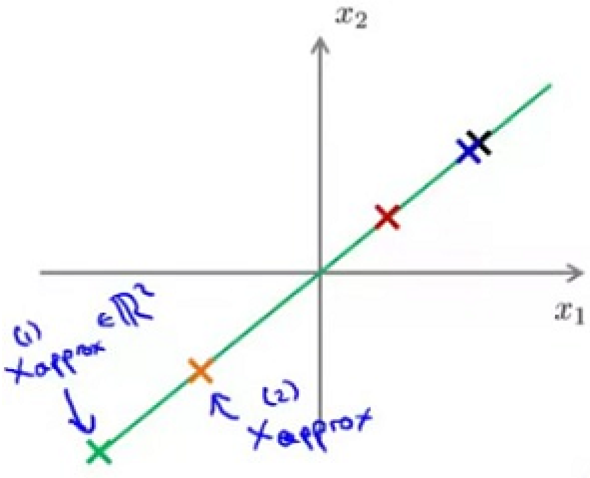
Say we have an example as follows



- We have our examples $(x^{(1)}, x^{(2)}, \dots, x^{(n)})$
  - Project onto z-surface
- Given a point $z^{(1)}$, how can we go back to the 2D space $z \in \mathbb{R} \rightarrow x \in \mathbb{R}^2$?
  - Considering: z = $(U_{\text{reduce}})^T$ * x
- To go in the opposite direction we must do:
  - $X_{\text{approx}} = U_{\text{reduce}} * z$
  - To consider dimensions:
    - U_{reduce} = $[n \ x \ k]$
    - z = $[k \ x \ 1]$
    - So X_{approx} = $[n \ x \ k] * [k \ x \ 1]$ = $[n \ x \ 1]$



So this creates the following representation:



And that's a pretty decent approximation to the original data. So that's how we go back from your low dimensional representation z, back to an uncompressed representation of the data, and we also call this process reconstruction of the original data where we think of trying to reconstruct the original value of x from the compressed representation.

We lose some of the information (i.e. everything is now perfectly on that line) but it is now projected into 2D space

Video Question: Suppose we run PCA with $k = n$, so that the dimension of the data is not reduced at all. (This is not useful in practice but is a good thought exercise). Recall that the percent / fraction of variance retained is given by: $\frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{n} S_{ii}}$. Which of the following will be true? Check all that apply.

U_{reduce} will be an $n \times n$ matrix.

$x_{\text{approx}} = x$ for every example x.

The percentage of variance retained will be 100%.

- We have that $\frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{n} S_{ii}} > 1$