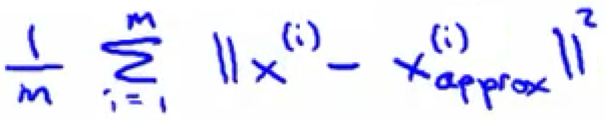
Choosing the Number of Principal Components

In the PCA algorithm we take n dimensional features and reduce them to some k dimensional feature representation, this number k is a parameter of the PCA algorithm, the number k is also called **the number of principle components** or the number of principle components that we've retained. So, how do we chose k?

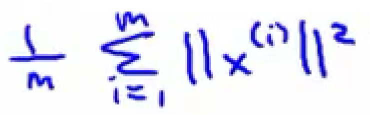
Choosing k (number of principal components)

PCA tries to minimize **averaged squared projection error**



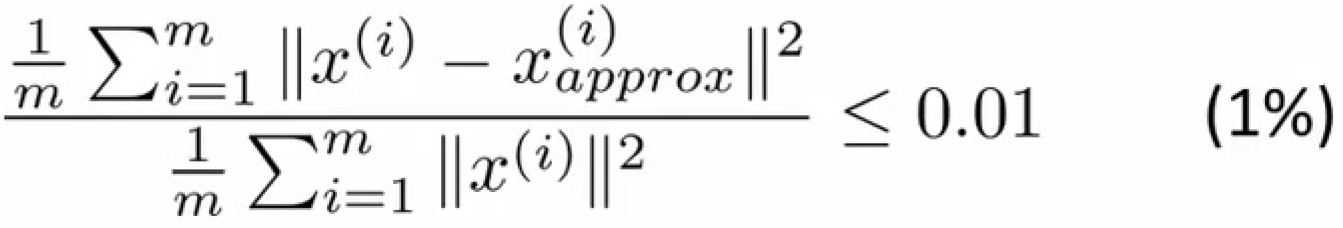
$$\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}-x^{(i)}_{approx}\|^2$$

Total variation in the data can be defined as the average over data saying how far are the training examples from the origin



$$\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}\|^2$$

When we're choosing k typical to use something like this:

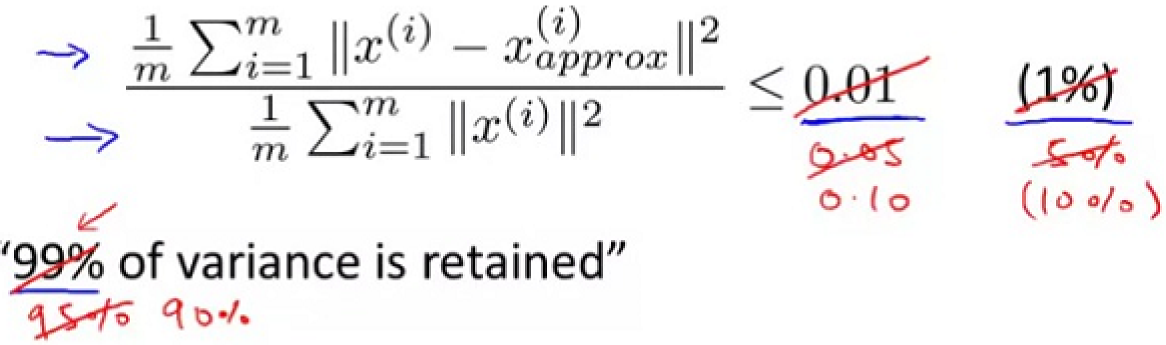


$$\frac{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}-x^{(i)}_{approx}\|^2}{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}\|^2} \leq 0.01 \qquad (1\%)$$

Tipically, choose k to be smallest value so that: **the ratio between these is less than 0.01**

Ratio between the averaged squared projection error divided by the total variation in data that was at most 1%

- Ratio between averaged squared projection error with total variation in data
  - Want ratio to be small - means **"99% of the variance is retained"**
- If it's small (0) then this is because the numerator is small
  - The numerator is small when $x^{(i)} = x^{(i)}_{\text{approx}}$



$$\begin{array}{l} \rightarrow \\ \rightarrow \end{array} \frac{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}-x^{(i)}_{approx}\|^2}{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}\|^2} \leq \underset{\substack{\cancel{0.05} \\ 0.10}}{\cancel{0.01}} \qquad \underset{\substack{\cancel{5\%} \\ (10\%)}}{\cancel{(1\%)}}$$

"~~99%~~ of variance is retained"

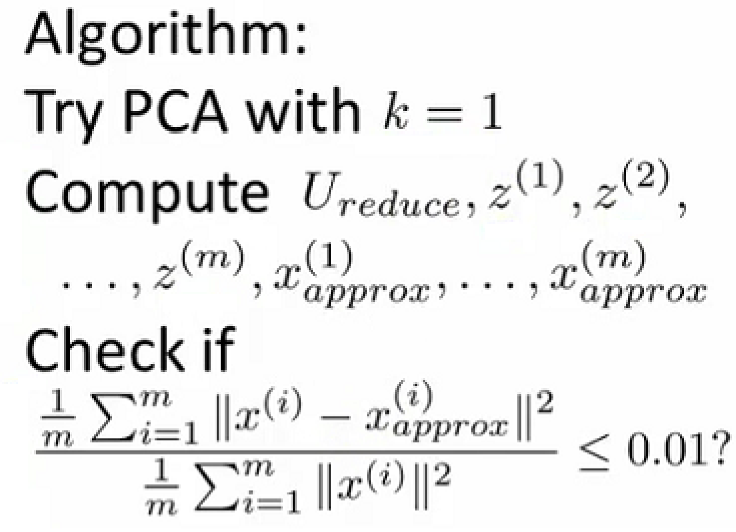
~~95%~~ 90%

In order to retain 99% of the variance, we can often reduce the dimension of the data significantly and still retain most of the variance. Because for most real life data sets many features are just highly correlated, and so it turns out to be possible to **compress the data** a lot and still retain 99% of the variance or 95% of the variance.

- So we chose k in terms of the ratio
  - Often can significantly reduce data dimensionality while retaining the variance

How we do implement this?

Here's one algorithm that we might use, if we want to choose the value of k, we might start off with k = 1 and then we run through PCA, so we compute U_{reduce}, $z^{(i)}$, $z^{(2)}$, up to $z^{(m)}$, we compute $x^{(1)}_{\text{approx}}$, $x^{(2)}_{\text{approx}}$ up to $x^{(m)}_{\text{approx}}$ and then we check if 99% of the variance is retained, si $k = 1$, no retiene el 99% de la varianza, seguimos intentando con distintos valores de k hasta encontrar un valor de k que contenga el 99% de la varianza.



Algorithm:

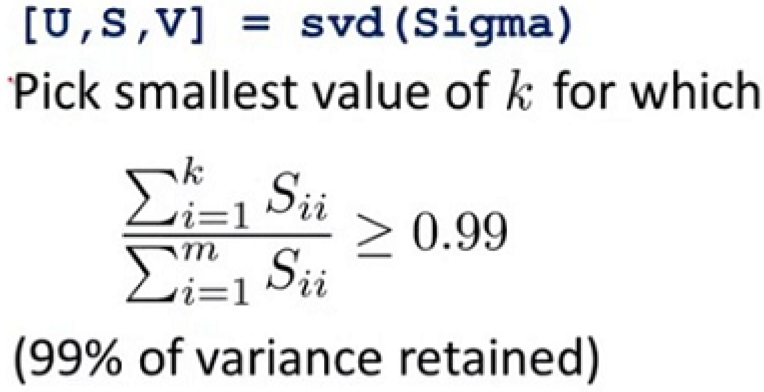
Try PCA with $k = 1$

Compute $U_{reduce}, z^{(1)}, z^{(2)}, \ldots, z^{(m)}, x^{(1)}_{approx}, \ldots, x^{(m)}_{approx}$

Check if

$$\frac{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)} - x^{(i)}_{approx}\|^2}{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}\|^2} \leq 0.01?$$

We can use svd to pick the smallest value of k



```
[U,S,V] = svd(Sigma)
```

Pick smallest value of k for which

$$\frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{m} S_{ii}} \geq 0.99$$

(99% of variance retained)

Video Question: Previously, we said that PCA chooses a direction $u^{(1)}$ (or k directions $u^{(1)}, \ldots, u^{(k)}$) onto which to project the data so as to minimize the (squared) projection error. Another way to say the same is that PCA tries to minimize:

- $\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}\|^2$
- $\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}_{\text{approx}}\|^2$

$$\boxed{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)} - x^{(i)}_{\text{approx}}\|^2}$$

- $\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)} + x^{(i)}_{\text{approx}}\|^2$