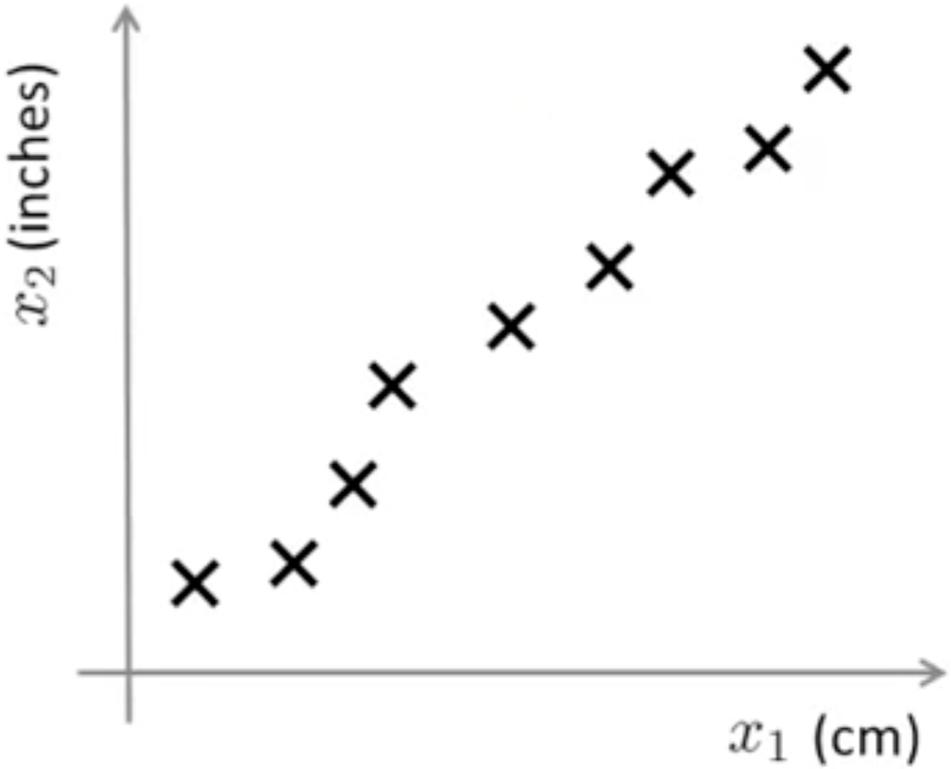
Statistical Methods in AI (CSE 471)
Lecture13: PCA

Vineet Gandhi
Centre for Visual Information Technology (CVIT)

PCA

- Useful linear transformation of the feature vector
- For dimensionality reduction we want to project the data onto a lower dimensional subspace (d dimensions to m dimensions)
- PAC is projection of data on to a subspace such that:
  - variance of projected data is maximized or
  - when approximating the original vector with its projection the MSE is minimized

Applications: data compression



Reduce the data from 2D to 1D

figure: Andrew Ng

Applications: data compression

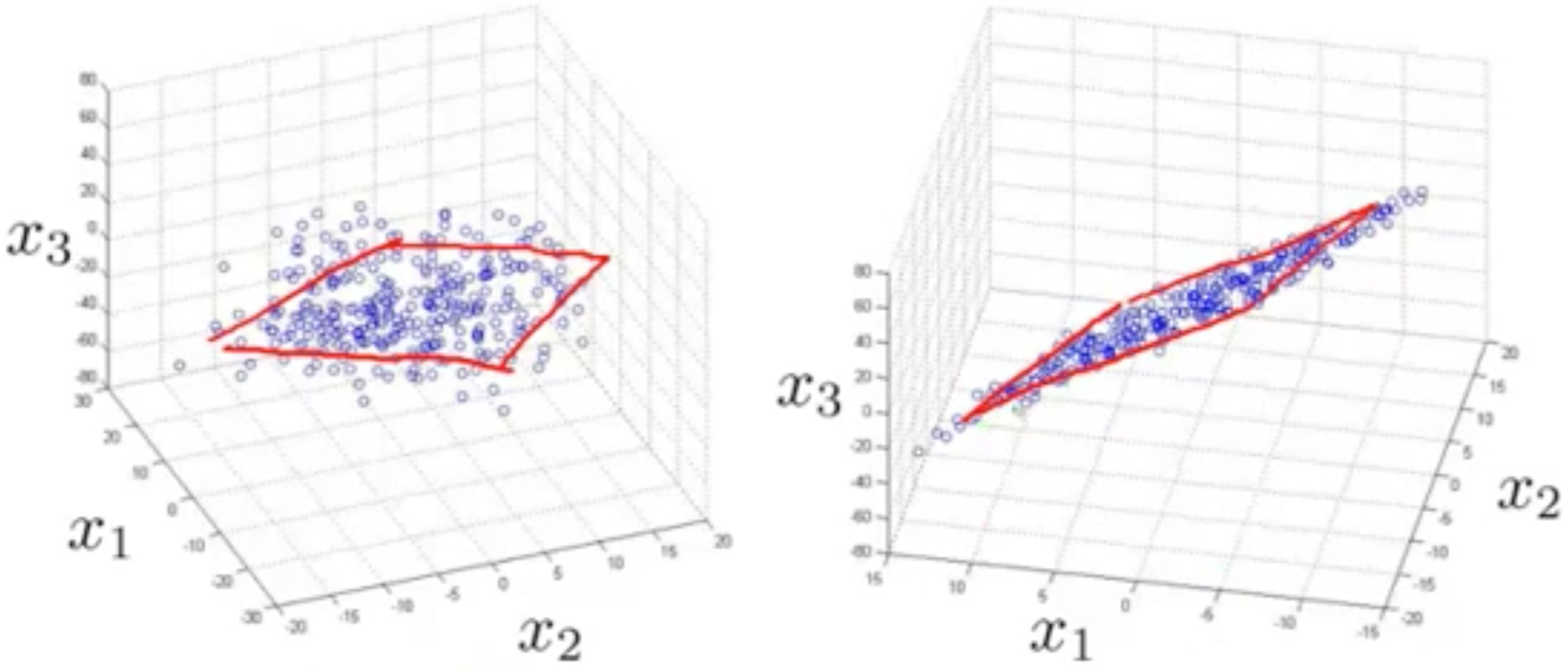


figure: Andrew Ng

Applications: visualisation

Country	GDP (trillions of US$)	Per capita GDP (thousands of intl. $)	Human Develop-ment Index	Life expectancy	Poverty Index (Gini as percentage)	Mean household income (thousands of US$)	...
Canada	1.577	39.17	0.908	80.7	32.6	67.293	...
China	5.878	7.54	0.687	73	46.9	10.22	...
India	1.632	3.41	0.547	64.7	36.8	0.735	...
Russia	1.48	19.84	0.755	65.5	39.9	0.72	...
Singapore	0.223	56.69	0.866	80	42.5	67.1	...
USA	14.527	46.86	0.91	78.3	40.8	84.3	...
...	...	...	...	...	...	...	

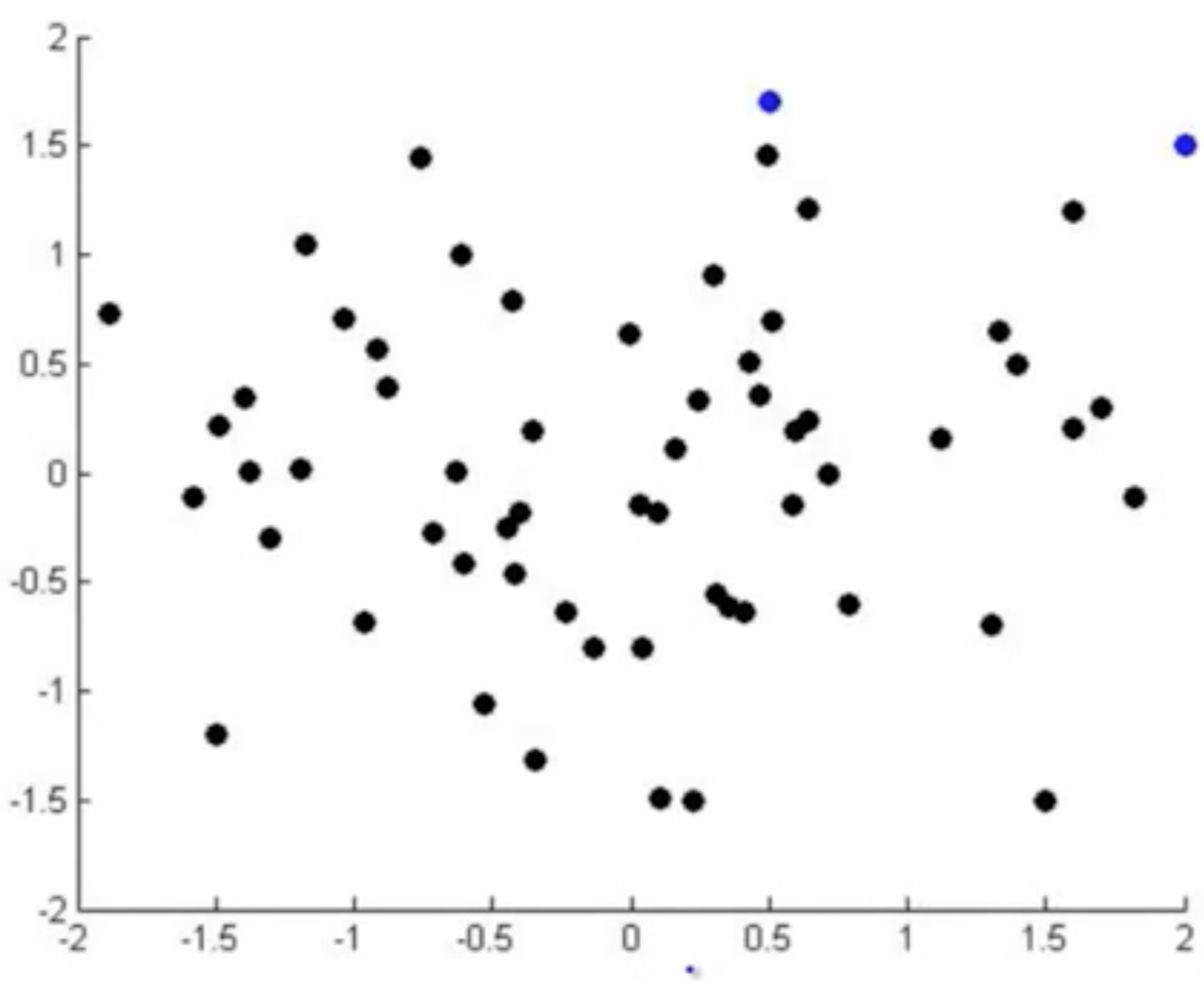
source: Andrew Ng, Wiki

Applications: visualisation

Country	z_1	z_2
Canada	1.6	1.2
China	1.7	0.3
India	1.6	0.2
Russia	1.4	0.5
Singapore	0.5	1.7
USA	2	1.5
…	…	…

source: Andrew Ng

Applications: visualisation



source: Andrew Ng

Searching the direction of maximum variance

$$X = [x_1, x_2, ...x_n]_{d \times n}$$

$$\max_{u_1} var(u_1^T X)$$ **Quadratic, no upper bound**

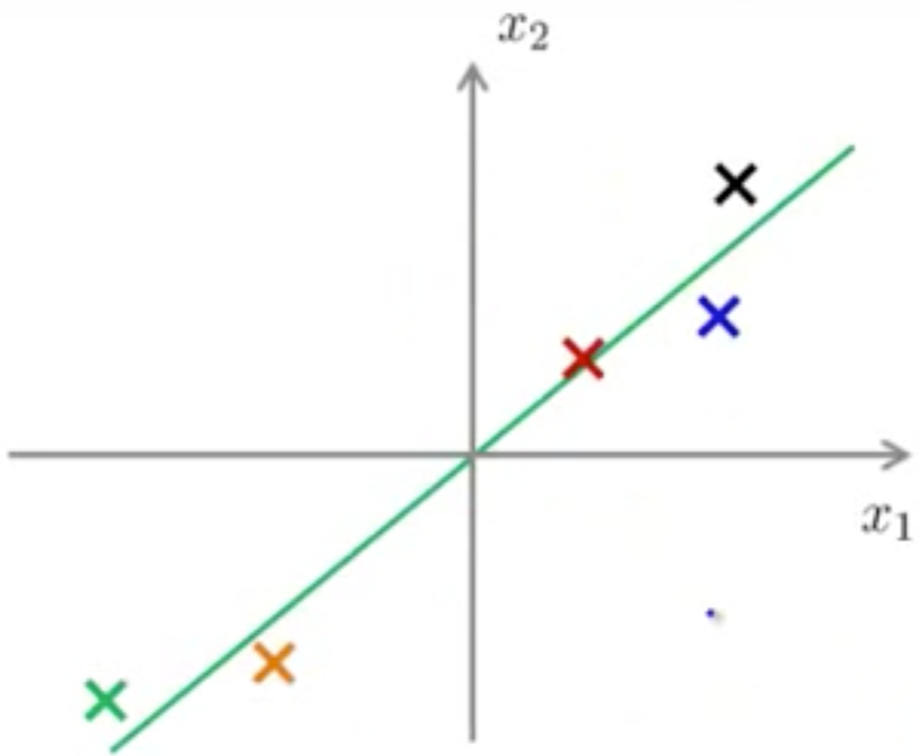
$$L(u_1, \lambda) = u_1^T S u_1 - \lambda(u^T u - 1)$$

Choosing k

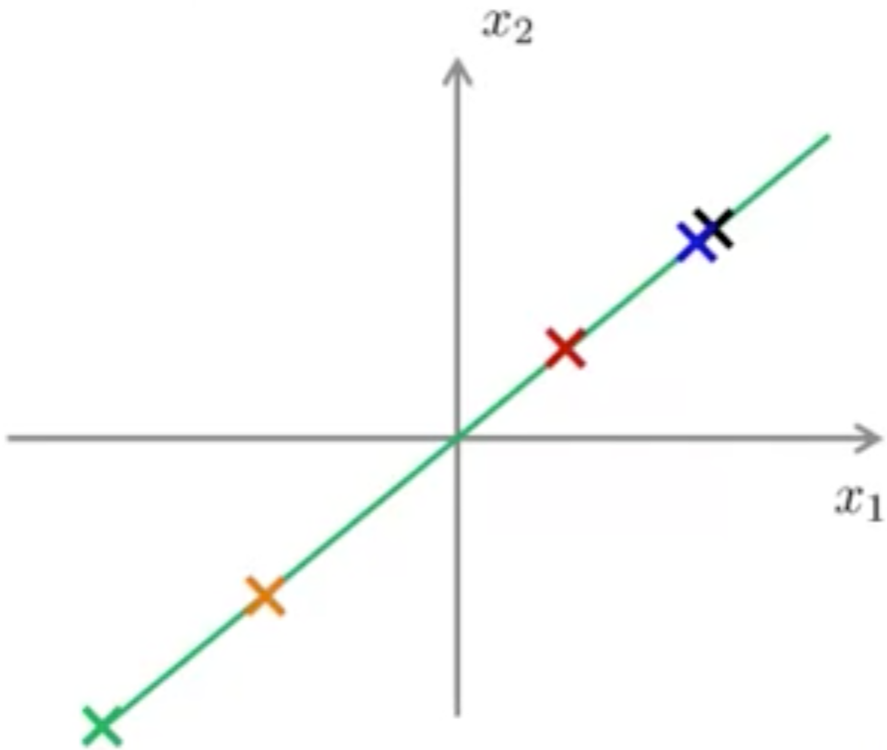
- Intrinsic and Extrinsic

```
[U,S,V] = svd(Sigma);
Ureduce = U(:,1:k);
z = Ureduce'*x;
```

Reconstruction



$z = Ureduce^T * X$



$Xapprox = Ureduce * z$