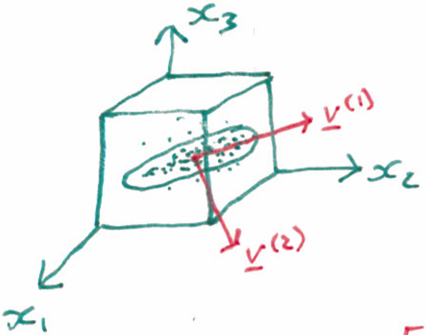
Principal Components Analysis (PCA)

- D-dimensional data → K-dimensional data
- $k=2$ (or $k=3$) → Visualization
- Fewer features to fit → Cheaper; → Less likely to overfit

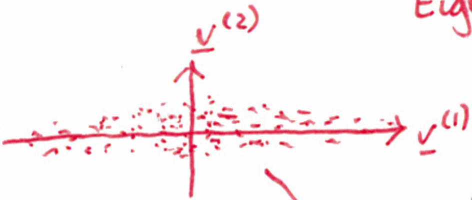
Example, D=3, k=2



$\underset{N \times D}{X}$ data matrix

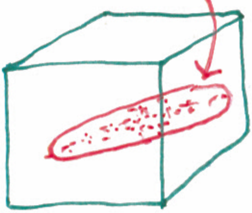
$$\underset{D \times K}{V} = \begin{bmatrix} | & | & & | \\ \underline{v}^{(1)} & \underline{v}^{(2)} & \cdots & \underline{v}^{(K)} \\ | & | & & | \end{bmatrix}$$

Eigenvectors of cov[X]



K-dim data $\underbrace{\underset{N \times D}{X} \; \underset{D \times K}{V}}_{N \times K}$

Reconstruct

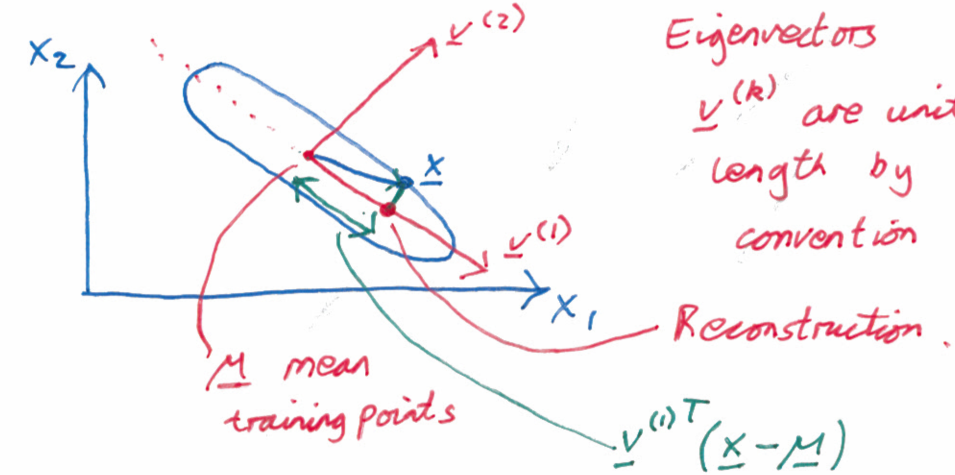


D-dimensional points, but in K-dimensional subspace

$$\hat{X} = (XV)V^T$$

Center Data

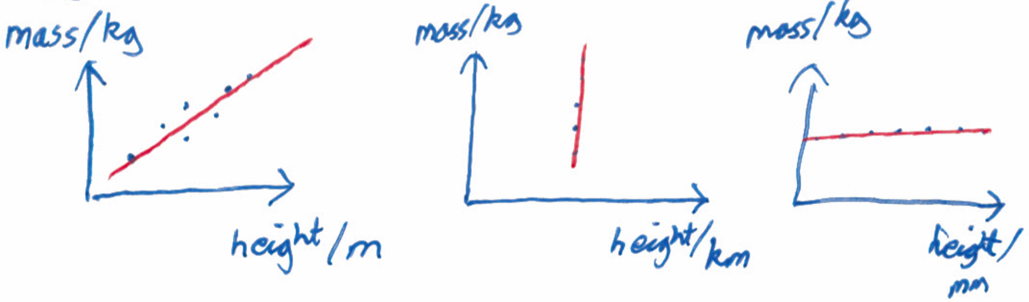
$D = 2, \quad k = 1$



Reconstruct:

$$\underline{v}^{(1)}\left[\underline{v}^{(1)T}(\underline{x}-\underline{\mu})\right] + \underline{\mu}$$

Units of Features matters



Gaussian model for PCA

Assume there is process in K-dimensional space

$$\underset{K\times 1}{\underline{h}^{(n)}} \sim N(0, \mathbb{I}_K)$$

$$\underset{D\times 1}{\underline{x}^{(n)}} = \underset{D\times K}{V}\;\underset{K\times 1}{\underline{h}^{(n)}}$$ + Gaussian noise, zero mean covariance $\sigma^2_{noise}\mathbb{I}_D$

$$\mathrm{cov}[V\underline{h}^{(n)}] = \mathbb{E}[V\underline{h}^{(n)}\underline{h}^{(n)\top}V^\top] = V\,\underbrace{\mathbb{E}[\underline{h}^{(n)}\underline{h}^{(n)\top}]}_{\mathbb{I}_K}V^\top = VV^\top$$

$$\underline{x} \sim N(\underline{0}, VV^\top + \sigma^2_{noise}\mathbb{I}_D)$$

$P(y=1\mid\bullet) = \mathrm{NaN}$

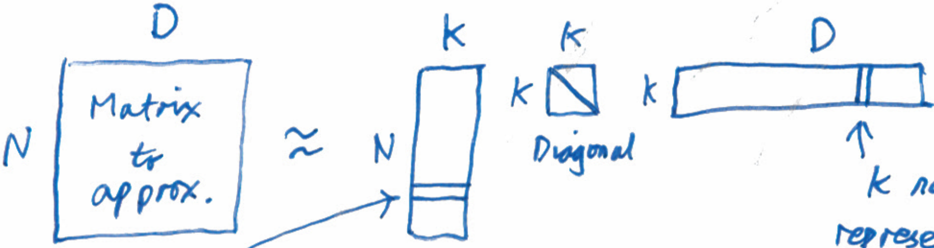
Here I fitted a mean too.

"Probabilistic PCA"

"Factor Analysis" – Noise has arbitrary diagonal covariance

Truncated SVD

SVD is some standard linear algebra method



k numbers representing the feature

$$X \approx \underbrace{U \quad S}_{\text{Projected data}} \quad V^T$$

V same as in PCA (If X centered)

k numbers representing a data item

http://netflixprize.com/

Truncated SVD

```
% PCA via SVD,
% for zero-mean X:
[U, S, V] = svd(X, 0);
U = U(:, 1:K);
S = S(1:K, 1:K);
V = V(:, 1:K);
X_kdim = U*S;
X_proj = U*S*V';
```

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1D} \\ X_{21} & X_{22} & \cdots & X_{2D} \\ X_{31} & X_{32} & \cdots & X_{3D} \\ X_{41} & \boxed{X_{42}} & \cdots & X_{4D} \\ X_{51} & X_{52} & \cdots & X_{5D} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{ND} \end{bmatrix} \approx$$

$$\begin{bmatrix} U_{11} & \cdots & U_{1K} \\ U_{21} & \cdots & U_{2K} \\ U_{31} & \cdots & U_{3K} \\ \boxed{U_{41}} & \boxed{\cdots} & \boxed{U_{4K}} \\ U_{51} & \cdots & U_{5K} \\ \vdots & \ddots & \vdots \\ U_{N1} & \cdots & U_{NK} \end{bmatrix} \begin{bmatrix} \boxed{S_{11}} & 0 & 0 \\ 0 & \boxed{\ddots} & 0 \\ 0 & 0 & \boxed{S_{KK}} \end{bmatrix} \begin{bmatrix} V_{11} & \boxed{V_{21}} & \cdots & V_{D1} \\ \vdots & \boxed{\vdots} & \ddots & \vdots \\ V_{1K} & \boxed{V_{2K}} & \cdots & V_{DK} \end{bmatrix}$$

$$X \approx U \quad S \quad V^\top$$