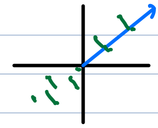
PCA - dimensional Reduction

maximize variance



1.) find a unit vector $w_{(1)}$, such that the projections of data samples on this vector have the largest variance

$$w_{(1)} = \underset{\|w\|=1}{\arg\max} \left\{ \sum_{i=1} (x_i \cdot w)^2 \right\} = \underset{\|w\|=1}{\arg\max} \left\{ \|Xw\|^2 \right\} = \underset{\|w\|=1}{\arg\max} \left\{ w^T X^T X w \right\}$$

$\rightarrow$ max $E[Y] = 0$

$$y_i = x_i \cdot w \qquad E[Y^2] - E[Y]^2 \approx 0$$

$$\underset{max}{=} \left(\sum_i y_i^2 \right)$$

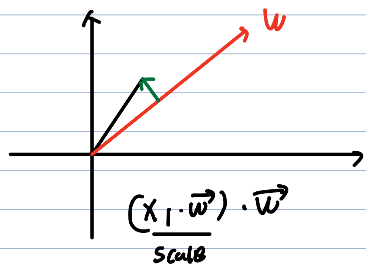
$D \qquad U\Lambda V^T \qquad V\Lambda^2 V^T$

Rayleigh quotient.

$$w_{(1)} = \underset{\|w\|=1}{\arg\max} \{ w^T X^T X w \} = \underset{\|w\|=1}{\arg\max} \left\{ \frac{w^T X^T X w}{w^T w} \right\}$$

since $X^T \cdot X$ is a positive semidefinite matrix, the max value of it is the largest eigenvalue of the matrix

$$= \frac{X^T V \Lambda^2 V^T X}{X^T X}$$



$$X_k = X - \sum_{s=1}^{k-1} \boxed{X w_{(s)} w_{(s)}^T}$$

project matrix