$X$    $N$ rows, $k$ columns

rows $\Leftrightarrow$ samples    columns $\Leftrightarrow$ features

assume that our data is centered

columns sum to zero

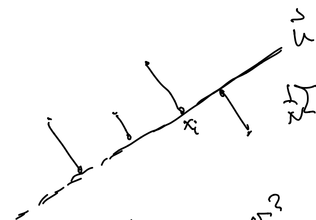
each row of $X$ gives a point in $\mathbb{R}^k$

$N$ points are a cloud of points in $\mathbb{R}^k$

$u$ is a unit vector in $\mathbb{R}^k$

$\sigma_u^2 = u^T D_0 u$

$D_0 = \frac{1}{N} X^T X$ covariance matrix.



$\frac{1}{N}\sum x_i^2 = \sigma_u^2$

How to find $u$ where $\sigma_u^2$ is maximum.

here $\|u\|^2 = 1$.

Constrained optimization. Lagrange multipliers.

$F(x_1, \ldots, x_k)$ your "objective" function: to be maximized

$g(x_1, \ldots, x_k) = 0$ your "constraint".

$S(x, \lambda) = F - \lambda g$.

$\frac{\partial S}{\partial x_i} = 0$ and $\frac{\partial S}{\partial \lambda} = 0$ which is the constraint $g = 0$.

$D_0 = (d_{ij})$

~~Constr~~ Objective: $\sigma_u^2 = u^T D_0 u$

$$[u_1 \ldots u_k]\, D_0 \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix}$$

$$D_0 \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{k} d_{ij} u_j \\ \\ \\ \end{bmatrix}$$

$$\sigma_u^2 = [u_1 \dots u_k] \begin{bmatrix} \sum_{j=1}^{k} d_{ij} u_j \\ \\ \\ \end{bmatrix}_{i=1,\dots,k} = \sum_{r=1}^{k} \sum_{s=1}^{k} u_r d_{rs} u_s$$

Constraint: $\|u\|^2 = 1$ $\quad u_1^2 + \dots + u_k^2 = 1.$

$$g(u_1,\dots,u_k) = u_1^2 + \dots + u_k^2 - 1$$

$$S(u_1,\dots,u_k,\lambda) = \sum_{r=1}^{k} \sum_{s=1}^{k} u_r d_{rs} u_s - \underline{\lambda(u_1^2 + \dots + u_k^2 - 1)}$$

$$\frac{\partial S}{\partial u_i} = \sum_{\substack{s=1 \\ s \neq i}}^{k} \underline{d_{is}} u_s + \sum_{\substack{r=1 \\ r \neq i}}^{k} \underline{d_{ri}} u_i + 2 u_{ii} d_{ii} - \lambda 2 u_i$$

$D_0$ is symmetric

$$= 2 \sum_{\substack{s=1 \\ s \neq i}}^{k} d_{is} u_s + \underline{2 u_{ii} d_{ii}} \qquad d_{ij} = d_{ji} \qquad - 2\lambda u_i$$

$$= 2 \sum_{s=1}^{k} d_{is} u_s - 2\lambda u_i \qquad \nabla\, u^T D_0 u \sim 2 D_0 u$$

$$\nabla_u S = 2 \begin{bmatrix} \sum_{s=1}^{k} d_{1s} u_s \\ \sum_{s=1}^{k} d_{2s} u_s \\ \vdots \end{bmatrix} - 2\lambda \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} = 2(D_0 - \lambda) u = 0.$$

$$\frac{\partial S}{\partial \lambda} = u_1^2 + \dots + u_k^2 - 1 = 0 \qquad \|u\|^2 = 1$$

1) $(D_0 - \lambda) u = 0$ OR $D_0 u = \lambda u$

u is an eigenvector for $D_0$ with eigenvalue $\lambda$

2) $\|u\|^2 = 1$

Theorem: The critical points of $\sigma_u^2 = u^T D_0 u$ such that $\|u\|^2 = 1$ occur when $u$ is an eigenvector for $D_0$ with eigenvalue $\lambda$ and $\|u\|^2 = 1$.

The value $\sigma_u^2 = u^T D_0 u = u^T \lambda u = \lambda \|u\|^2 = \lambda$.

The maximum variance is the largest eigenvalue of $D_0$.

The minimum variance " " smallest eigenvalue of $D_0$.

[Fact: $D_0$ is symmetric so it has real eigenvalues]