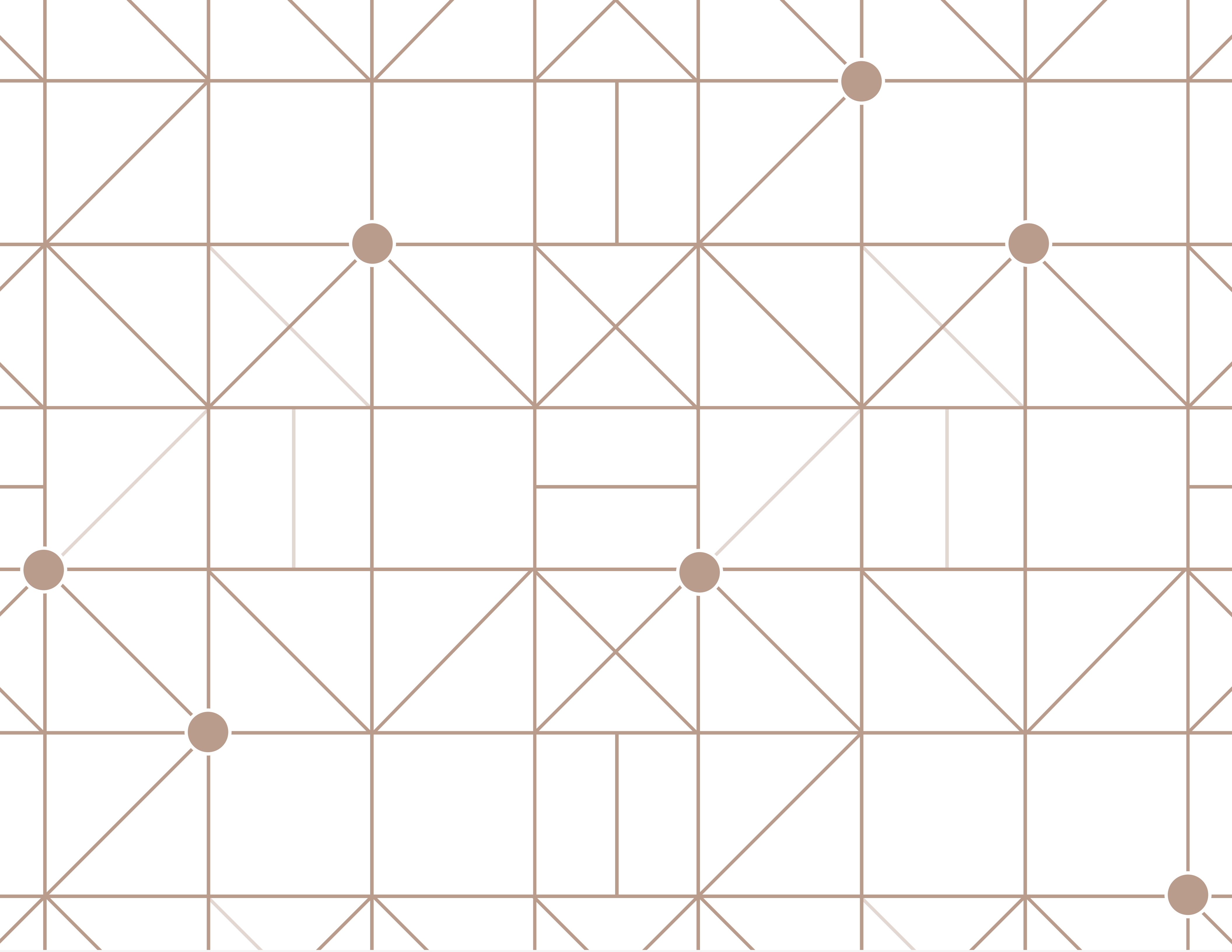


Datapoints (vectors)

	f_1	f_2	...	f_d
x_1				
x_2				
x_3				
⋮				
x_N				

new representation on t features

A: $\boxed{t \ll d}$

$N \times d \longrightarrow N \times t$

each new feature g_t is a linear comb. of the original features

B: new data matrix

$$\begin{array}{c|c} & t \text{ feat} \\ x_1 & \\ \vdots & \\ x_N & \end{array} \sim X_{orig}$$

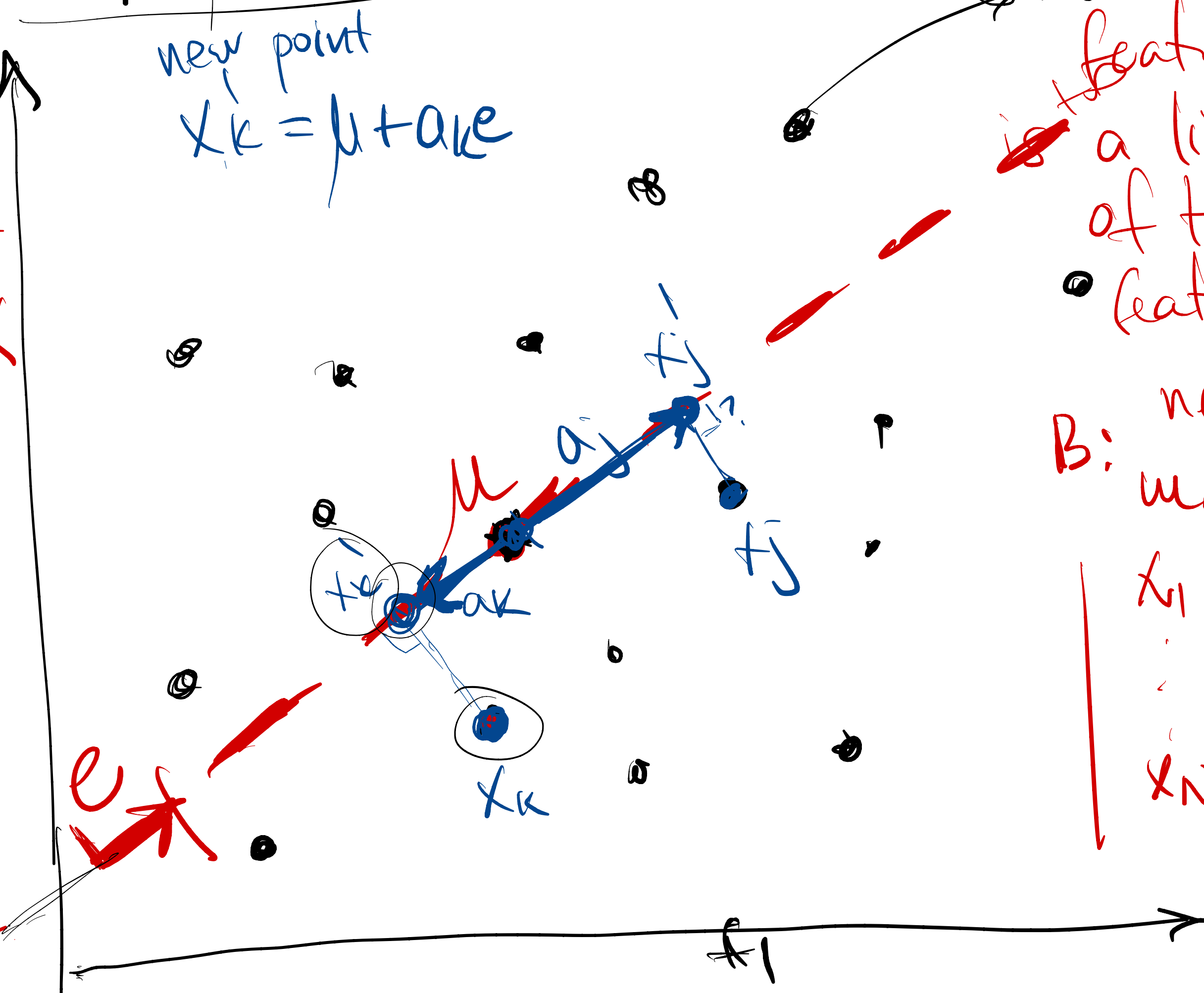
d=2 visualization

① t=0 NOFEAT

→ represent whole X on 1 point

$\mu = \text{mean}(X) = E[X]$

$|e|=1$ direction of line



$t=1 \Rightarrow 1$ new dim $\Rightarrow$ new representation will be on 1 line

guess: that "one line" now rep

- passes through μ
- visually correspond to longest-direction-stretch of data

geom line: $e, \underbrace{a_1 a_2 \dots a_N}_{\text{new coordinates}}$

SSE / SQ LOSS

error $J(a_1, a_2, \dots a_N, e) = \sum_{k=1}^{N} \| \underbrace{\mu + a_k e}_{\text{new point}} - \underbrace{x_k}_{\text{ori's point}} \|^2 = \sum_k \| a_k e - (x_k - \mu) \|^2$

$x_k = \mu + a_k e$

$x_k - \mu = a_k \cdot e$

$(x_k - \mu) e = a_k$

$$= \boxed{\sum_k a_k^2 \|e\|^2 - 2 \sum_k a_k e^T (x_k - \mu) + \sum_k \| x_k - \mu \|^2}$$

$J = \min \Rightarrow \frac{\partial J}{\partial a_k} = 0 \Leftrightarrow a_k - e^T (x_k - \mu) = 0 \Rightarrow a_k = e^T (x_k - \mu)$ projection

new data mean $(a_1, a_2, \dots a_N) = E[\mu + a_k e] = \mu + E[a_k e] = \mu + e^T E[x_k - \mu] e = \mu$

$E[x_k - \mu] = 0$

$$J\left(e \quad \text{implicit projections } x \to e = a\right) = \sum_k a_k^2 - 2\sum_k a_k^2 + \sum_k \|x_k - \mu\|^2$$

$a_k \cdot e^T (x_k - \mu)$

a_k

$$= -\sum_k \underbrace{\left(e^T (x_k - \mu)\right)^2}_{a_k^2} + \sum_k \|x_k - \mu\|^2$$

$$= \boxed{-\sum_k e^T (x_k - \mu)(x_k - \mu)^T e} + \cancel{\sum_k \|x_k - \mu\|^2}$$

$$= -e^T \left[\sum_k (x_k - \mu)(x_k - \mu)^T\right] e \quad \to \text{covar matrix } \Sigma$$

$$-e^T \Sigma e$$

sigma

minimize $J(e)$

$\Updownarrow$

$\min -e^T \Sigma e \;/\; \text{MAX } e^T \Sigma e$

$\Updownarrow$

max variance (projections)

MOST IMPORTANT: what is the best e? want MAX

MAX $\mathrm{Var}[\text{projections} = \text{new representations}]$

$$E\Big[\big(\underbrace{\mu + a_k e}_{\text{proj}} - E[\mu + a_k e]\big)^2\Big] = E[(\mu + a_k e - \mu)^2] = E[(a_k e)^2]$$

$$= E\big[\overbrace{e^T(x_k - \mu)}^{a_k e} \cdot e^T(x_k - \mu)\big] = E\big[e^T (x_k - \mu)(x_k - \mu)^T e\big]$$

$$= \underset{1\times d}{e^T}\ \underset{(d\times d)}{\Sigma}\ \underset{d\times 1}{e} \quad \text{MAX}$$

sigma covar matrix

$$\underbrace{\Sigma}_{\text{sigma}} = \underset{\text{sum of matrices}}{\sum_k} \underbrace{(x_k - \mu)}_{d\times 1} \underbrace{(x_k - \mu)^T}_{1\times d} =$$

$$\begin{array}{c|ccc}
 & & j & \\
\hline
 & \sum_k (x_k^1 - \mu^1)(x_k^1 - \mu^1) & & \sum_k (x_k^1 - \mu^1)(x_k^d - \mu^d) \\
i & & & \\
 & \sum_k (x_k^d - \mu^d)(x_k^1 - \mu^1) & & \sum_k (x_k^d - \mu^d)(x_k^d - \mu^d)
\end{array}$$

$\sigma_{ii} = \sum_k (x_k^i - \mu^i)(x_k^i - \mu^i) = \mathrm{var}(\text{feat } i)$

$\boxed{\sigma_{ij} = \sum_k (x_k^i - \mu^i)(x_k^j - \mu^j)} = \sigma_{ji}$

linear correl (feat i, feat j)

Constrained optimization pb Σ_i = ~~cor~~ cov matrix = fixed

$$\underset{e}{\text{maximize}} \quad e^T \Sigma_i e$$

subject to $\|e\| = 1 \iff e^T e = 1$

Lagrangian

Lag Multip

OBJ CONSTRAINED

$$L = \max\left[\underbrace{e^T \Sigma_i e}_{\text{OBJ}} - \alpha \underbrace{(e^T e - 1)}_{\text{CONSTRAINED}}\right]$$

$$\frac{\partial L}{\partial e} = 0 \iff 2\Sigma_i e - 2\alpha e = 0$$

same direction as e

$$\boxed{\Sigma_i e = \alpha e}$$

sigma scalar

$\parallel$ e does not change direction when multip by Σ_i !!

d×d d×1

$$\boxed{M \circ V}$$

matrix vector

usually changes direction of the vector

$M \cdot v$ = direction of v $\Longleftrightarrow$ v = eigen vector for M

M = fixed

λv

scalar

λ = eigen value

$\Sigma = \mathrm{covar}(X)$ (1) symmetric $\Sigma_{ij} = \Sigma_{ji}$ very special

$\boxed{\text{(2) pos def} \quad v^T \Sigma v \geq 0}$

$\Downarrow$

allows spectral decomposition

e_i = eigen vectors of Σ
normalized

$$\Sigma = \underbrace{\begin{bmatrix} | & | & & | \\ e_1 & e_2 & \cdots & e_d \\ | & | & & | \end{bmatrix}}_{\text{eigenvect}^T} \underbrace{\begin{bmatrix} \alpha_1 & & & 0 \\ & \alpha_2 & & \\ & & \ddots & \\ 0 & & & \alpha_d \end{bmatrix}}_{\text{diag}(\text{eigen values})} \underbrace{\begin{bmatrix} - & e_1 & - \\ - & e_2 & - \\ & \vdots & \\ - & e_d & - \end{bmatrix}}_{\text{eigenvect}}$$

$\alpha_1 \geq \alpha_2 \geq \cdots \alpha_d \geq 0$

WANT t dim $\Longrightarrow$ take top t eigen values

make the rest 0