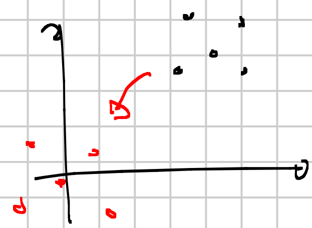
Principal component analysis

Note Title 2024-10-25

$x_1, x_2, \ldots x_n \in \mathbb{R}^m$ $\quad A = [x_1 | x_2 | x_3 \cdots | x_n] \in \mathbb{R}^{m \times n}$

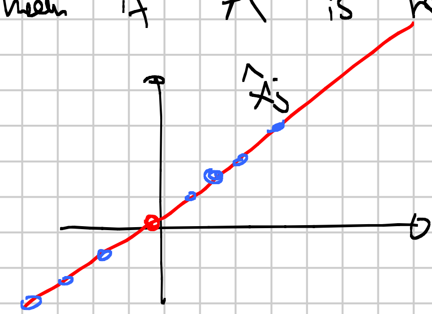
$\mu = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$ $\quad$ "De-meaned data" $\hat{x}_j = x_j - \mu \quad j = 1, 2, \ldots n$



$$\hat{A} = [\hat{x}_1 | \hat{x}_2 \cdots | \hat{x}_n] = USV^T$$

(In some sources: $U\left(\frac{1}{n-1} SS^T\right)U^T$ is an eigendecomposition of $\frac{1}{n-1}\hat{A}\hat{A}^T$ "covariance matrix", less accurate).

What does it mean if $\hat{A}$ is rank-1?

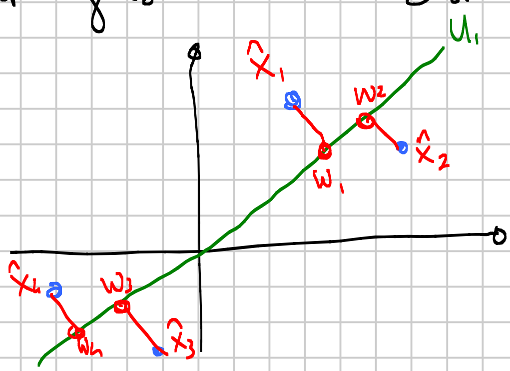


All $\hat{x}_j$ multiple of the same vector $\Rightarrow$ all data points $\hat{x}_j$ are aligned on one line through the origin

(Similarly, rank $k \Leftrightarrow$ data $\hat{x}_j$ all on a k-dimensional hyperplane).

So, in general, I can approximate $\hat{A}$ with $X_1 = u_1 \sigma_1 v_1^T$

and that gives me the best rank-1 approximation of the $\hat{x}_j$



u_1 gives minimum

$$\sum \|\hat{x}_j - w_j\|^2$$

Dimensionality reduction: if data have large dimension m, I can construct the "most informative" plot on k dimensions by SVD: $\min \sum \|\hat{x}_j - w_j\|^2$.

$\boxed{1D}$ $w_j = u_1 \alpha_j$ $\quad \alpha_j = u_1^T \hat{x}_j \quad [\alpha_1\ \alpha_2 \dots \alpha_n] = u_1^T \hat{A}$.

(coordinates or scores)

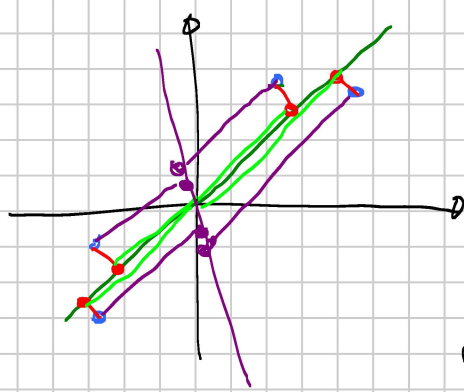
$\boxed{2D}$ $w_j = u_1 \alpha_j + u_2 \beta_j$ $\quad [\alpha_1 \dots \alpha_n] = u_1^T \hat{A}$

$[\beta_1 \dots \beta_n] = u_2^T \hat{A}$

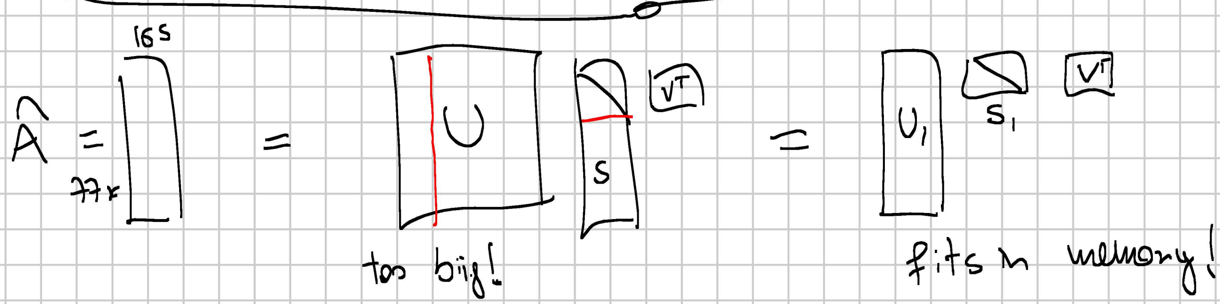
Statistical property:

Among all plots that are obtained by projecting the $\hat{x}_j$ on a line (plane, hyperplane of fixed dimension), w_j is the one that has maximum variance $\sum w_j w_j^T$

(or, without de-meaning, $\sum (w-\mu)(w-\mu)^T$)



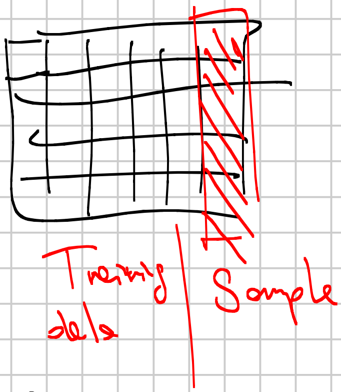
the variance of the red points (projections on the optimal line) is larger than the variance of the projection on any other line (e.g. purple points)



$$x_1 = \mu + u_1\alpha + u_2\beta + \dots + u_n\delta$$

ϕ ϕ

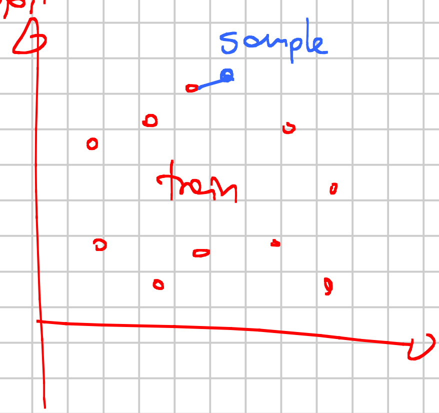
Image recognition example



Each face can be represented as

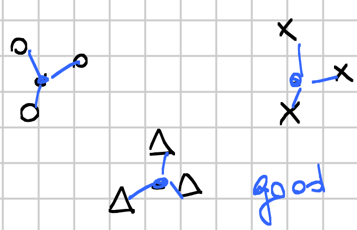
$$x_j = \mu + u_1\alpha + u_2\beta + \underbrace{u_3\gamma + \dots}_{3,4,5,6,7} u_n\delta.$$

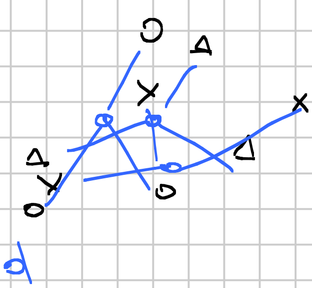
project train + sample on this 5-dim space



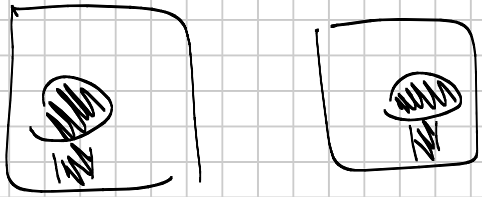
$$\text{distances} = \left[\, \|\text{scores}(:,1) - \text{samplescores}\|^2, \;\dots\; \|\text{scores}(:,n) - \text{samplescores}\|^2 \,\right]$$

Lots of limitations!

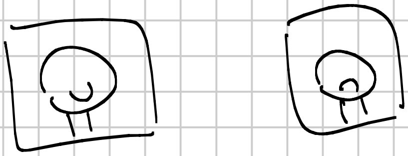




- based <u>only</u> on Frobenius distance between images



very far apart, cannot tell they are the same individual

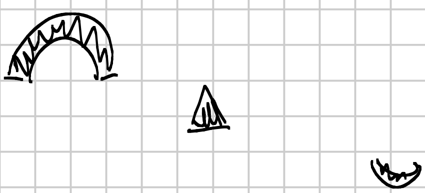


are very close, cannot tell expressions apart

- tries to flatten out information to 2 dimensions, possibly losing info from other axes.
- The SVD cannot be generalised easily / meaningfully to more than 2 dimensions

Even computing the rank of a 3-dim tensor is NP-Hard!

- Cannot add constraints! E.g. nonnegativity: one could want to recognize components that are positive instead of orthogonals



$$\min_{\substack{rk\, X \leq k \\ X \text{ positive}}} \|\hat{A} - X\|$$

Or change norms: $\min \|A - X\|_\infty$

All these give optimization problems that are much harder to solve.