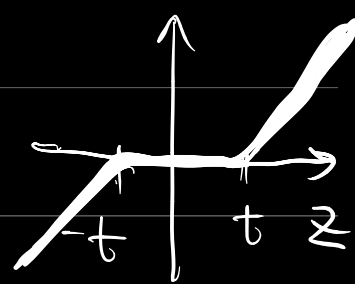
Putting together:

$$\beta_j^{lasso} = \begin{cases} \beta_j^{ls} - \frac{\lambda}{2} & \beta_j^{ls} - \frac{\lambda}{2} > 0 \\ 0 & \frac{\lambda}{2} \geq \beta_j^{ls} \geq -\frac{\lambda}{2} \\ \beta_j^{ls} + \frac{\lambda}{2} & \beta_j^{ls} + \frac{\lambda}{2} < 0 \end{cases}$$

Introduce the <u>soft thresholding function</u>

$$S_t(z) \overset{def}{=} \operatorname{sgn}(z)(|z| - t)_+ = \begin{cases} z - t & z > t \\ 0 & t \geq z \geq -t \\ z + t & -t > z \end{cases}$$

$(t > 0)$



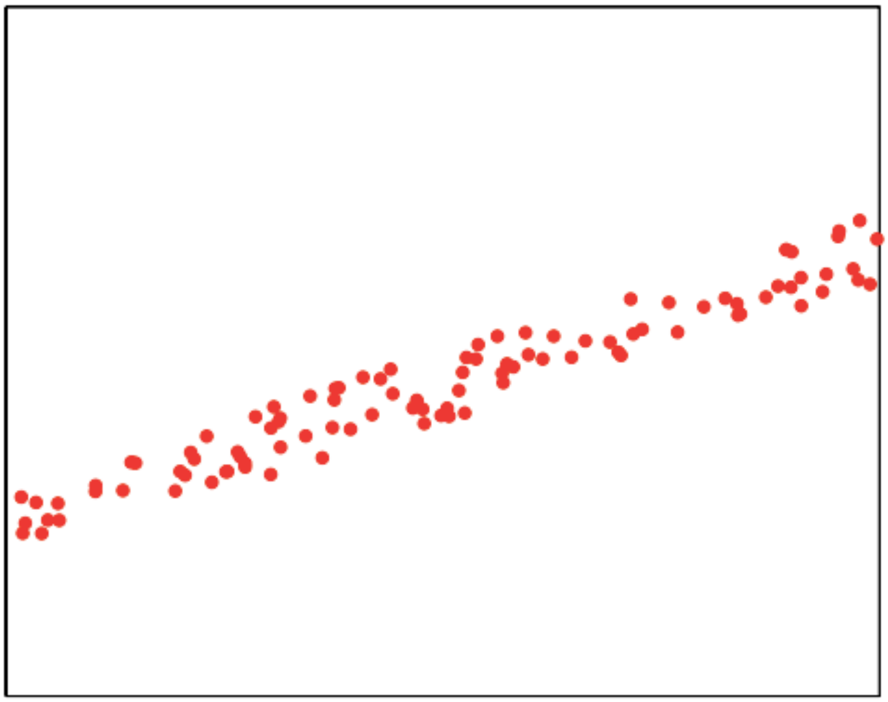
where $\operatorname{sgn}(z) \overset{def}{=} \begin{cases} 1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases}$

$$y_+ \overset{def}{=} \begin{cases} y & y \geq 0 \\ 0 & y < 0 \end{cases}$$

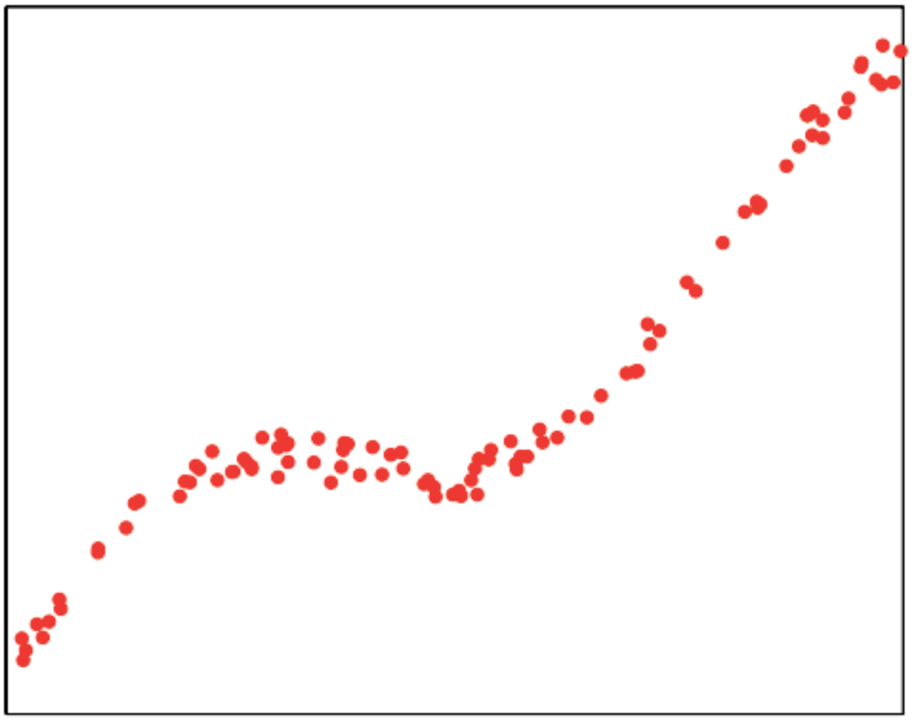
Then $\beta_j^{lasso} = S_{\frac{\lambda}{2}}(\beta_j^{ls})$

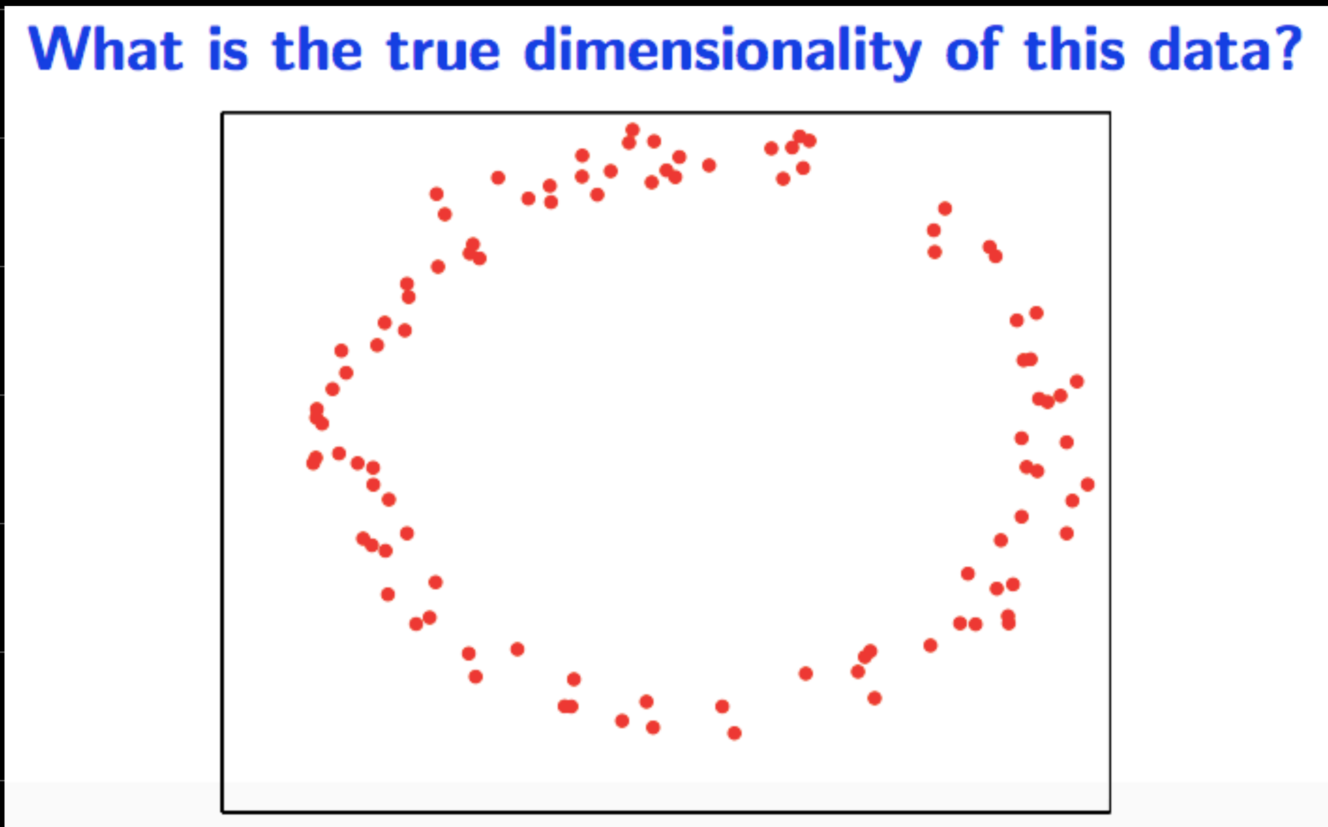
3 Dimension Reduction.

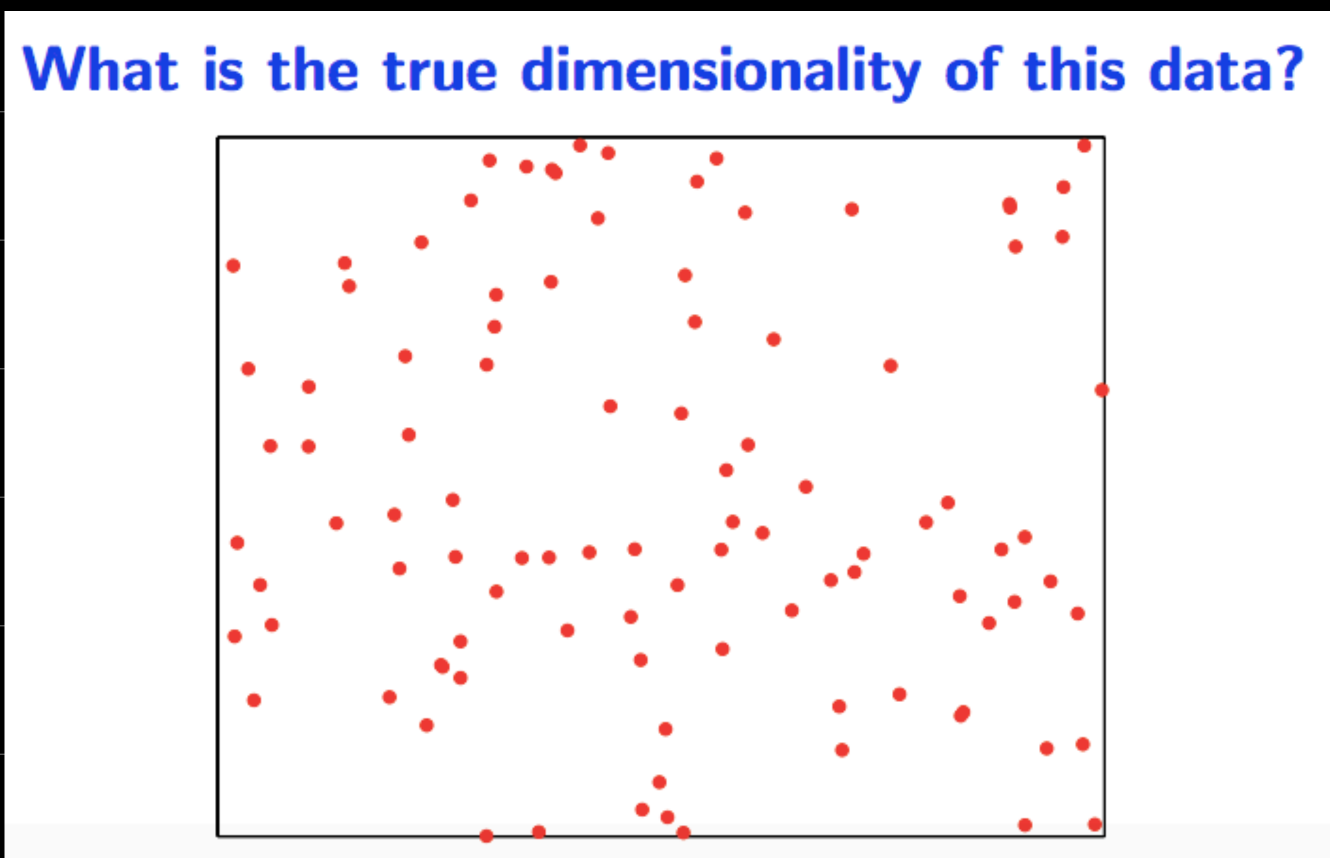
What is the true dimensionality of this data?



What is the true dimensionality of this data?







Goal: find "the most important" features for a task to reduce the number of features.

3.1 Principal Component Analysis (PCA)

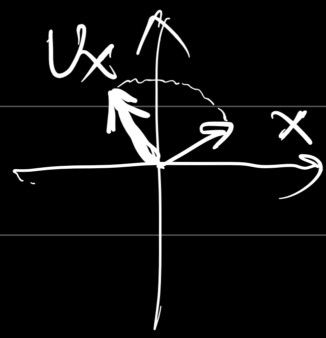
Goal of PCA: find low-dimensional representation of the sample points that maximizes the data variance.

3.1.2 Math Prep.

Recall that an orthogonal matrix $U_{n\times n}$ $\left(UU^T = U^TU = I\right)$ preserves the Euclidean norm of a vector,

i.e. $\|Ux\| = \|x\| \quad \forall x \in \mathbb{R}^n$

(because $\|Ux\|^2 = (Ux)\cdot(Ux)$
$= (Ux)^T Ux = x^T \underbrace{U^TU}_{=I} x$
$= x^Tx = \|x\|^2$)



Def: For a symmetric matrix A and a non-zero vector u, the ratio

$$\frac{u^T A u}{u^T u} = \frac{u^T A u}{\|u\|^2}$$

is called the Rayleigh quotient.

Let $A \in \mathbb{R}^{n \times n}$ be symmetric with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and associated normalized eigenvectors $u^{(1)}, \ldots, u^{(n)}$. Then the eigenvalue decomp

$$A = U \Lambda U^T = \begin{pmatrix} | & & | \\ u^{(1)} & \cdots & u^{(n)} \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} | & & | \\ u^{(1)} & \cdots & u^{(n)} \\ | & & | \end{pmatrix}^T$$

Remark: Notice $\frac{u^T A u}{u^T u} = \frac{u^T A u}{\|u\|^2} = \left(\frac{u}{\|u\|}\right)^T A \left(\frac{u}{\|u\|}\right)$

so it suffices to consider the Rayleigh quotient for unit vectors.

Prop.
$$\lambda_n \leq \frac{u^T A u}{u^T u} \leq \lambda_1$$

Each "=" holds if and only if u is the associated eigenvector.