Mathematical Tools
for Neural and Cognitive Science

Fall semester, 2017

Section 2: Least Squares

Least Squares (outline)

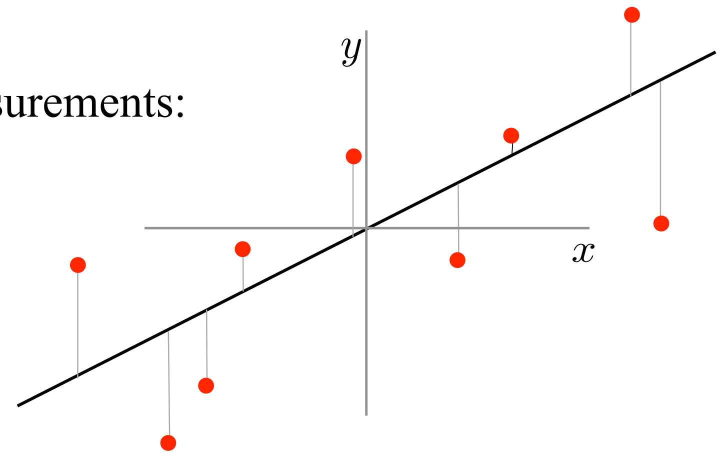
- Standard regression: Fit data with weighted sum of regressors. Solution via calculus, orthogonality, SVD
- Choosing regressors, overfitting
- Robustness: weighted regression, iterative outlier trimming, robust error functions, iterative re-weighting
- Constrained regression: linear, quadratic constraints
- Total Least Squares (TLS) regression, and Principle Components Analysis (PCA)

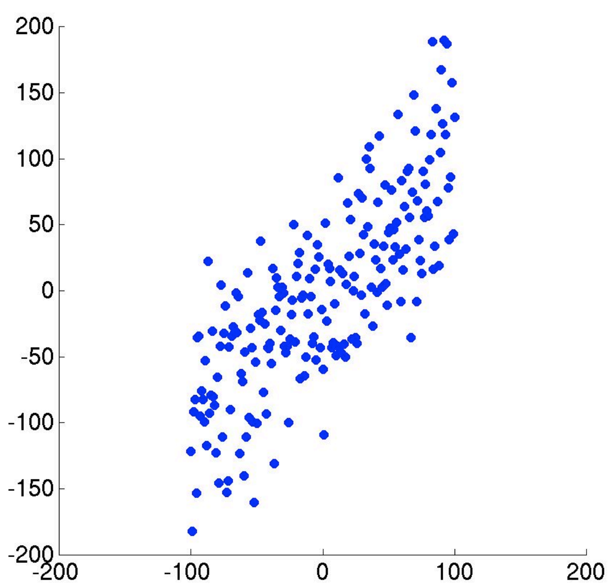
Least squares regression:

$$\min_{\beta} \sum_n (y_n - \beta x_n)^2$$

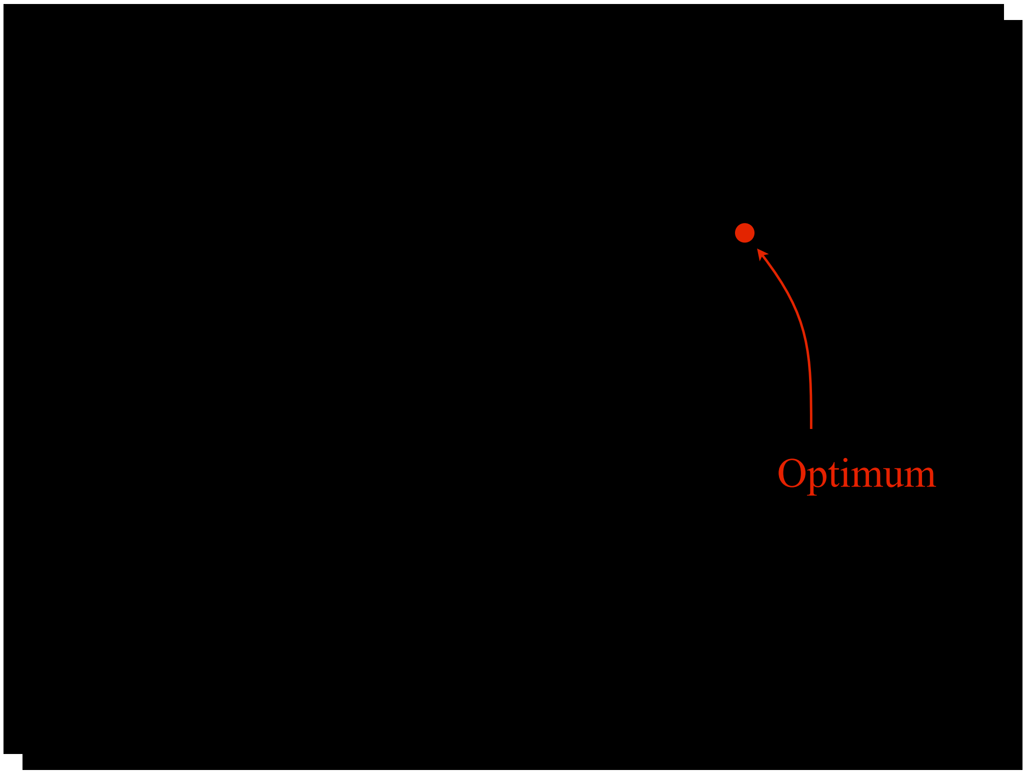
"objective function"

In the space of measurements:





"objective function"



$$\min_{\beta} \sum_{n} (y_n - \beta x_n)^2$$

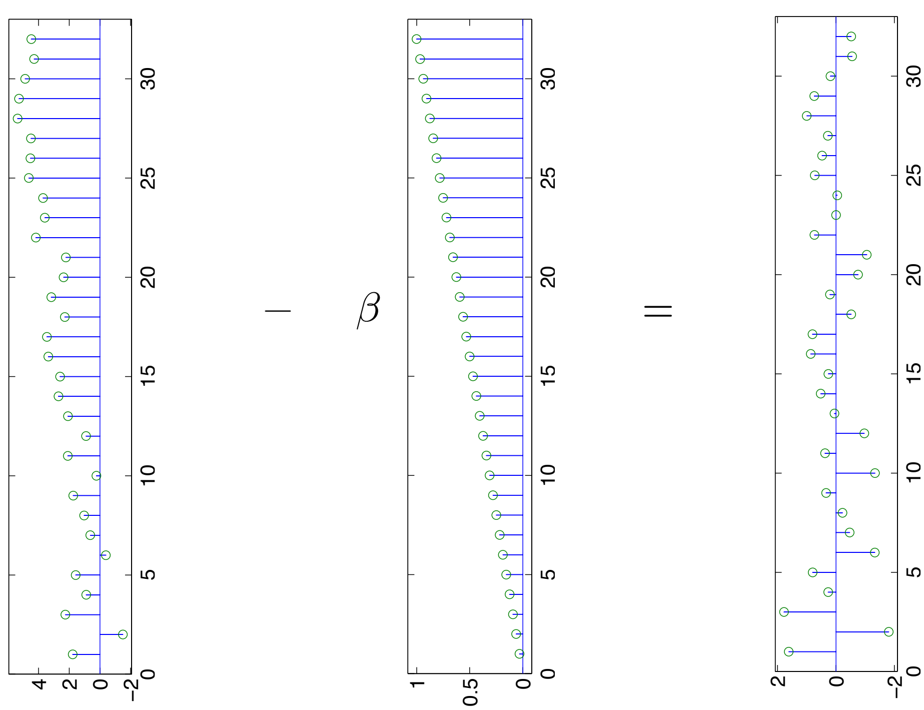
can solve this with calculus...

[on board]

... or linear algebra:

$$\min_{\beta} \sum_{n} (y_n - \beta x_n)^2 \;=\; \min_{\beta} ||\vec{y} - \beta\vec{x}||^2$$

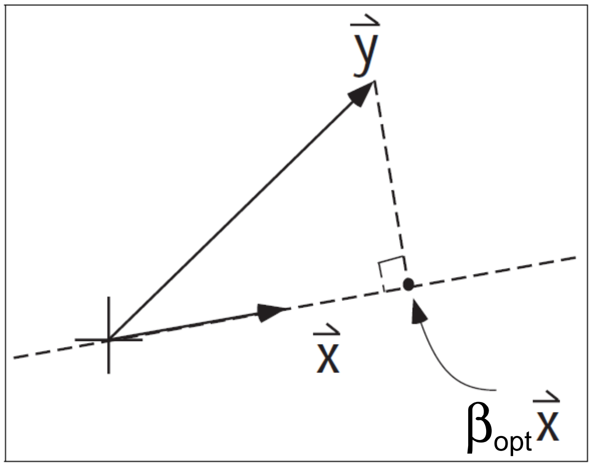
Observation $\vec{y}$ | Regressor $\vec{x}$ | Residual error



... or linear algebra:

$$\min_{\beta} \sum_{n} (y_n - \beta x_n)^2 \quad = \min_{\beta} ||\vec{y} - \beta \vec{x}||^2$$

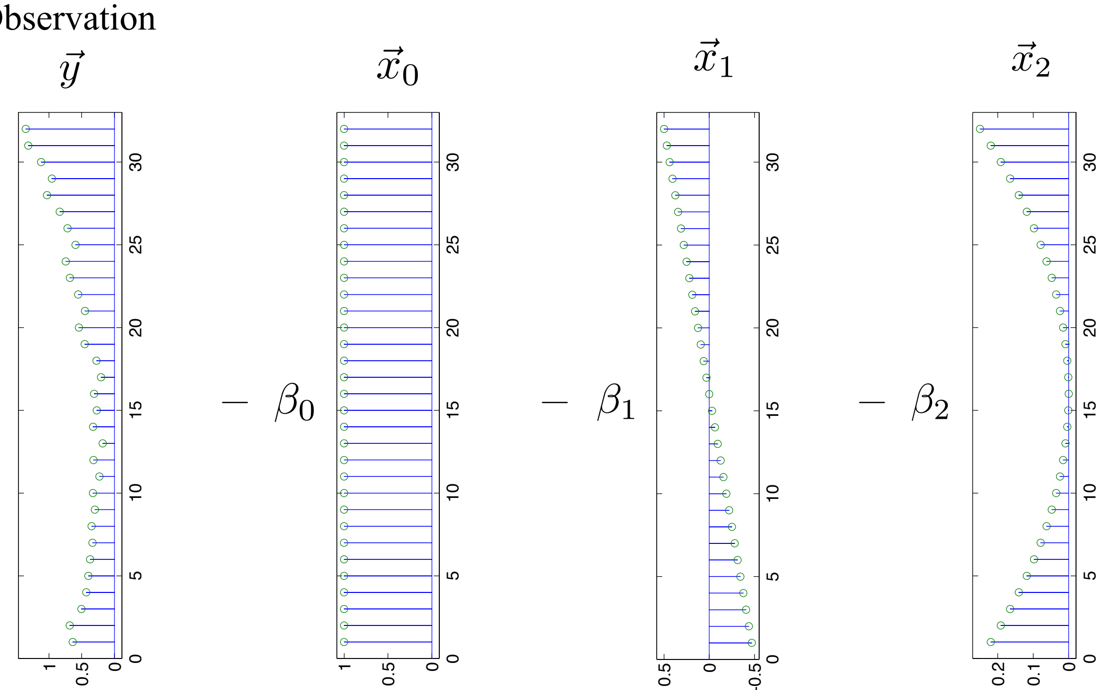
Geometry:
(note: this is *not* the measurement space of previous plots)



Multiple regression:

$$\min_{\vec{\beta}} ||\vec{y} - \sum_{k} \beta_k \vec{x}_k||^2 = \min_{\vec{\beta}} ||\vec{y} - X\vec{\beta}||^2$$

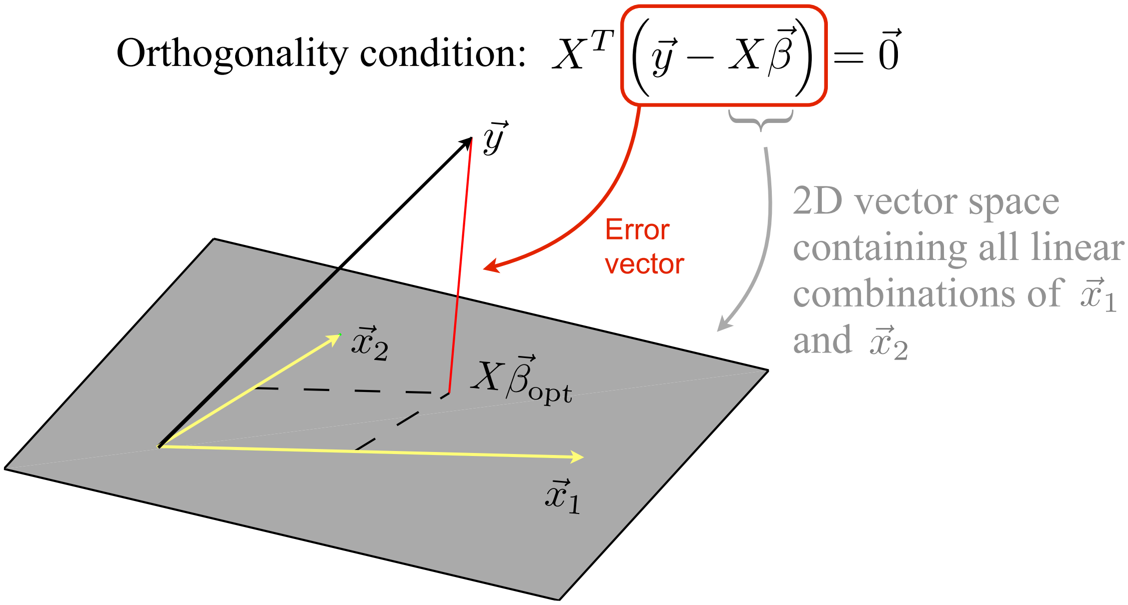
Observation



Solution via the Orthogonality Principle

Construct matrix X, containing columns $\vec{x}_1$ and $\vec{x}_2$

Orthogonality condition: $X^T \left(\vec{y} - X\vec{\beta}\right) = \vec{0}$



Alternatively, use SVD...

11

$$\min_{\vec{\beta}} ||\vec{y} - X\vec{\beta}||^2 = \min_{\vec{\beta}} ||\vec{y} - USV^T\vec{\beta}||^2$$

$$= \min_{\vec{\beta}} ||U^T\vec{y} - SV^T\vec{\beta}||^2$$

$$= \min_{\vec{\beta}^*} ||\vec{y}^* - S\vec{\beta}^*||^2$$

$$\text{where} \quad \vec{y}^* = U^T\vec{y}, \quad \vec{\beta}^* = V^T\vec{\beta}$$

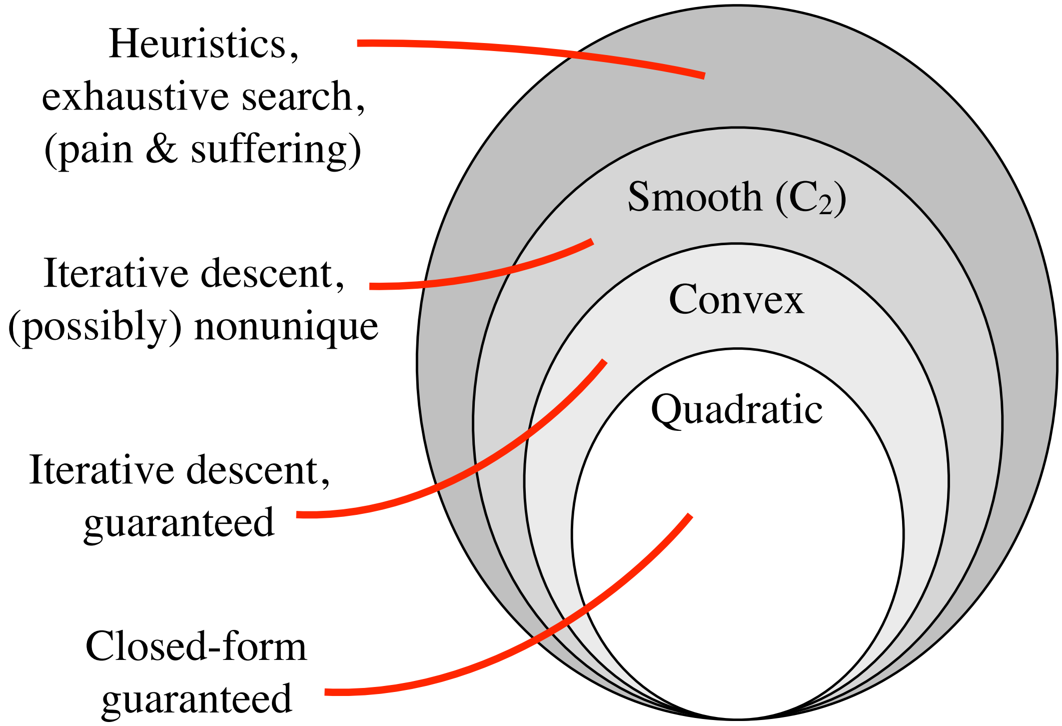
Solution: $\beta^*_{\text{opt},k} = y^*_k / s_k, \quad \text{for each } k$

or $\vec{\beta}^*_{\text{opt}} = S^{\#}\vec{y}^*$

[on board: transformations, elliptical geometry]

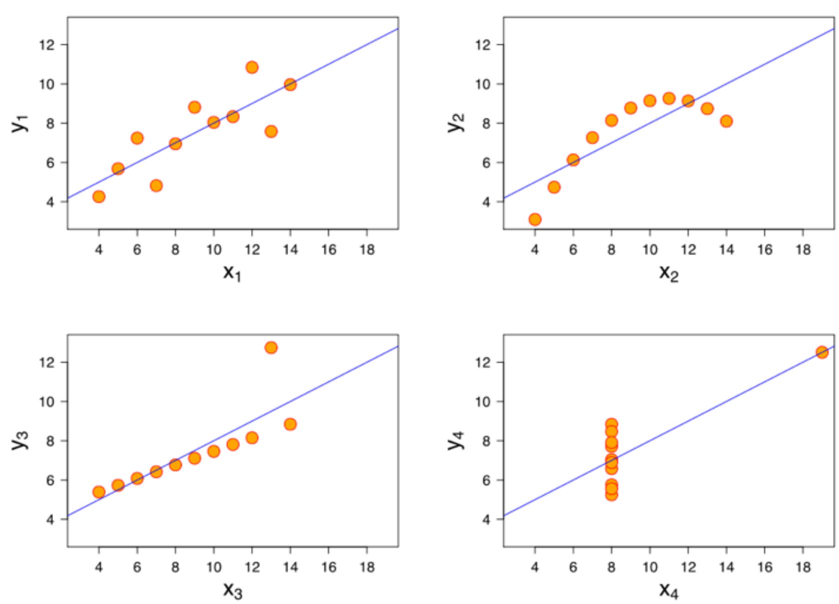
12

Optimization problems



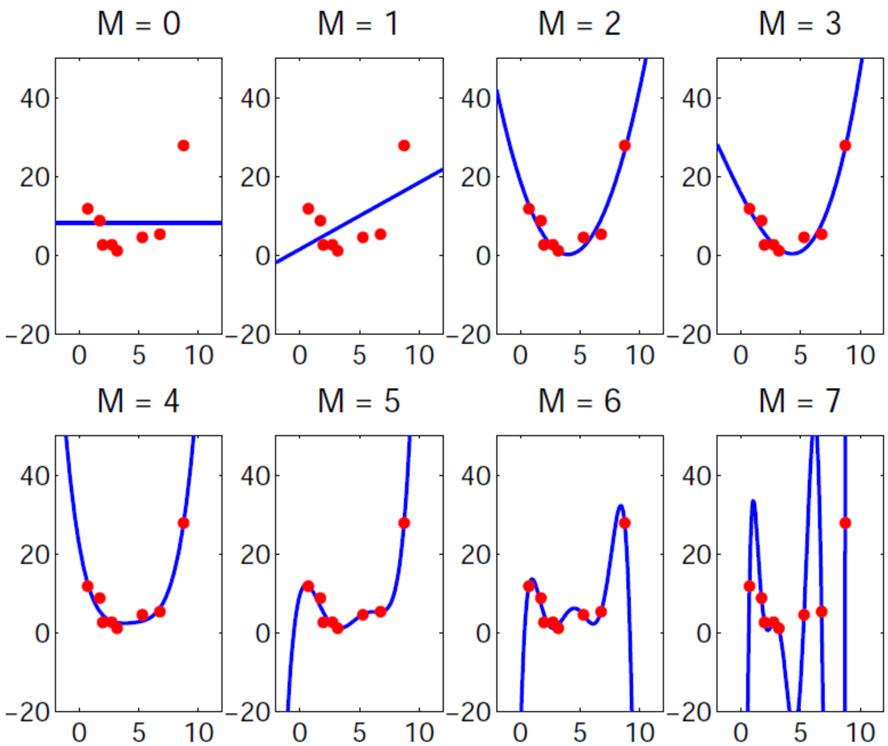
Interpretation: what does it mean?

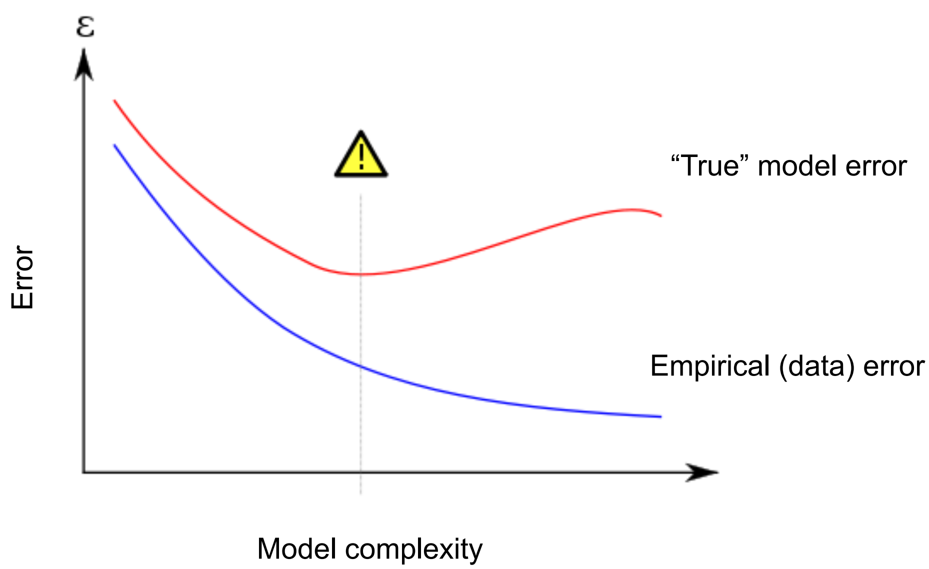
Note that these all give the same regression fit:



[Anscombe, 1973]

Polynomial regression - how many terms?





(to be continued, when we get to “statistics”...)

Weighted Least Squares

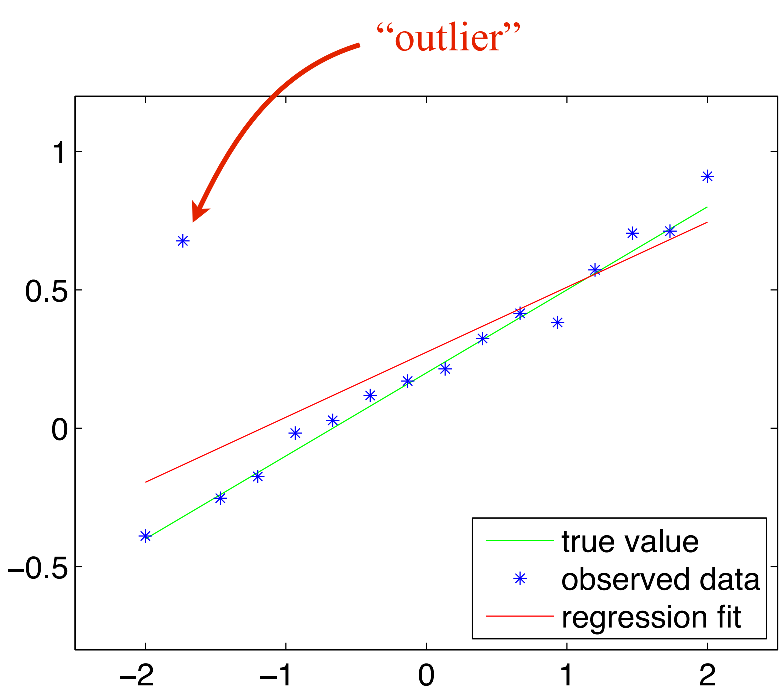
$$\min_{\beta} \sum_{n} \left[w_n (y_n - \beta x_n) \right]^2$$

$$= \min_{\beta} ||W(\vec{y} - \beta \vec{x})||^2$$

diagonal matrix

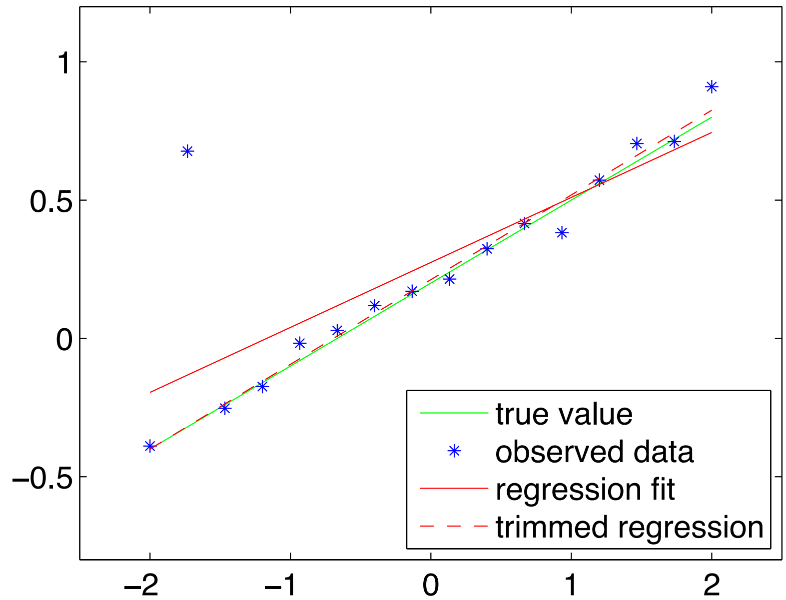
Solution via simple extensions of basic regression solution
(i.e., let $\vec{y}^* = W\vec{y}$ and $\vec{x}^* = W\vec{x}$ and solve for β)

17



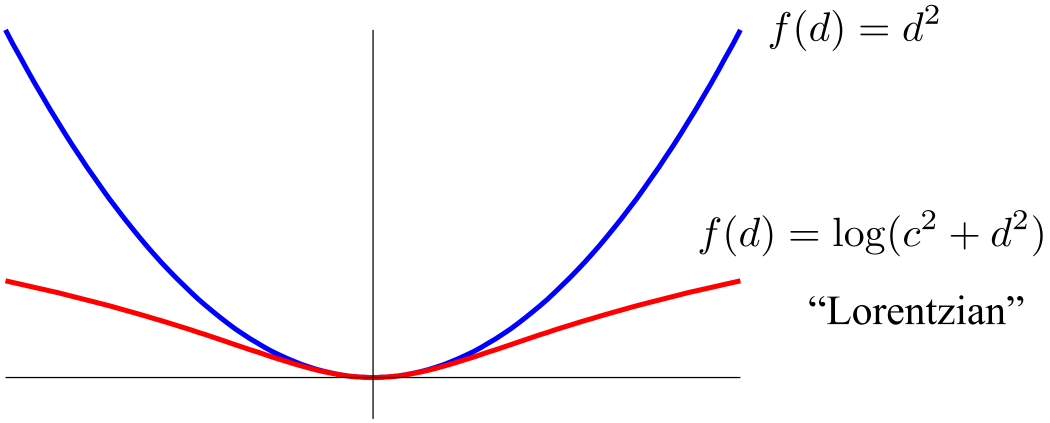
Solution 1: "trimming"...

18



When done iteratively (discard the outlier, re-fit, repeat), this is a so-called "greedy" method. When do you stop?

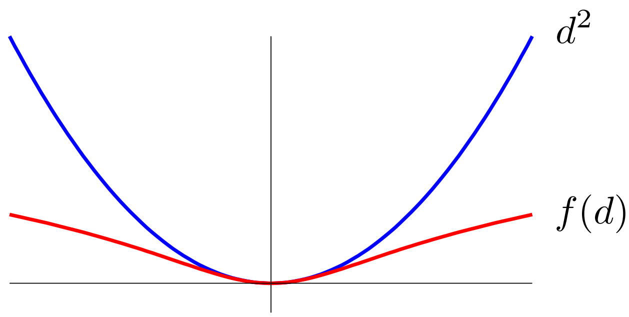
Solution 2: Use a "robust" error metric.
For example:



Note: generally can't obtain solution directly (i.e., requires an iterative optimization procedure).

In some cases, can use iteratively re-weighted least squares (IRLS)...

Iteratively Re-weighted Least Squares (IRLS)



initialize: $w_n^{(0)} = 1$

$$\beta^{(i)} = \arg\min_{\beta} \sum_n w_n^{(i)} \left[(y_n - \beta x_n)\right]^2$$

$$w_n^{(i+1)} = \frac{f(y_n - \beta^{(i)} x_n)}{(y_n - \beta^{(i)} x_n)^2}$$

iterate

(one of many variants)

Constrained Least Squares

Linear constraint:

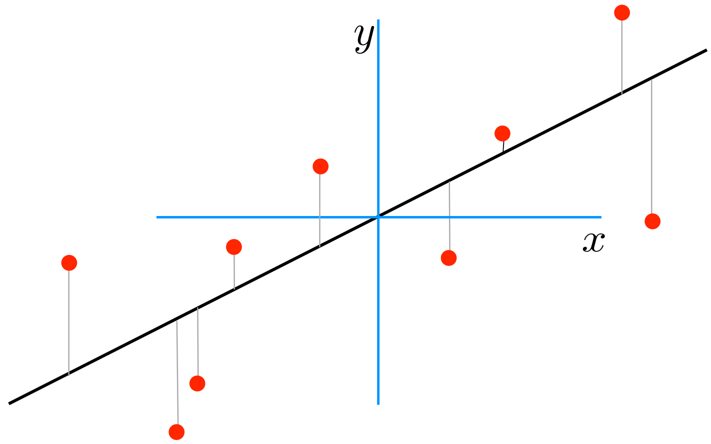
$$\min_{\vec{\beta}} ||\vec{y} - X\vec{\beta}||^2, \quad \text{where } \vec{c} \cdot \vec{\beta} = \alpha$$

Quadratic constraint:

$$\min_{\vec{\beta}} ||\vec{y} - X\vec{\beta}||^2, \quad \text{where } ||C\vec{\beta}||^2 = 1$$

Both can be solved exactly using linear algebra (SVD)...
[on board, with geometry]

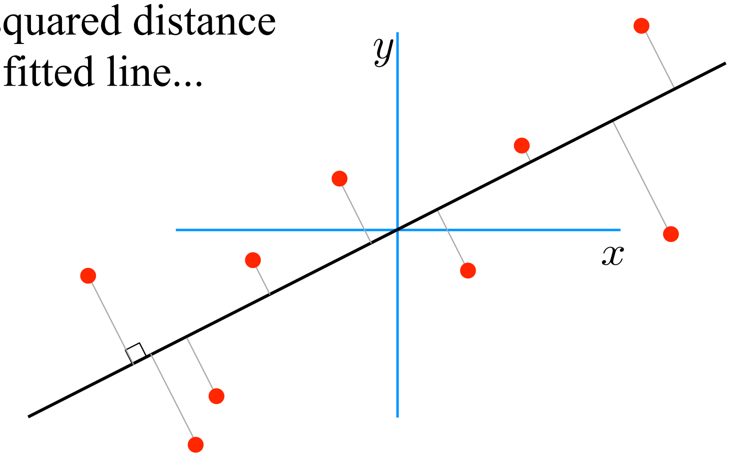
Standard Least Squares regression objective:
Squared error of the "dependent" variable



$$\min_{\beta} ||\vec{y} - \beta\vec{x}||^2$$

Total Least Squares Regression
(a.k.a "orthogonal regression")

Error is squared distance from the fitted line...

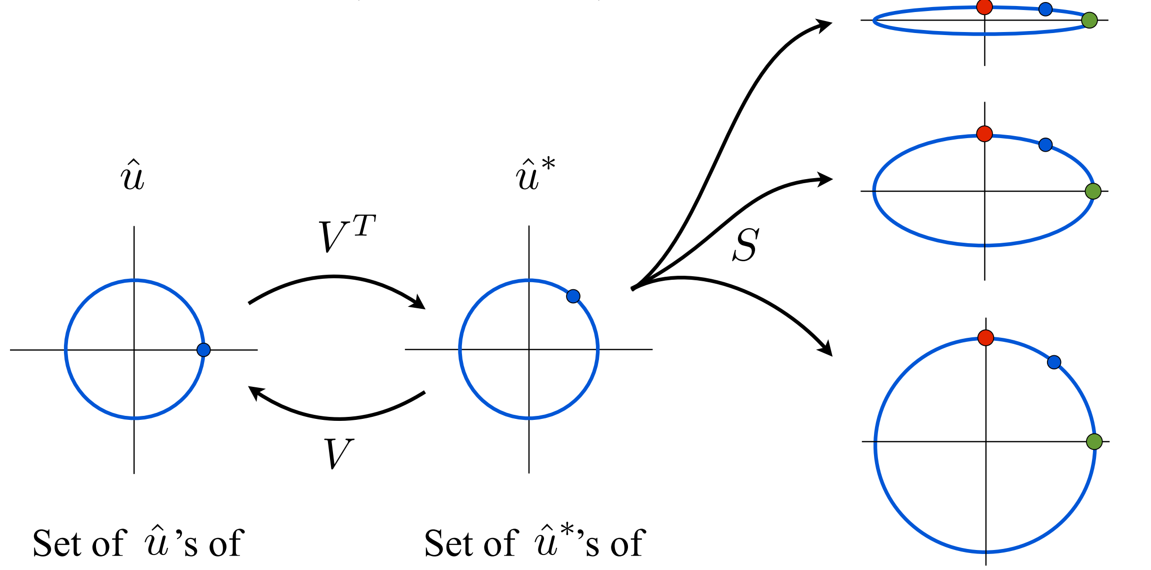


$$\min_{\hat{u}} ||D\hat{u}||^2, \quad \text{where } ||\hat{u}||^2 = 1$$

Note: "data" now includes both x and y coordinates

$$||USV^T\hat{u}||^2 = ||SV^T\hat{u}||^2 = ||S\hat{u}^*||^2 = ||\vec{u}^{**}||^2,$$

$$\text{where } D = USV^T, \quad \hat{u}^* = V^T\hat{u}, \quad \vec{u}^{**} = S\hat{u}^*$$



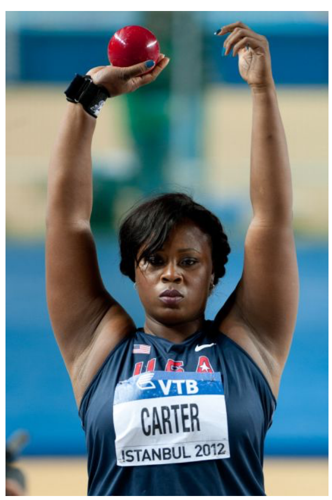
Set of $\hat{u}$'s of length 1 (i.e., unit vectors)

Set of $\hat{u}^*$'s of length 1 (i.e., unit vectors)

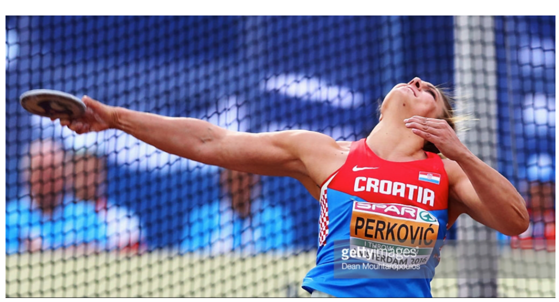
First two components of $\vec{u}^{**}$ (rest are zero!), for three example S's.

Olympic gold medalists (Rio, 2016)

Thomas Röhler (Germany)



Michelle Carter (USA)



Sandra Perković (Croatia)

3D geometry:
Javelin, Discus, Shotput

Eigenvectors/eigenvalues

Define symmetric matrix:

$$
\begin{aligned}
C &= M^T M \\
&= (USV^T)^T (USV^T) \\
&= V S^T U^T U S V^T \\
&= V (S^T S) V^T
\end{aligned}
$$

- "rotate, stretch, rotate back"
- matrix C "summarizes" the shape of the data

$\vec{v}_k$, the *k*th columns of V, is an *eigenvector* of *C*:

$$
\begin{aligned}
C\vec{v}_k &= V(S^T S) V^T \vec{v}_k \\
&= V(S^T S)\hat{e}_k \\
&= s_k^2 V \hat{e}_k \\
&= s_k^2 \vec{v}_k
\end{aligned}
$$

- output is a rescaled copy of input
- scale factor s_k^2 is called the *eigenvalue* associated with $\vec{v}_k$