Mathematical Tools
for Neural and Cognitive Science

Fall semester, 2019

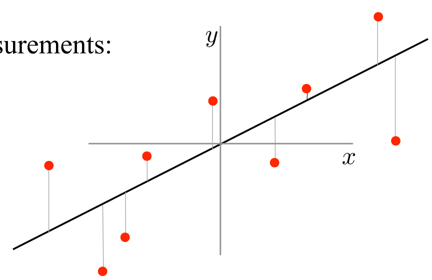
Section 2: Least Squares

Least squares regression:

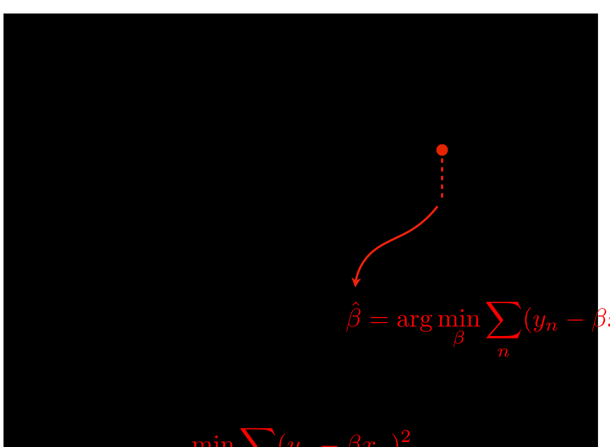
“objective” or “error" function

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

In the space of measurements:



[Gauss, 1795 - age 18]



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

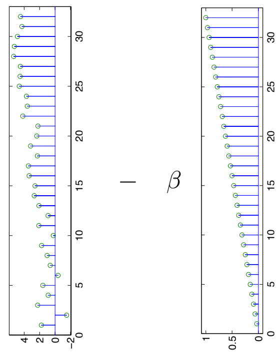
$$\min_{\beta} \sum_n (y_n - \beta x_n)^2$$

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

can solve this with calculus… *[on board]*

... or with linear algebra!

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

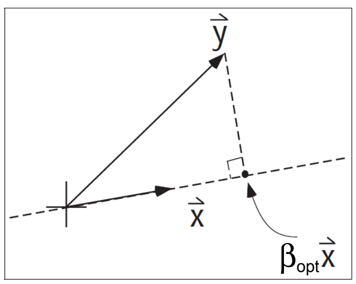


Observation $\vec{y}$ Regressor $\vec{x}$

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

Geometry:

Note: this is not the two-dimensional *(x,y)* measurement space of previous plots!



Note: partition of sum of squared data values:

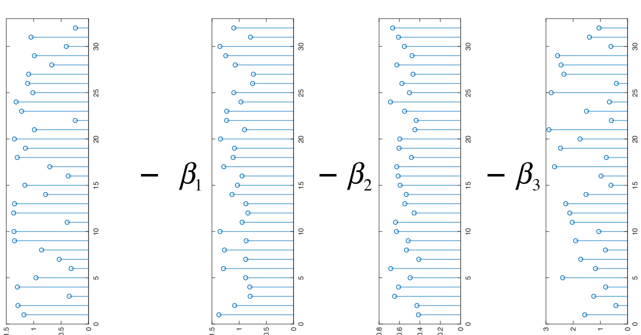
$$||\vec{y}||^2 = ||\beta_{\text{opt}}\vec{x}||^2 + ||\vec{y} - \beta_{\text{opt}}\vec{x}||^2$$

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

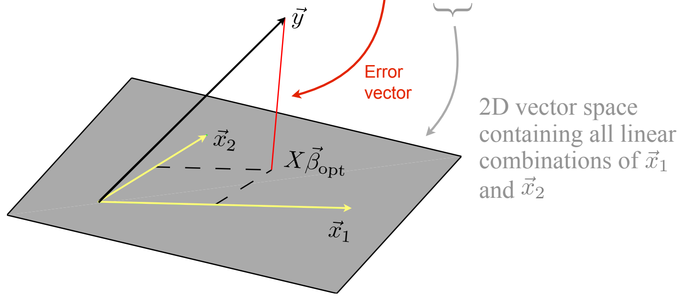
$\vec{y}$ $\vec{x}_1$ $\vec{x}_2$ $\vec{x}_3$



Solution via the “Orthogonality Principle”:

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

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



Alternatively, can solve using SVD...

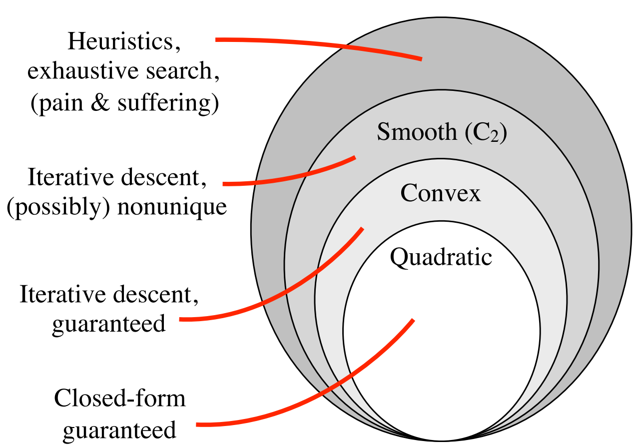
$$\min_{\vec{\beta}} ||\vec{y} - X\vec{\beta}||^2 = \min_{\vec{\beta}} ||\vec{y} - USV^T\vec{\beta}||^2$$
$$= \min_{\vec{\beta}} ||U^T\vec{y} - SV^T\vec{\beta}||^2$$
$$= \min_{\vec{\beta}^*} ||\vec{y}^* - S\vec{\beta}^*||^2$$
$$\text{where} \quad \vec{y}^* = U^T\vec{y}, \quad \vec{\beta}^* = V^T\vec{\beta}$$

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

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

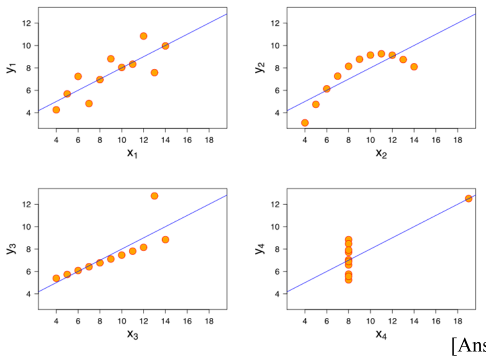
[on board: transformations, elliptical geometry]

Optimization problems



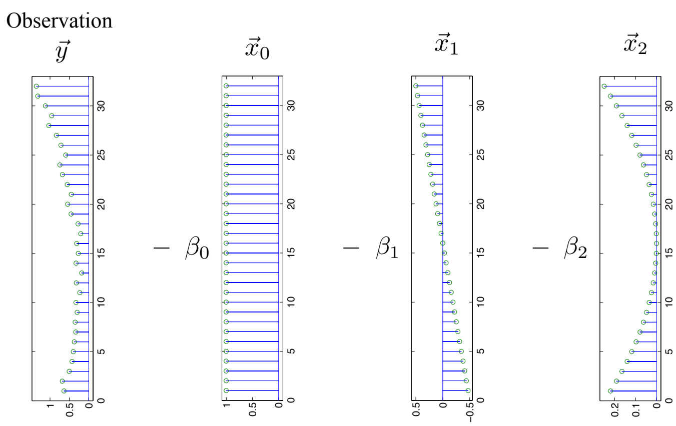
Note: fitting with a line does not guarantee data actually lie along a line…

These 4 data sets give the same regression fit, and same error:

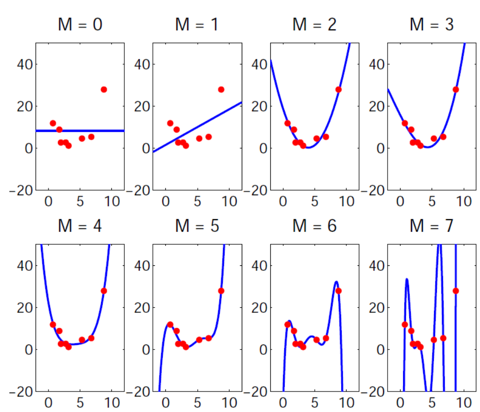


[Anscombe, 1973]

Polynomial regression



Polynomial regression - how many terms?



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

Weighted Least Squares

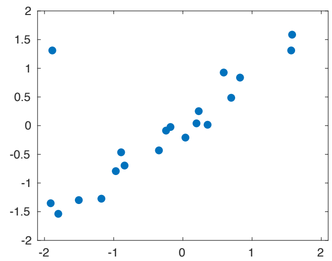
$$\min_{\beta} \sum_{n} \left[w_n (y_n - \beta x_n) \right]^2$$

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

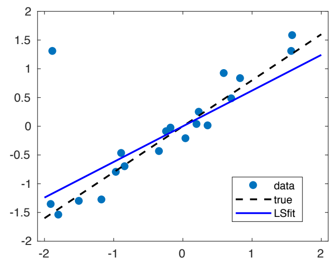
diagonal matrix

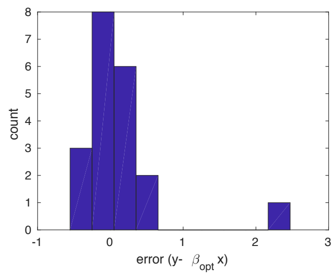
Solution via simple extensions of basic regression solution
(i.e., let $\vec{y}^* = W\vec{y}$ and $\vec{x}^* = W\vec{x}$ and solve for β)

Outliers

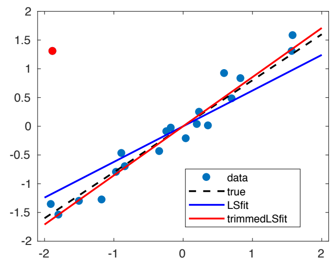


Outliers



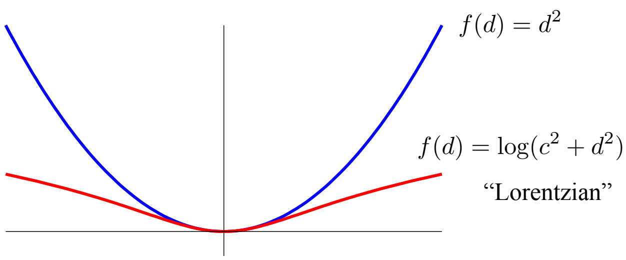


Solution 1: "trimming"… discard points with "large" error.
Note: a special case of weighted least squares.



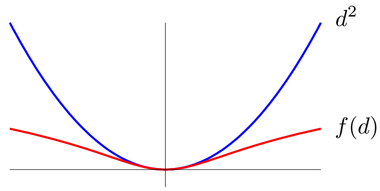
Trimming can be done iteratively (discard outlier, re-fit, repeat), a so-called "greedy" method. When do you stop?

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

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

iterate

iterate

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

(one of many variants)

Constrained Least Squares

Linear constraint:

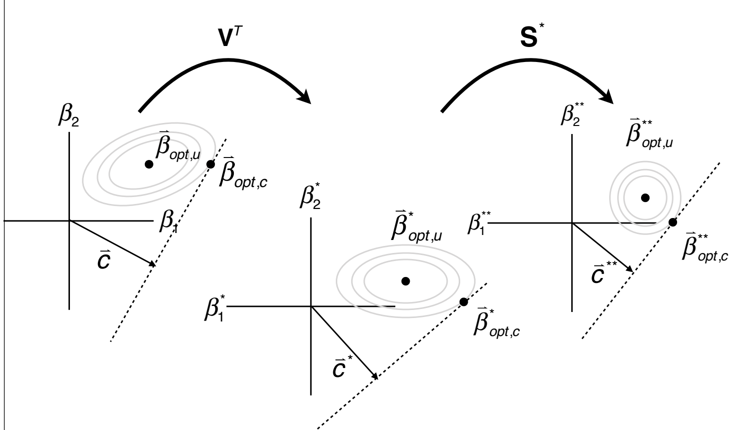
$$\min_{\vec{\beta}} ||\vec{y} - X\vec{\beta}||^2, \quad \text{where } \vec{c} \cdot \vec{\beta} = \alpha$$

Quadratic constraint:

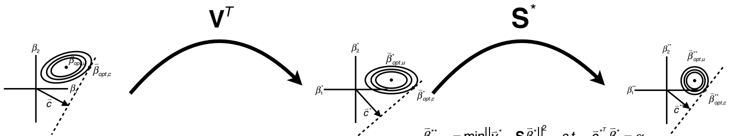
$$\min_{\vec{\beta}} ||\vec{y} - X\vec{\beta}||^2, \qquad \text{where} ||\beta||^2 = 1$$

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

Constrained Least Squares



Constrained Least Squares



$$\vec{\beta}_{opt,c} = \min_{\vec{\beta}} \left\| \vec{y} - \mathbf{X}\vec{\beta} \right\|^2 \quad s.t. \quad \vec{c}^T\vec{\beta} = \alpha$$
$$= \min_{\vec{\beta}} \left\| \vec{y} - \mathbf{USV}^T\vec{\beta} \right\|^2$$

$$\vec{\beta}^{**}_{opt,c} = \min_{\vec{\beta}^{**}} \left\| \vec{y}^* - \mathbf{S}\vec{\beta}^* \right\|^2 \quad s.t. \quad \vec{c}^{*T}\vec{\beta}^* = \alpha$$
$$= \min_{\vec{\beta}^{**}} \left\| \vec{y}^{**} - \vec{\beta}^{**} \right\|^2 \quad s.t. \quad \vec{c}^{*T}\left(\mathbf{S}^{*-1}\vec{\beta}^{**}\right) = \alpha = \vec{c}^{**T}\vec{\beta}^{**}$$

$\vec{y}^{**}$ = top n rows of $\vec{y}^*$

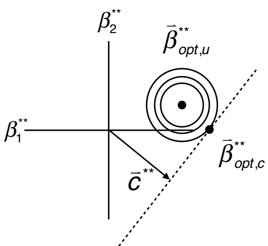
$\mathbf{S}^*$ = top n rows of $\mathbf{S}$

$\vec{\beta}^{**} = \mathbf{S}^*\vec{\beta}^*$

$\vec{c}^{**} = \left(\vec{c}^{*T}\mathbf{S}^{*-1}\right)^T = \mathbf{S}^{*-1^T}\vec{c}^* = \mathbf{S}^{*-1}\vec{c}^*$

$$\vec{\beta}^*_{opt,c} = \min_{\vec{\beta}^*} \left\| \mathbf{U}^T\left(\vec{y} - \mathbf{USV}^T\vec{\beta}\right) \right\|^2 \quad s.t. \quad \vec{c}^T\vec{\beta} = \alpha$$
$$= \min_{\vec{\beta}^*} \left\| \vec{y}^* - \mathbf{S}\vec{\beta}^* \right\|^2 \quad s.t. \quad \vec{c}^T\mathbf{V}\vec{\beta}^* = \alpha = \vec{c}^{*T}\vec{\beta}^*$$
$$\vec{y}^* = \mathbf{U}^T\vec{y}, \quad \vec{\beta}^* = \mathbf{V}^T\vec{\beta}, \quad \vec{c}* = \left(\vec{c}^T\mathbf{V}\right)^T = \mathbf{V}^T\vec{c}$$

Constrained Least Squares



$$\vec{\beta}^{**}_{opt,c} = \vec{\beta}^{**}_{opt,u} - \gamma\vec{c}^{**}$$

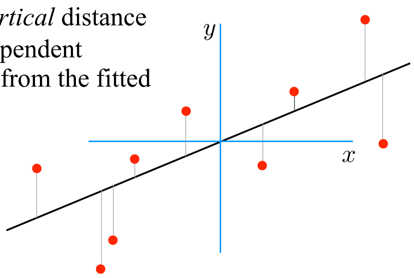
and satisfies the constraint

$$\vec{\beta}^{**}_{opt,c} \bullet \vec{c}^{**} = \alpha$$
$$= \vec{\beta}^{**}_{opt,u} \bullet \vec{c}^{**} - \gamma\vec{c}^{**} \bullet \vec{c}^{**}$$
$$\gamma = \frac{\vec{\beta}^{**}_{opt,u} \bullet \vec{c}^{**} - \alpha}{\vec{c}^{**} \bullet \vec{c}^{**}} = \frac{\vec{y}^{**} \bullet \vec{c}^{**} - \alpha}{\vec{c}^{**} \bullet \vec{c}^{**}}$$

Solution: $\gamma \rightarrow \vec{\beta}^{**}_{opt,c} \xrightarrow{\text{Stretch by } \mathbf{S}^{*-1}} \vec{\beta}^*_{opt,c} \xrightarrow{\text{Rotate by } \mathbf{V}} \vec{\beta}_{opt,c}$

Standard Least Squares regression

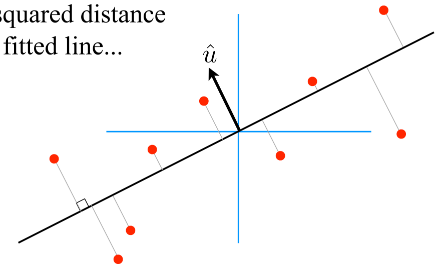
Error is *vertical* distance (in the "dependent variable") from the fitted line...



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

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

Error is squared distance from the fitted line...



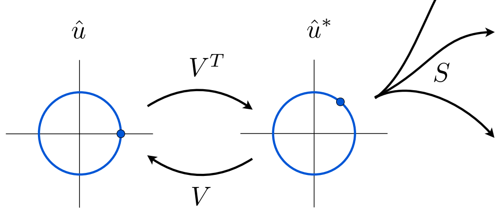
expressed as: $\min_{\hat{u}} ||D\hat{u}||^2, \quad \text{where } ||\hat{u}||^2 = 1$

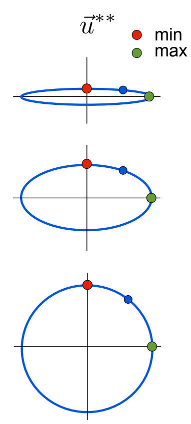
Note: "data" matrix D now includes both x and y coordinates

Variance of data D, projected onto axis $\hat{u}$:

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

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



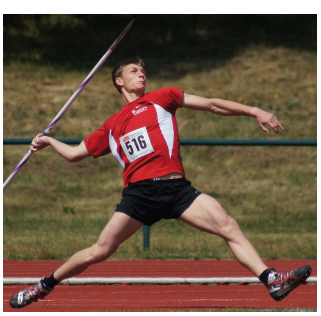


Set of $\hat{u}$'s of length 1 (i.e., unit vectors)

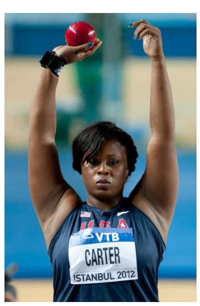
Set of $\hat{u}^*$'s of length 1 (i.e., unit vectors)

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

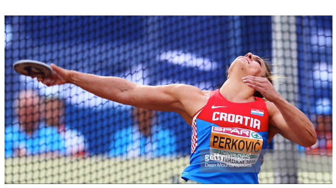
Olympic gold medalists (Rio, 2016)



Thomas Röhler (Germany)



Michelle Carter (USA)



Sandra Perković (Croatia)

3D geometry:
Javelin, Discus, Shotput

Eigenvectors/eigenvalues

Define symmetric matrix:

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

- "rotate, stretch, rotate back"
- matrix C "summarizes" the shape of the data with an ellipsoid: principal axes are columns of V, dimensions are elements of S

$\hat{v}_k$, the kth column of V, is an *eigenvector* of C:

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

- *eigenvectors* are vectors that are rescaled by the matrix (i.e., direction is unchanged) - this is true for all columns of V
- scale factor s_k^2 is called the *eigenvalue* associated with $\hat{v}_k$

Principal Component Analysis (PCA)

The shape of a data cloud can be summarized with an ellipse (ellipsoid) using a simple procedure:

(1) Subtract mean of all data points, to re-center around origin

(2) Assemble centered data vectors in rows of a matrix, D

(3) Compute the SVD of D:

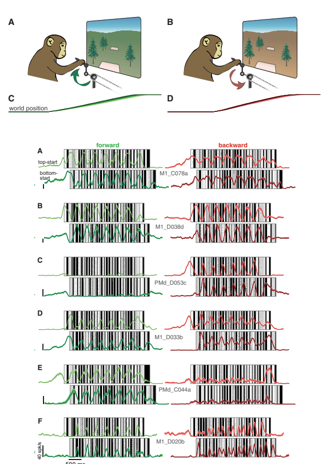
$$D = USV^T$$

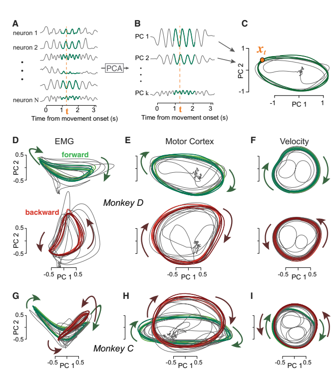
or compute eigenvectors of $C = D^T D$:

$$C = V \Lambda V^T$$

(4) Columns of V are the principal components (axes) of the ellipsoid, diagonal elements s_k or $\sqrt{\lambda_k}$ are the corresponding sizes of the ellipsoid

Example: PCA for dimensionality reduction and visualization





[Russo et. al., 2018]