Section 2: Least Squares

Least squares regression:

\[
\min_\beta \sum_n (y_n - \beta x_n)^2
\]

In the space of measurements:

\[
\hat{\beta} = \arg \min_\beta \sum_n (y_n - \beta x_n)^2
\]
\[ \min_{\beta} \sum_{n} (y_n - \beta x_n)^2 \] can solve this with calculus...

... or with linear algebra!

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

Geometry:

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

Note: partition of sum of squared data values:

\[ ||\vec{y}||^2 = ||\beta_{opt} \vec{x}||^2 + ||\vec{y} - \beta_{opt} \vec{x}||^2 \]

**Multiple regression:**

\[ \min_{\beta} ||\vec{y} - \sum_{k} \beta_k \vec{x_k}||^2 = \min_{\beta} ||\vec{y} - X \vec{\beta}||^2 \]
Solution via the “Orthogonality Principle”:

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

Orthogonality: $X^T (\tilde{y} - X\tilde{\beta}) = \tilde{0}$

Alternatively, can solve using SVD...

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

Solution: $\tilde{\beta}_{\text{opt},k} = y_k^*/s_k$, for each $k$

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

[on board: transformations, elliptical geometry]

Optimization problems

- Heuristics, exhaustive search, (pain & suffering)
- Iterative descent, (possibly) nonunique
- Iterative descent, guaranteed
- Closed-form guaranteed
- Smooth ($C^2$)
- Convex
- Quadratic
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:

![Graphs](Anscombe, 1973)

Polynomial regression

Observation

\[
\begin{align*}
\tilde{y} &
\end{align*}
\]

- $\beta_0$
- $\beta_1$
- $\beta_2$

Polynomial regression - how many terms?

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

\[
\min_\beta \sum_n [w_n(y_n - \beta x_n)]^2
\]

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

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 \( \beta \))
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:

\[ f(d) = d^2 \]

\[ f(d) = \log(c^2 + d^2) \]

“Lorentzian”

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

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

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

Constrained Least Squares

Linear constraint:

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

Quadratic constraint:

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

Both can be solved exactly using linear algebra (SVD)... on board, with geometry
**Constrained Least Squares**

\[ \beta_{opt.c} = \beta_{opt,u} - \gamma \tilde{c}^c \]

and satisfies the constraint

\[ \tilde{\beta}_{opt.c} \cdot \tilde{c}^c = \alpha \]

\[ \gamma = \frac{\tilde{\beta}_{opt.u} \cdot \tilde{c}^c - \alpha}{\tilde{c}^c \cdot \tilde{c}^c} \]

Solution: \( \gamma \rightarrow \tilde{\beta}_{opt.c} \)

**Constrained Least Squares**

\[ \tilde{\beta}_{opt.u} = m \tilde{y} - U \tilde{\beta} \]

\[ m \tilde{y} - US^T \tilde{\beta} \]

\[ \tilde{y} = \text{top } n \text{ rows of } \tilde{y} \]

\[ S \text{ = top } n \text{ rows of } S \]

\[ \tilde{\beta} = S \tilde{\beta} \]

\[ \tilde{c} = [\tilde{c}^c]^T = S^T \tilde{c} = S^T \tilde{c} \]

**Standard Least Squares regression**

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

\[ \text{arg min}_\beta ||\tilde{y} - \beta \tilde{x}||^2 \]
Total Least Squares Regression
(a.k.a “orthogonal regression”)

Error is squared distance from the fitted line...

expressed as: \( \min_{\hat{u}} \|D\hat{u}\|^2 \), where \( \|\hat{u}\|^2 = 1 \)

Note: “data” matrix \( D \) now includes both \( x \) and \( y \) coordinates

Variance of data \( D \), projected onto axis \( \hat{u} \):
\[
\|USV^T \hat{u}\|^2 = \|SV^T \hat{u}\|^2 = \|S\hat{u}^*\|^2 = \|\hat{u}^{**}\|^2,
\]
where \( D = USV^T, \hat{u}^* = V^T \hat{u}, \hat{u}^{**} = S\hat{u}^* \)

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 \( \hat{u}^{**} \) (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:

\[ C = D^T D \]

\[ = (USV^T)^T (USV^T) \]

\[ = VS^T U^T USV^T \]

\[ = V(S^T S)V^T \]

- "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 \( S \)

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

\[ C\hat{v}_k = V(S^TS)V^T \hat{v}_k \]

\[ = V(S^TB)\hat{e}_k \]

\[ = \hat{e}_k \]

\[ = s_k^2 \hat{v}_k \]

- 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

[Image: PCA for dimensionality reduction and visualization]