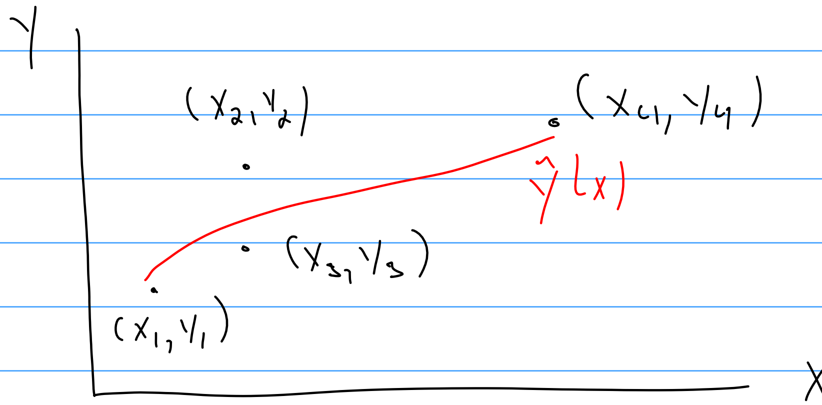
# Linear Regression

Consider n- data points $(x_i, y_i)$ $i \in [1, n]$




We want to find a linear regression, which is a linear combination of regressor functions $f(x)$,

$$\hat{y}(x) = \sum_{i=1}^{p} \underset{\substack{\uparrow \\ \text{Scalar}}}{a_i} \underset{\substack{\uparrow \\ \text{function}}}{f_i(x)} = a_1 f_1(x) + a_2 f_2(x) + \dots + a_p f_p(x)$$

$f_i(x)$ could be nonlinear.

ex.) let $f(x) = \{1, x, x^2\}$

$$\hat{y}(x) = \sum_{i=1}^{3} a_i f_i(x) = a_1 (1) + a_2 (x) + a_3 x^2$$

ex.) $f(x) = \{\sin(x), \cos(x), \sin(2x), \cos(2x)\}$

$$\hat{y}(x) = a_1 \sin(x) + a_2 \cos(x) + a_3 \sin(2x) + a_4 \cos(2x)$$

Ideally $\hat{y}(x_i) = y_i$ :

$$\hat{y}(x_1) = a_1 \underset{\uparrow \text{ known}}{f_1(x_1)} + a_2 f_2(x_1) + \dots + a_p f_p(x_1) = y_1$$

$$\hat{y}(x_2) = a_1 f_1(x_2) + \dots + a_p f_p(x_2) = y_2$$

Write as:

$$\begin{bmatrix} f_1(x_1) & f_2(x_1) & \dots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_p(x_2) \\ \vdots & & & \\ f_1(x_n) & f_2(x_n) & \dots & f_p(x_n) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\underset{(n \times p)}{\underline{F}} \quad \underset{(p \times 1)}{\underline{a}} = \underset{(n \times 1)}{\underline{y}}$$

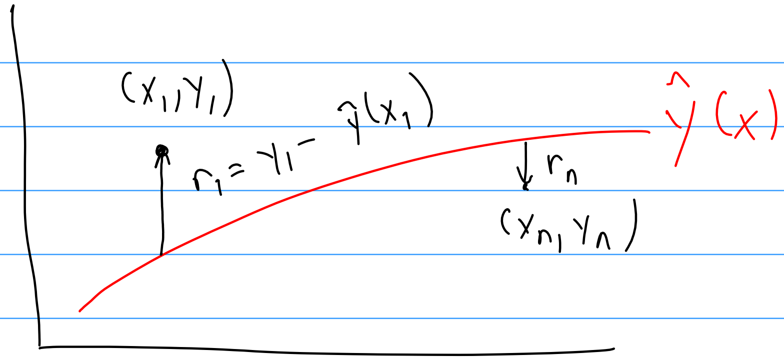
If $n = p$ & $x_i \neq x_j$ if $i \neq j$
$\underline{F}^{-1}$ exists

In general $n \neq p$ & $x_i = x_j$ if $i \neq j$
$\Rightarrow$ $F^{-1}$ does not exist

No solution to $\underline{F}\,\underline{a} = \underline{y}$

In this case find the best approximate solution.

the $\hat{\underline{a}}$ that minimizes $\|\underset{\text{residual}}{\underline{r}}\|_2 = \|\underline{y} - \underline{F}\,\hat{\underline{a}}\|_2$




It turns out that the solution to

$$\rightarrow \underline{F}^T \underline{F}\, \hat{\underline{a}} = \underline{F}^T \underline{y} \quad \leftarrow \text{Normal Equations}$$

that gives the least-squares solution to $\boxed{\underline{F}\,\underline{a} = \underline{y}}$

Notes: 1) will cover later: involves projections onto matrix subspaces.

2) Never form the Normal Equations!

Here's why:

a) $\kappa(\underline{A}^T) = \kappa(\underline{A})$

b) $\kappa(\underline{A}\,\underline{B}) = \kappa(\underline{A})\,\kappa(\underline{B})$

$$\Rightarrow \kappa(\underline{A}^T \underline{A}) = \kappa(\underline{A}^T)\,\kappa(\underline{A}) = (\kappa(\underline{A}))^2$$

$\Rightarrow$ If $\kappa(\underline{A})$ is large then $\kappa(\underline{A}^T \underline{A})$ is huge!

We will show indirect methods to solve this (QR, SVD, etc.)

3) $\underline{F}\,\underline{a} = \underline{y}$
$(n \times p)(p \times 1) = (n \times 1)$

$\underline{F}^T \underline{F}\,\underline{a} = \underline{F}^T \underline{y}$
$(p \times n)(n \times p)(p \times 1) = (p \times n)(n \times 1)$
$(p \times p)(p \times 1) = (p \times 1)$

$\underline{F}^T \underline{F}$ is symmetric & real(?)

$(\underline{F}^T \underline{F})^{-1}$ will always exist

4) In Matlab, the backslash operation:
$\underline{x} = \underline{A} \backslash \underline{b}$ for $\underline{x} = \underline{A}^{-1} \underline{b}$
will give the least-squares solution.

# Nonlinear Regression

What if you have n-Data points but you think the function $\frac{a_1 x}{x + a_2}$ best fits it?

Common examples: $e^{-(a_1 x + a_2)}$

Fit the function $\hat{y}(x, \underline{a})$

ex.) $\hat{y}(x, \underline{a}) = \frac{a_1 x}{x + a_2}$ $\qquad \underline{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$

Minimize the objective function

$$S(\underline{a}) = \sum_{i=1}^{n} r_i^2 \qquad \underbrace{r_i}_{\text{residual}} = y_i - \hat{y}(x_i, \underline{a})$$

You can use <u>any</u> minimization method.

---

## Gauss-Newton Method

Apply Newton method to

$$S(\underline{a}) = \sum_{i=1}^{n} r_i^2$$

$$\underline{a}_{k+1} = \underline{a}_k - \underline{H}_k^{-1} g_k$$

$$g_k = \nabla S(\underline{a}_k) \quad : \quad g_j = \left.\frac{\partial S}{\partial a_j}\right|_{\underline{a}_k}$$

$\underline{H}_k$ = Hessian of $S$ @ $\underline{a}_k$

$$H_{jk} = \frac{\partial^2 S}{\partial a_j \partial a_k} = \frac{\partial g_j}{\partial a_k}$$

Now look at $S = \sum r_i^2(\underline{a})$

$$g_j = 2\sum_{i=1}^{n} r_i \frac{\partial r_i}{\partial a_j} = 2\sum_{i=1}^{n} r_i J_{ij} = 2\underline{J}^T \underline{r}$$

$\underline{J}$ = Jacobian of $\underline{r}(\underline{a})$

$$H_{jk} = 2\sum_{i=1}^{m}\left(\frac{\partial r_i}{\partial a_j}\frac{\partial r_i}{\partial a_k} + \underbrace{r_i \frac{\partial^2 r_i}{\partial a_j \partial a_k}}_{\text{usually noisy, neglect it}}\right)$$

$$\Rightarrow H_{jk} \simeq 2\sum_{i=1}^{n} \frac{\partial r_i}{\partial a_j}\frac{\partial r_i}{\partial a_k} = 2\sum_{i=1}^{n} J_{ij} J_{ik} = 2\underline{J}^T\underline{J}$$

$$\Rightarrow \underline{a}_{k+1} = \underline{a}_k - \underline{H}_k^{-1} g_k$$

$$= \underline{a}_k - \cancel{\tfrac{1}{2}}(\underline{J}_k^T \underline{J}_k)^{-1}(\cancel{2}\,\underline{J}_k^T \underline{r}_k)$$

$$\underline{a}_{k+1} = \underline{a}_k - (\underline{J}_k^T \underline{J}_k)^{-1} (\underline{J}_k^T \underline{r}_k)$$

Write as $\underline{a}_{k+1} - \underline{a}_k = \underline{\delta}_k = -(\underline{J}_k^T \underline{J}_k)^{-1} \underline{J}_k^T \underline{r}_k$

$\underline{\delta}_k$ is least-squares solution to

$$\underline{J}_k \underline{\delta}_k = -\underline{r}_k$$

$$(\underline{J}_k^T \underline{J}_k \underline{\delta}_k = -\underline{J}_k^T \underline{\delta}_k)$$

Algorithm: Given $\underline{a}_k$ compute $\underline{J}_k$, $\underline{r}_k$

Solve $\underline{J}_k \underline{\delta}_k = -\underline{r}_k$

Set $\underline{a}_{k+1} = \underline{a}_k + \alpha_k \underline{\delta}_k$

Advantage is no Hessian calculation.

but, if $\left| r_i \frac{\partial^2 r_i}{\partial a_i \partial a_k} \right| \ll \left| \frac{\partial r_i}{\partial a_j} \frac{\partial r_i}{\partial a_k} \right|$

Does not hold might not converge.

Usually it is small given appropriate $\underline{a}_0$.