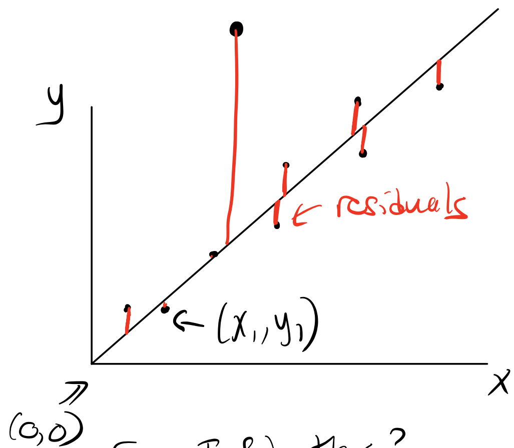
Linear Regression

Target variable is a linear function of the feature(s)
(y) $(x_1, x_2, \dots x_d)$



1 dimension: $y = ax$

d dimensions: $y = w_1 x_1 + w_2 x_2 + \dots + w_d x_d$

Can I fit this?

What is the "best" fit?

sum of squared residuals minimized

$$y_1 = a x_1$$

$$r_i = (a x_i - y_i)^2$$

$$\min_a \sum (a x_i - y_i)^2$$

How do I fit this? Let's do the general (d-dim) case

$$y = w_1 x_1 + w_2 x_2 + \dots + w_d x_d$$

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_d \end{pmatrix} \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

Translate into Linear algebra notation:

$$y_i = \vec{x}_i \cdot \vec{w} = \vec{x}_i^{\,T} \vec{w}$$

That's just one datapoint... we want to fit all of them!

$$\vec{y} = \underbrace{\begin{bmatrix} \leftarrow x_1^T \rightarrow \\ \leftarrow x_2^T \rightarrow \\ \vdots \\ \leftarrow x_n^T \rightarrow \end{bmatrix}}_{n \times d\,!} \vec{w}$$

$\vec{y} = X\vec{w}$

known known unknown

Solve for $\vec{w}$:

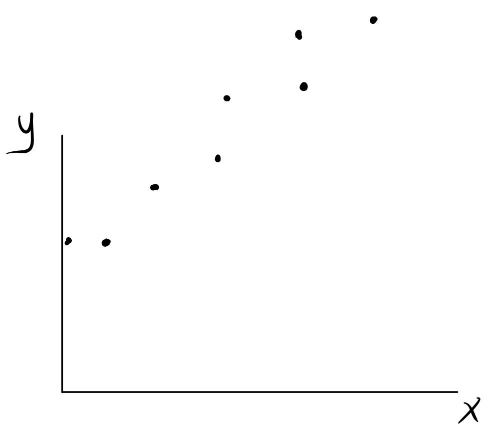
MATH 204/304: $(X^TX)^{-1}(X^T\vec{y}) = \vec{w}$

DATA 371: Spoilers...

"best fit" in linear algebra speak: $\min_{\vec{w}} \|X\vec{w} - y\|_2^2$

Tricks

- Predict multiple y? do ↑ repeatedly



1 dimension: $y = Wx$

Can I fit this?

$\vec{x}^T\vec{w} = \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_0 \end{bmatrix}$

Want: $y = W_1 x + W_0$

Trick 1: $x' \leftarrow \begin{bmatrix} x \\ 1 \end{bmatrix}$

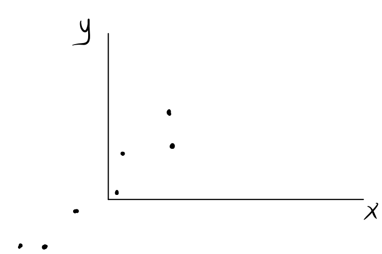
augment feature vector with a 1

"bias trick"

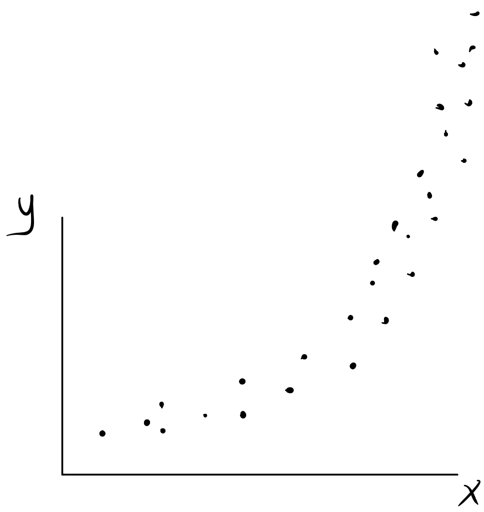
Trick 2:

$\vec{x}' \leftarrow \vec{x} - \bar{x}$

$\vec{y}' \leftarrow \vec{y} - \bar{y}$



center the data



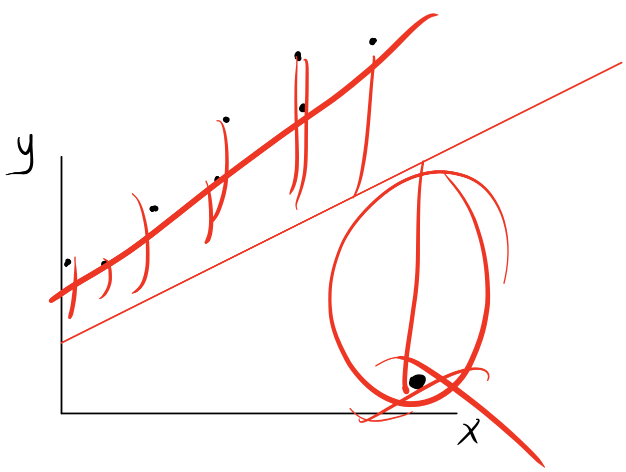
Can I fit this?

Trick: if $\vec{x} = [x_1, x_2, \dots x_d]$

$$\vec{x}' \leftarrow [x_1 \; x_1^2 \; x_1^3 \; x_2 \; x_2^2 \; x_2^3 \; \dots \; x_d, \; x_d^2 \; x_d^3]$$

$$w_1 x_1 + w_2 x_1^2 + w_3 x_1^3 + \dots$$

Limitations



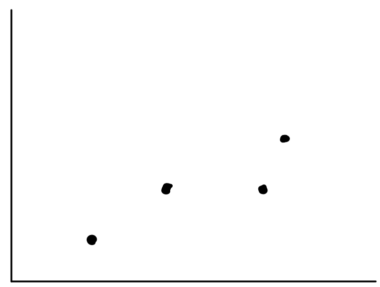
Can I fit this?

- outliers

- linearly dependent columns

Question:

if $\vec{x}' \leftarrow [x_1 \ x_1^2 \ x_1^3 \ x_2 \ x_2^2 \ x_2^3 \ \dots \ x_d, \ x_d^2 \ x_d^3]$,



What model will it find?

Trick: "regularization" - punish complicated models

Standard linear regression: $\min_{\vec{w}} \|X\vec{w} - y\|$

"ridge regression": $\min_{\vec{w}} \|X\vec{w} - y\| + \|\vec{w}\|$

"LASSO": $+ \|\vec{w}\|_1$