Lecture 3:

- Linear models
- Optimization perspective.
- BBO

Recall:

$(x_i, y_i)_{i=1}^{n}$

$x_i \in \mathbb{R}^m$

$y_i \in \mathbb{R}$

$x_{n+1} \longrightarrow \hat{y}_{n+1}$

$\hat{y}_{n+1} = f(x_{n+1})$

↳ "model"

$f_w \rightarrow$ parametric model

↳ weight or model params

$\hat{y}_{n+1} = \cancel{m} x_{i,1} + b$ (slope)

Linear Regression

- Assume a parametric form

$$f_w(x_i) = w_1 x_{i,1} + w_2 x_{i,2} + \cdots + w_m x_{i,m}$$

$$= \sum_{j=1}^{m} w_j x_{i,j} \quad x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,m})$$

$$= w \cdot x_i$$

$$= w^T x_i$$

$f_w(x_i) = w_1 x_{i,1} + w_2 x_{i,2}$

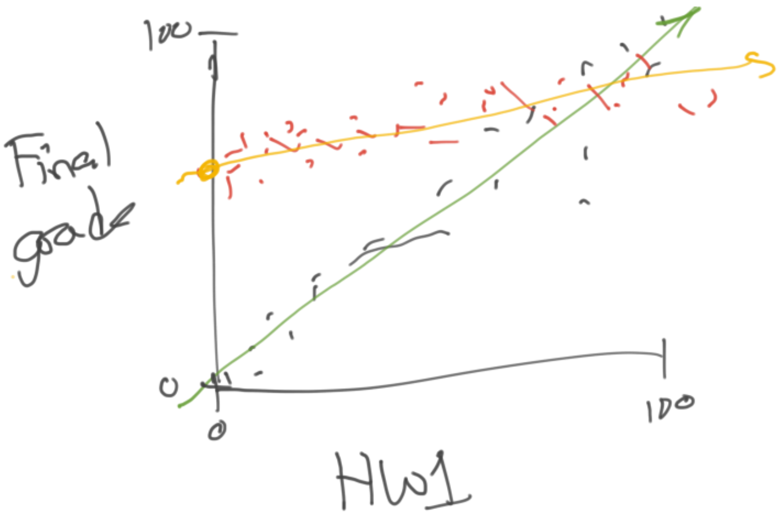
slope   always $= 1$

- Sometimes include extra weight as y-offset, equivalent to:

$$m \leftarrow m + 1$$

$$x_{i,m} = 1 \text{ always.}$$

↳ new m, equals old m, plus one.

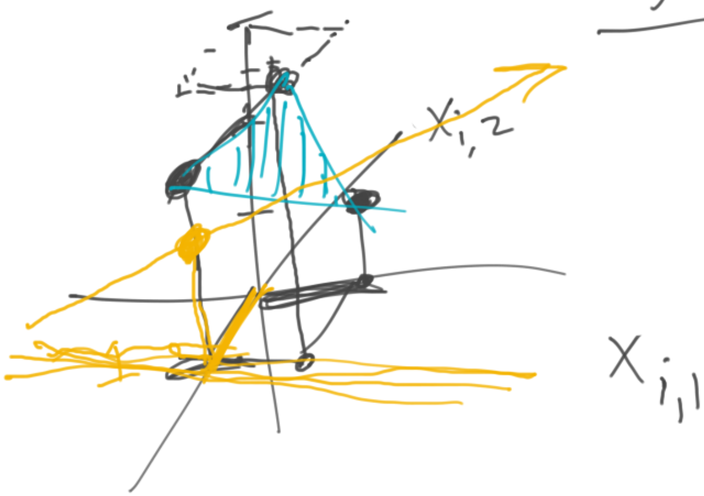


$Z \sim D$

$m = 1$

$x_{i,1} = $ HW1 score.

$x_{i,2} = 1$

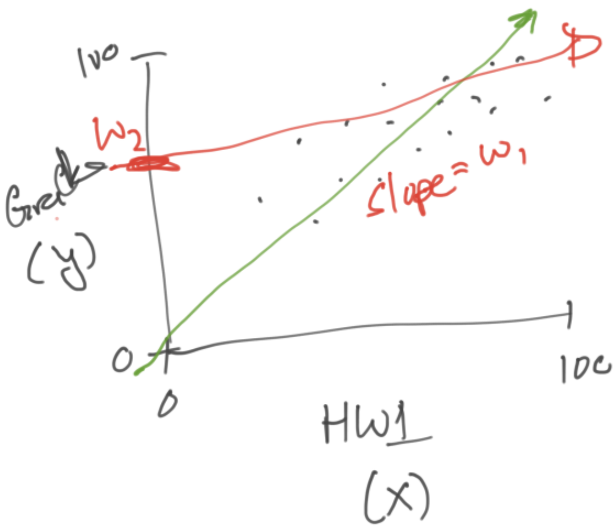


$$f_w(x_i) = \overset{1}{\overset{\uparrow}{w_1}}\, x_{i,1} + \overset{2}{\overset{\uparrow}{w_2}}\, x_{i,2}$$

$(1,0) \to 1$

$(0,1) \to 2$

$(1,1) \to 3$



$$\hat{y}_i = f(x_i)$$

$$f_w(x_i) = w_1 x_i + w_2$$

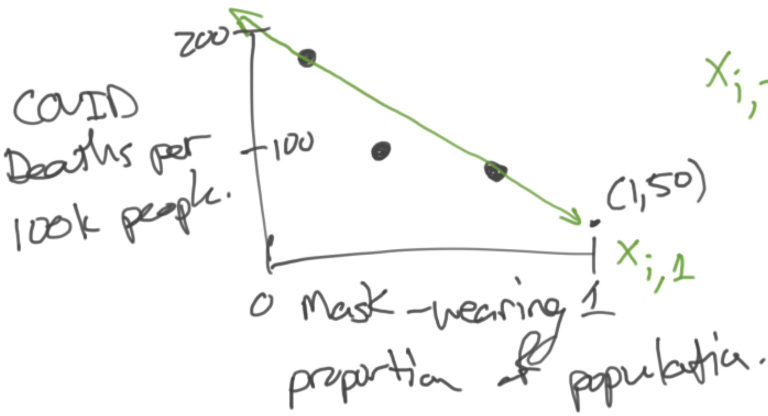
→ $m = 1$
↳ # of features.

$$f_w(x_i) = \sum_{j=1}^{m} w_j x_{i,j} = w_1 x_{i,1} + w_2 x_{i,2}$$

$$x_{i,2} = 1$$

$\forall i$.

Example:



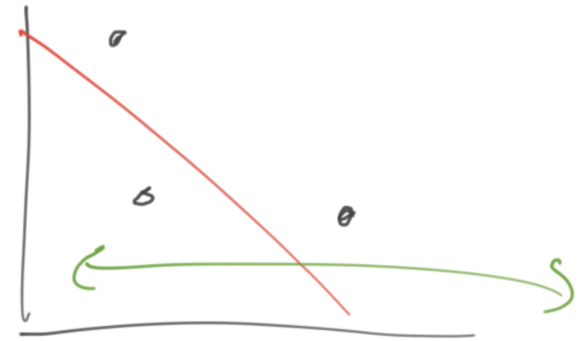
$m = 2$

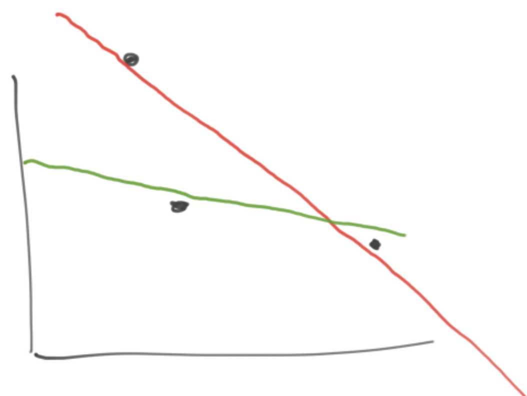
$x_{i,2} = 1$ always.

slope = $\frac{\text{rise}}{\text{run}}$

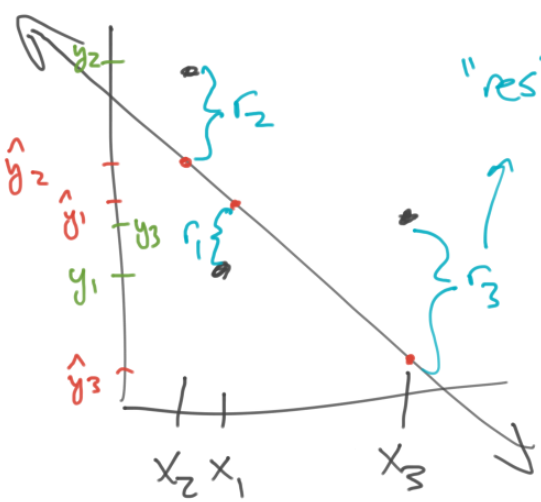
$w_1 = -150$ $w_2 = 200$

- What line is the "best fit"?





Most common: "least squares"



"residual"

$$r_i = y_i - \hat{y}_i$$
$$= y_i - f(x_i)$$
$$= y_i - f_w(x_i)$$
$$= y_i - (w_1 x_{i,1} + w_2 \overset{1}{\cancel{x_{i,2}}})$$

Best fit → smallest residual magnitude

↳ take $abs(r_i)$

Which is better:

i	line 1 *	line 2 *
1	1	3
2	3	1
3	-20	-1

loss function $\ell(\text{model})$ = how bad the model is.

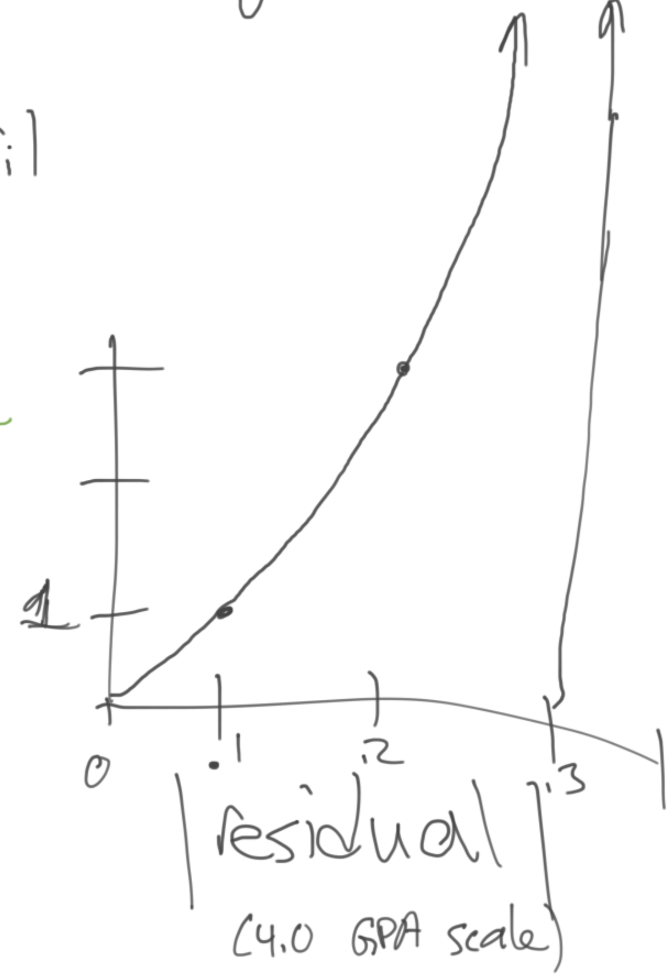
$$\ell(\text{model}) = \sum_{i=1}^{n} r_i$$ × negative r_i not good.

$$\ell(\text{model}) = \sum_{i=1}^{n} |r_i|$$

least squared loss function

$$\boxed{\ell(\text{model}) = \sum_{i=1}^{n} r_i^2}$$

- common assumption: cost of residuals scales with the square of the residual.



$$\ell(w) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} \big(\underbrace{y_i - f_w(x_i)}_{r_i}\big)^2$$

model parameters that define a model

Hyperparameters so far:

- Model parameterization (linear)
- Choice of loss function.

"best", "optimal"

$$w^* \in \underset{w \in \mathbb{R}^m}{\arg\min}\ \ell(w)$$

"in"

returns a set.

- called an "optimization problem"

ℓ is called a loss function

- Takes models as input (model parameters)
- Outputs how "bad" the model is
- "Best fit" parametric model minimizes loss function.

Linear Regression Summary

- Parametric vs. nonparametric.
    - alternatives
- $(x_i, y_i)_{i=1}^{n}$
- Parametric model $\hat{y}_i = f_w(x_i)$
- Linear $f_w(x_i) = w^T x_i = \sum_{j=1}^{m} w_j x_{i,j}$
- Goal: Find the best-fit model w.r.t. loss function ℓ

  with respect to

  $w^* \in \underset{w}{\operatorname{argmin}}\ \ell(w)$

  ↳ loss function.
- least squares loss: $\ell(w) = \sum_{i=1}^{n} (y_i - f_w(x_i))^2$