Linear Regression
- Assume a parametric form

\[ f_w(x_i) = w_1 x_{i,1} + w_2 x_{i,2} + \ldots + w_m x_{i,m} \]

\[ = \sum_{j=1}^{m} w_j x_{i,j} \rightarrow x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,m}) \]

\[ = w^T x_i \]

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

\[ m \leftarrow m + 1 \]

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

\[ \text{new m, equals old m, plus one.} \]

\[ \hat{y}_{n+1} = \text{model}(x_{n+1}) \]

\[ \text{slope always} = 1 \]

Recall:
\[ f_w(x_i) = w_1 x_{i,1} + w_2 x_{i,2} \]

\[ (x_{i,j}, y_i)_{i=1}^{n} \]

\[ x_i \in \mathbb{R}^m \]
\[ y_i \in \mathbb{R} \]
\[ x_{n+1} \rightarrow \hat{y}_{n+1} \]

\[ \hat{y}_{n+1} = f_w(x_{n+1}) \]

- Linear models
- Optimization perspective.
- BBO
Final grade

$Z \in \mathbb{D}$

$m = 4$

$x_{i,1} = \text{HW1 score}$

$x_{i,2} = 1$

\[ f_{w}(x_i) = w_1 x_{i,1} + w_2 x_{i,2} \]

\(<1, 0> \rightarrow 1 \]

\(<0, 1> \rightarrow 2 \]

\(<1, 1> \rightarrow 3 \]
\[ y_i = f(x_i) \]
\[ f(x_i) = \omega_1 x_i + \omega_2 \]
\[ f_w(x_i) = \frac{1}{m} \sum_{j=1}^{m} \omega_j x_{i,j} = \omega_1 x_{i,1} + \omega_2 x_{i,2} \]
\[ X_{i,2} = \frac{1}{\omega_1} \]

\( ^* \) m = 1
b# of features.
Example:

\[ m = 2 \]

\[ x_{i,z} = 1 \text{ always.} \]

\[ \text{Slope} = \frac{\text{rise}}{\text{run}} \]

COVID Deaths per 100k people.

0 mask-wearing 1 proportion of population.

\[ w_1 = -150 \quad w_2 = 200 \]

- 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 - \hat{f}(x_i)
= y_i - (w_1 x_1 + w_2 x_2)
```

Best fit = smallest residual magnitude

which is better:

<table>
<thead>
<tr>
<th></th>
<th>line 1</th>
<th>line 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>-20</td>
<td>-1</td>
</tr>
</tbody>
</table>

take \text{abs}(r_i)
loss function \( l(model) \) = how bad the model is.
\[ l(model) = \sum_{i=1}^{n} r_i \times \text{negative } r_i \text{ not good.} \]
\[ l(model) = \sum_{i=1}^{n} r_i^2 \]

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

[Diagram showing a graph with residual on the y-axis and square of residual on the x-axis.]
\[ l(w) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - f_w(x_i))^2 \]

Model parameters that define a model

Hyperparameters so far:
- Model parameterization (linear)
- Choice of loss function

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

\[ w^* \in \text{argmin} \ l(w) \]

"best", "optimal"

"in" returning \( w \rightarrow \text{TR}^m \)

-called an "optimization problem"
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 \(\ell\)

\[ w^* = \arg \min_{w} \ell(w) \]

- Least squares loss: \(\ell(w) = \frac{1}{n} \sum_{i=1}^n (y_i - f_w(x_i))^2\)