Foda LII

Linear Regression
Input: \((X, y) = \{(x_i, y_i), (x_2, y_2), \ldots (x_m, y_m)\}

\(X \in \mathbb{R}^{m \times 1}\)
\(y \in \mathbb{R}^m\)

\(X_i \in \mathbb{R}\)
\(y_i \in \mathbb{R}\)

explanatory variable
dependent variable

\(f : \mathbb{R}^d \rightarrow \mathbb{R}^d\)

\(f(x_i) = y_i\)

linear function
\(y_i = a \times x_i + b\)

slope
<table>
<thead>
<tr>
<th>height (in)</th>
<th>weight (lbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>160</td>
</tr>
<tr>
<td>68</td>
<td>170</td>
</tr>
<tr>
<td>60</td>
<td>110</td>
</tr>
<tr>
<td>70</td>
<td>178</td>
</tr>
<tr>
<td>65</td>
<td>155</td>
</tr>
<tr>
<td>61</td>
<td>120</td>
</tr>
<tr>
<td>74</td>
<td>223</td>
</tr>
<tr>
<td>73</td>
<td>215</td>
</tr>
<tr>
<td>75</td>
<td>235</td>
</tr>
<tr>
<td>67</td>
<td>164</td>
</tr>
<tr>
<td>69</td>
<td>?</td>
</tr>
</tbody>
</table>

The table shows the relationship between height (in inches) and weight (in pounds) for a group of individuals. The graph on the right illustrates a linear function of the form $y = ax + b$, with a predicted weight for a height of 69 inches, $f(69) = 182$ pounds.
Measure Error

residual for model \( l: \mathbb{R} \rightarrow \mathbb{R} \)
on data point \((x_i, y_i)\)

\[ r_i = (y_i - \hat{y}_i) = (y_i - l(x_i)) \]

units of \( r_i \) same as \( y_i \)
Overall error

Sum \( \sum_{i=1}^{n} r_i < 0 \)

\( r_i = y_i - \hat{y}_i \)

\( \|r\|_1 = \sum_{i=1}^{n} |r_i| \)

Sum of Squared Errors

\( \sum_{i=1}^{n} r_i^2 \)

Closed form algorithm

\( \text{SSF}(x, s, l) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - l(x_i))^2 \)

\( y_i = l(x_i) + \epsilon \sim N(0, \sigma^2) \)

\( l(x) = ax + b \)
Input \((x, y) = \{(x_i, y_i), \ldots, (x_n, y_n)\} \in \mathbb{R} \times \mathbb{R}\)

Goal Find \(l(x) = ax + b\) value \((a, b)\)

Do minimize 

\[ g(a, b) = \sum_{i=1}^{n} (y_i - (a x_i + b))^2 \]
Solving for $a, b$

1. \[ \tilde{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]
2. \[ \tilde{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \]

- Normalize data (center)

\[ \tilde{x} = (x_1 - \tilde{x}, x_2 - \tilde{x}, \ldots, x_n - \tilde{x}) \]
\[ \tilde{y} = (y_1 - \tilde{y}, y_2 - \tilde{y}, \ldots, y_n - \tilde{y}) \]

1. Calculate the angle $\theta = \arccos(\langle \tilde{x}, \tilde{y} \rangle)$
2. Calculate $b = \tilde{y} - a \tilde{x}$

where $\| \tilde{x} \| \| \tilde{y} \| \cos(\theta) = \frac{|\tilde{y}|}{\| \tilde{x} \|} \cos(\theta)$