FoDA

L11

. Linear Regression

. explanatory & dependent variables
Data labeled data \((X, y)\) actual observations

\(X \in \mathbb{R}^n\) and \(y \in \mathbb{R}^n\)

\((X, y) = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}\)

\(X = \text{explanatory variable}\)

\(y = \text{dependent variable}\)

\(y = \hat{\theta}(x) = ax + b\)

\(a, b\) - slope and intercept.
• Measure Error based on Prediction
(want to use on data we don't have yet!)
How to Measure Error?

**Residual**

\[ y = l(x) \]

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

**Sum of Squared Errors**

\[ \text{SSE}(\{x_i, y_i\}) = \sum_{i=1}^{N} (y_i - y_c)^2 \]

Example:

\[ y_i = 3 \]
\[ y_c = 4 \]
\[ r_i = 3 - 4 = -1 \]
Sum of Squared Errors

\[
SSE(x, g) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

Why?

- Squaring makes non-negative.
- Norm (L2-norm) \( \| \mathbf{r} \|^2 = r_1^2 + r_2^2 + \ldots \)

\( \hat{y}_i \) is the predicted value, \( y_i \) is the true label, \( x \) is the model, \( n \) is the number of data points.
Start w/ Bayesian Inference

Assume Normal Noise on $y$: $N(l(x_i), \sigma^2)$

$\rightarrow$ Negative Log-Likelihood

$\rightarrow$ SSE $((x_i, y_i), \lambda)$

Easy to Solve for optimal $(a^*, b^*) = \arg\min_{a,b} \text{SSE}(x_i, y_i, \lambda)$

$\Rightarrow$ "closed form"

$\Rightarrow$ gradient descent $\Rightarrow$ convex
How to Solve for $a^*, b^*$:

$$a^*, b^* = \underset{a, b \in \mathbb{R}}{\text{arg min}} \quad SSE((x, y), (a, b))$$

$$l_{a,b}(x) = a x + b$$

1. **average**
   $$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

2. **set** "center"  
   $$\bar{x} = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x})$$
   $$\bar{y} = (y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y})$$

3. **a**
   $$a^* = \frac{\langle \bar{y}, \bar{x} \rangle}{\sum_{i=1}^{n} \bar{x}_i^2} = \frac{\sum_{i=1}^{n} \bar{y}_i \bar{x}_i}{\sum_{i=1}^{n} \bar{x}_i} \cos \theta$$

   $$b^* = \bar{y} - a^* \bar{x}$$