Linear Regression

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul
SLIDES ADAPTED FROM LAUREN HANNAH
Data are the set of inputs and outputs, $\mathcal{D} = \{ (x_i, y_i) \}_{i=1}^{n}$
In linear regression, the goal is to predict $y$ from $x$ using a linear function.
Examples of linear regression:

- given a child’s age and gender, what is his/her height?
- given unemployment, inflation, number of wars, and economic growth, what will the president’s approval rating be?
- given a browsing history, how long will a user stay on a page?
Linear Regression

\[ (x_i, y_i) \]

\[ f(x) = \beta_0 + \beta_1 x \]
Multiple Covariates

Often, we have a vector of inputs where each represents a different feature of the data

\[ \mathbf{x} = (x_1, \ldots, x_p) \]

The function fitted to the response is a linear combination of the covariates

\[ f(\mathbf{x}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j \]
Multiple Covariates

- Often, it is convenient to represent $\mathbf{x}$ as $(1, x_1, \ldots, x_p)$
- In this case $\mathbf{x}$ is a vector, and so is $\mathbf{\beta}$ (we’ll represent them in bold face)
- This is the dot product between these two vectors
- This then becomes a sum (this should be familiar!)

\[
\mathbf{\beta} \mathbf{x} = \beta_0 + \sum_{j=1}^{p} \beta_j x_j
\]
Hyperplanes: Linear Functions in Multiple Dimensions

Hyperplane
Covariates

- Do not need to be raw value of $x_1, x_2, \ldots$
- Can be any feature or function of the data:
  - Transformations like $x_2 = \log(x_1)$ or $x_2 = \cos(x_1)$
  - Basis expansions like $x_2 = x_1^2$, $x_3 = x_1^3$, $x_4 = x_1^4$, etc
  - Indicators of events like $x_2 = 1\{−1 \leq x_1 \leq 1\}$
  - Interactions between variables like $x_3 = x_1 x_2$
- Because of its simplicity and flexibility, it is one of the most widely implemented regression techniques
Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates
- We just find the point on the line that corresponds to the new input:

$$\hat{y} = \beta_0 + \beta_1 x$$ (1)
Prediction

• After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates
• We just find the point on the line that corresponds to the new input:

$$\hat{y} = \beta_0 + \beta_1 x$$ (1)
Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates.
- We just find the point on the line that corresponds to the new input:

$$\hat{y} = 1.0 + 0.5x$$ (1)

$$y = 1.0 + 0.5x$$
Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates.
- We just find the point on the line that corresponds to the new input:

$$\hat{y} = 1.0 + 0.5 \times 5$$  \hspace{1cm} (1)
Prediction

- After finding $\hat{\beta}$, we would like to predict an output value for a new set of covariates
- We just find the point on the line that corresponds to the new input:
  \[ \hat{y} = 3.5 \]

\[ y = 1.0 + 0.5x \]
\[ x = 5.0 \]

$\hat{y}$ = 3.5 (1)
Outline

Example
Example: Old Faithful
Example: Old Faithful

We will predict the time that we will have to wait to see the next eruption given the duration of the current eruption.

![Graph showing the relationship between current eruption time and waiting time.](image-url)