STATS / DATA SCI 315
Lecture 02

Basic Elements of Linear Regression

Kinds of ML problems

- *Supervised learning*: have labeled data
    - Regression: Label is a real number
    - Classification: Label is chosen from a finite set
- *Unsupervised learning*: have only unlabeled data
- *Reinforcement learning*: interact with an environment and receive rewards

Regression

- Modeling the relationship between input variables and a real-valued output
- Input variables also called:
  - Features
  - Covariates
  - independent variables
- Output variable also called:
  - Label
  - Target
  - Dependent variable
- ``Regress y on **x**" means "Run a regression with **x** as input, y as output"

Examples

- predicting prices (of homes, think of Zillow's Z-estimate)
- predicting length of stay (for patients in the hospital)
- demand forecasting (for retail sales)

Linear regression

- A special but important case of regression
- Model the relationship of y, the output variable, as linear in **x**
- Wish to estimate the prices of houses (in dollars) based on their area (in square feet) and age (in years)
- *Linearity assumption*: target (price) can be expressed as a weighted sum of the features (area and age):

$$\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b$$

Weights and bias

- w_{area} and w_{age} are called *weights*
- b is called a *bias* (also called an *offset* or *intercept*)
- Strictly speaking, our model for price involves an *affine transformation*
- Affine = Linear + bias
- What makes a model good?
- How do we find good values for the weights and bias?

Training Dataset

- Need a dataset where we know the sale price, area, and age for each home
- This is called a *training dataset* or *training set*
- Put one sale info on each row
- Each row is called an *example* (or *data point*, *data instance*, *sample*)
- Each example has
    - A label (price)
    - Features (area, age)

Choosing weights and bias based on training data

- Choose the weights and the bias such that our model predictions best fit the true prices observed in the data
- Long form of our linear model:

  $$\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b$$

- If we had *d* features instead of just two:

  $$\hat{y} = w_1 x_1 + \ldots + w_d x_d + b$$

- The “hat” on top of y denotes that it is an estimate

More compact notation

- Collect all features into a vector $\mathbf{x} \in \mathbb{R}^d$ and all weights into a vector $\mathbf{w} \in \mathbb{R}^d$
- Use dot product to express model compactly:

$$\hat{y} = \mathbf{w}^\top \mathbf{x} + b$$

- Entire dataset of n examples is referred to as the *design matrix* $\mathbf{X} \in \mathbb{R}^{n \times d}$
- $\mathbf{X}$ contains one row for every example and one column for every feature
- Prediction vector $\hat{\mathbf{y}} \in \mathbb{R}^n$ can be expressed via the matrix-vector product:

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w} + b$$