Midterm Review

Hongchang Gao

Spring 2024

1. Vector

- Definition
- Operation
  - Addition,
  - Subtraction,
  - Multiplication
- Norm
  - Three norm
- ~~Orthogonal Vectors~~

2. Matrix

- Definition
- Operation: dimension
  - Addition
  - Subtraction
  - Multiplication
  - Transposition
- ~~Orthogonal matrix~~
- ~~Symmetric matrix~~

3. Exploratory Data Analysis

- Visualize numerical features
  - Line plot
  - Hist plot
  - Boxplot
- Long tail vs non-long tail distribution
- Explore pairs of features
- ~~Visualize categorical features~~

4. Data Preprocessing

- Missing values
  - Numerical features
  - Categorical features
- Convert categorical features to numerical features
  - Label encoding
  - One-hot encoding
  - Ordinal encoding
- Feature scaling/normalization
  - Min-max normalization
  - Z-score normalization

5. Linear regression

- Supervised Learning vs unsupervised learning
- Regression vs classification
- Build model
- Loss function $\min_w \frac{1}{n}\|y - Xw\|_2^2$
  - Ridge regression, Lasso
- Evaluation
  - MAE
  - MSE
  - RMSE
- Overfitting

6. Logistic regression

- Build model
  - Sigmoid function
  - Softmax function
- Loss function
  - Cross-entropy loss function

$$\min_{\mathbf{w}} -\sum_{i=1}^{n} \{y_i \log p_i + (1 - y_i) \log(1 - p_i)\}$$

- Evaluation (Binary classification & multi-class classification)
  - Accuracy
  - Recall, Precision
  - Micro vs macro
- Model selection
  - Threefold split
  - Cross-validation

Midterm Exam

- 9:30 – 10:50 am, Feb. 29, BEURY, 00162
- Closed-book exam
- Question set 1:
  - 8 questions, 3 points per question
  - True or false, multiple choice
- Question set 2:
  - 7 questions, 3 points per question
  - Fill in the blank
- Question set 3:
  - 3 questions, 5 points per question