

Assignment 3: Classification, Logistic Regression, and Gradient Descent

Machine Learning

Fall 2019

💡 Learning Objectives

- Learn about the framing of the classification problem in machine learning.
- Learn about the logistic regression algorithm.
- Learn about gradient descent for optimization.
- Some C&E topic.

🔄 Prior Knowledge Utilized

- Supervised learning problem framing.
- Training / testing splits.

1 The Classification Problem

2 Perceptron?

3 Top-down View of Logistic Regression

4 Mathematical Foundations

4.1 Probability

4.2 Logistic function

4.3 Log-loss

4.4 Chain Rule for Gradients

5 Gradient Descent

5.1 Visualization

6 Algorithm Derivation

Todo: this is easier with the identities of the derivative of a logistic function.

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} e(\mathbf{w}) \quad (1)$$

$$e(\mathbf{w}) = \sum_{i=1}^n y_i \log \frac{1}{1 + e^{-\mathbf{w}^\top \mathbf{x}_i}} + (1 - y_i) \log \frac{1}{1 + e^{\mathbf{w}^\top \mathbf{x}_i}} \quad (2)$$

$$= \arg \min_{\mathbf{w}} \sum_{i=1}^n -y_i \log \left(1 + e^{-\mathbf{w}^\top \mathbf{x}_i} \right) - (1 - y_i) \log \left(1 + e^{\mathbf{w}^\top \mathbf{x}_i} \right) \quad (3)$$

$$\nabla e(\mathbf{w}) = \sum_{i=1}^n \frac{y_i \mathbf{x}_i}{1 + e^{-\mathbf{w}^\top \mathbf{x}_i}} - \frac{(1 - y_i) \mathbf{x}_i}{1 + e^{\mathbf{w}^\top \mathbf{x}_i}} \quad (4)$$

$$= \sum_{i=1}^n \mathbf{x}_i \left(\frac{y_i}{1 + e^{-\mathbf{w}^\top \mathbf{x}_i}} - \frac{(1 - y_i)}{1 + e^{\mathbf{w}^\top \mathbf{x}_i}} \right) \quad (5)$$