University of California, Los Angeles — Assignment 3
Winter 2025

Math 156: Machine Learning

Instructions. Please submit a pdf with neatly written solutions. See the syllabus for instructions on coding problems.

1. Book Exercise 4.1
2. Book Exercise 4.14
3. Implement a program to train a binary logistic regression model using mini-batch SGD. Use the logistic regression model we derived in class, corresponding to Equation (4.90) from the textbook, and where the feature transformation $\boldsymbol{\phi}$ is the identity function.

   The program should include the following hyperparameters:

   - Batch size
   - Fixed learning rate
   - Maximum number of iterations

4. In this problem, you will run a logistic regression model for classification on a breast cancer dataset.

   (a) Download the Wisconsin Breast Cancer dataset from the UCI Machine Learning Repository [1] or scikit-learn's built-in datasets [2].

   (b) Split the dataset into train, validation, and test sets.

   (c) Report the size of each class in your training (+ validation) set.

   (d) Train a binary logistic regression model using your implementation from problem 3. Initialize the model weights randomly, sampling from a standard Gaussian distribution. Experiment with different choices of fixed learning rate and batch size.

   (e) Use the trained model to report the performance of the model on the test set. For evaluation metrics, use accuracy, precision, recall, and F1-score.

   (f) Summarize your findings.

Suggested problems (not for submission)

1. Book Exercise 4.12
2. Consider the model $y(\mathbf{x}, \mathbf{w}) = \mathrm{sign}(\mathbf{w}^\top \mathbf{x})$. Suppose a single sample $\mathbf{x}_1$ with target value $t_1 = 1$ is given. Show that the set of all weight vectors $\mathbf{w}$ that correctly classifies this point is convex.

[1] `https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic`
[2] `https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html`

3. Given i.i.d. data $D = \{(\mathbf{x}_1, t_1), \cdots, (\mathbf{x}_N, t_N)\}$. Consider the logistic regression model with a target variable $t \in \{-1, 1\}$. If we define $p(t = 1|y) = \sigma(y)$ where

$$y(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}) + b,$$

$\sigma(a) = \frac{1}{1+e^{-a}}$ is the logistic sigmoid function and $\boldsymbol{\phi}$ denotes a fixed feature-space transformation. Show that the negative log likelihood, with the addition of a quadratic regularization term, takes the form

$$\sum_{n=1}^{N} \ln(1 + e^{-y(\mathbf{x}_n)t_n}) + \lambda \|\mathbf{w}\|^2.$$

4. Book Exercises 6.5, 6.7, 6.8