SAAS CX Homework 9

§1 Introduction

Problem 1.1. Whereas regression involves assigning each data point a continuous value, classification involves assigning each data point to a discrete class.

Problem 1.2. Linear regression deals with continuous values with unbounded range, whereas in classification we are looking for discrete classes (for example, 0 or 1 values).

Problem 1.3. Generative models attempt to fully model all variables probabilistically (i.e., assumming each data point is draw from a class-conditional and modeling the distribution within each class), whereas discriminative models only predicts the label (either by estimating posterior probability directly or simply creating a decision boundary).

Problem 1.4. For each of the following models, (i) categorize the model as generative or discriminative, (ii) describe how the decision boundary is determined, (iii) identify how optimization is performed to determine the solution for the parameters / weights.

- Logistic Regression
  - (i) discriminative
  - (ii) calculate predicted probabilties using linear prediction and the sigmoid function
  - (iii) Gradient Descent on the cost function involving cross entropy loss.
- Gaussian Discriminant Analysis
  - (i) generative
  - (ii) model each class as a Gaussian distribution
  - (iii) simply compute empirical means and covariances using maximum likelihood estimation (MLE)
- Support Vector Machine
  - (i) discriminative
  - (ii) directly create the decision boundary by maximizing the objective function involving the margin
  - (iii) solving the Quadratic Program optimization problem (for example, using interior point method)

Problem 1.5. In logistic regression, we utilize the cost function involving the cross entropy loss.

$$L = -\sum_{i=1}^{n}(y_i ln(p_i) + (1 - y_i) ln(1 - p_i))$$

where $p_i = s\left(X_i \cdot w\right)$ is the posterior probability computed from taking the sigmoid of the linear combination of features.