The Elements of Statistical Learning

November 22, 2020

Linear Methods for Classification

Core concepts covered in this chapter:

- Decisions boundaries
- Indicator Matrices
- Masking
- Linear Discriminant Analysis
- Quadratic Discriminant Analysis

We call this set of classification procedures linear because their decision boundaries are linear. A function mapping an input x to a class k is called *discriminant functions* $\delta_k(x)$. Decision boundaries are defined by the set $\{x : \delta_k(x) = 0\}$. The decision boundary is linear if δ_k or some monotone transformation of δ_k are linear (*logit* transformations for instance).

Linear Regression for an Indicator Matrix

The goal here is to fit a transformed linear regression model to predict a categorical variable. Y is here a $N \times K$ *indicator response matrix* where $\forall i \in \{1, \dots, n\}, \forall k \in \{1, ..., K\}, Y_{i,k} = 1$ *if* $G = k$ *and 0 otherwise.* Hence Y is a matrix of 0's and 1's with each row having a single 1. The model is fitted as usual and we find an regressor $\hat{\beta} = (X^T X)^{-1} X^T Y$ of size $p + 1 \times K$. To classify a new observation x, we compute the fitted output $\hat{f}(x) = (1, x^T)\hat{\beta}$, a K vector. Classify the observation according to the largest component of $\hat{f}(x)$: $\hat{G}(x) = argmax_{k \in \mathcal{G}} \hat{f}_k(x)$. One important issue this this method when $K \geq 3$ is that some classes maybe masked by others because of the rigidity of the model. The goal of the following methods is to keep decisions boundaries linear while avoiding classes masking each other.

Linear Discriminant Analysis

What we are interested in is $P(G = k|X = x)$. The law of $G|X$ is unknown but we use Bayes theorem and class densities to estimate it. We denote $\pi_k = P(G = k)$ the prior law of G and $f_k(x) = P(X = x|G = k)$ the class-conditional density of X in class $G = k$. Bayes theorem gives us $P(G = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{i=1}^{K} \pi_i f_i(x)}$.

Regularized Discriminant Analysis

Computations for LDA

Reduced-Rank Linear Discriminant Analysis

Logistic Regression

Fitting Logistic Regression Models

Quadratic Approximations and Inference

L_1 Regularized Logistic Regression

Logistic Regression or LDA?

Separating Hyperplanes

Rosenblatt's Perceptron Learning Algorithm

Optimal Separating Hyperplanes