The aim of this work is to derive maximum likelihood estimates when learning standard simple generative and discriminative classifiers. The considered generative approach for classification is the Gaussian Discriminant Analysis, and the discriminative one consists in the binary logistic regression classifier.

Learning of a Gaussian density model

The objective here is to use Gaussian Discriminant Analysis for data classification. Both Linear and Quadratic Discriminant analysis assume a Gaussian density as a pdf for each class. The parameter of each class k have then to be estimated from the training data set of each class.

Consider the problem of QDA (LDA can be derived in a closely similar way) for multivariate data where each class has a Gaussian pdf described by its own mean vector $\boldsymbol{\mu}_k$ and co-variance matrix $\boldsymbol{\Sigma}_k$, given by

$$f\left(\mathbf{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \mathcal{N}\left(\mathbf{x}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \frac{1}{(2\pi)^{d/2} \left|\boldsymbol{\Sigma}_k\right|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{x}_i - \boldsymbol{\mu}_k\right)^T \boldsymbol{\Sigma}_k^{-1}\left(\mathbf{x}_i - \boldsymbol{\mu}_k\right)\right). \quad (1)$$

1. Given an i.i.d set of multidimensional Gaussian observations $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, derive the ML estimates for $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$.
2. Comment the bias of the maximum likelihood estimators ? how to correct the possible bias for the covariance estimator (you can use results seen last year in the the case of ML estimators for univariate Gaussian)

Learning of a binary logistic regression model

Logistic regression is discriminative classifier. It therefore directly models $p(y|x)$. Consider the simple binary case where the output (class) $y \in \{0, 1\}$. The binary logistic regression model is thus given by the following sigmoid function

$$p(y = 1|\mathbf{x}_i; \boldsymbol{w}) = \pi(\mathbf{x}_i; \boldsymbol{w}) = \frac{\exp(\boldsymbol{w}^T\mathbf{x}_i)}{1 + \exp(\boldsymbol{w}^T\mathbf{x}_i)}. \quad (2)$$

The objective is to fit $\boldsymbol{w}$ for learn this model from a labeled data set.

Given a training set of independent observations $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, derive the ML estimate for $\boldsymbol{w}$. Note that in this binary case, the conditional probability of class 0 is given by $p(y = 0|\mathbf{x}_i; \boldsymbol{w}) = 1 - \pi(\mathbf{x}_i; \boldsymbol{w})$. You shall use the IRLS algorithm (Newton-Raphson) to maximize the conditional log-likelihood.