Prof. Dr. Rudolf Mathar, Dr. Gholamreza Alirezaei, Dr. Arash Behboodi

Exercise 10
- Proposed Solution -

Friday, January 27, 2017

Solution of Problem 1

a)

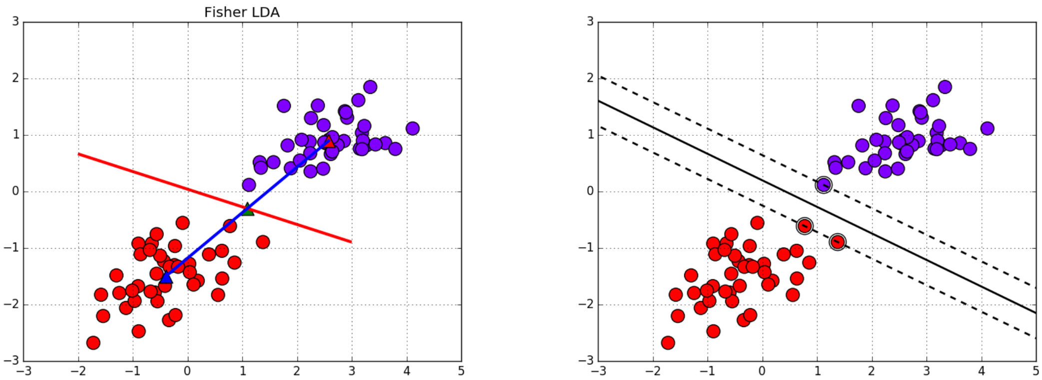

Figure 1: Two separable classes: SVM versus LDA

b) The constant C in the soft margin classifier has been chosen equal to 6.

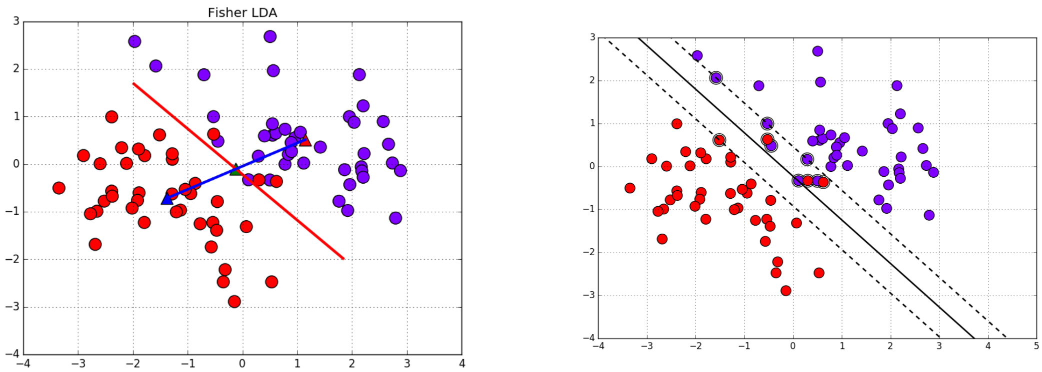

Figure 2: Two non-separable classes: SVM versus LDA

Solution of Problem 2

The constant C in the soft margin classifier has been chosen equal to 6. Note that in this case, unlike ML-discriminant analysis, there is a region which cannot be attributed to a class.

a)

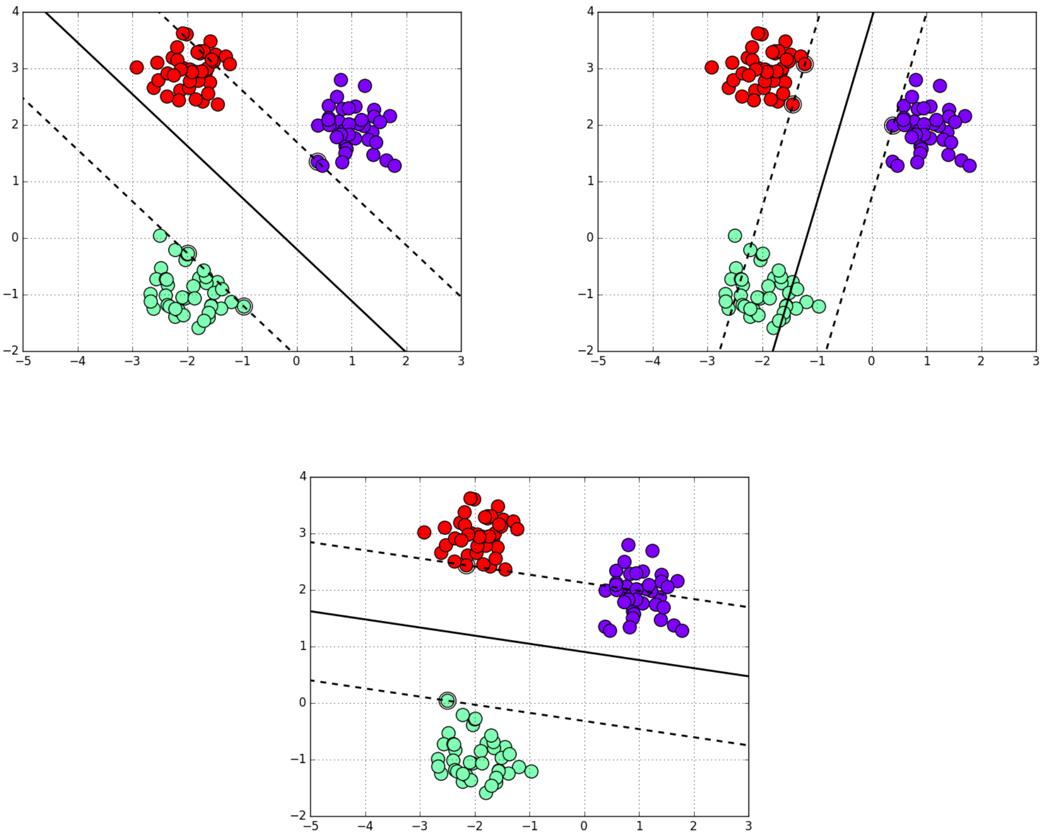

Figure 3: SVM applied to each pair of classes

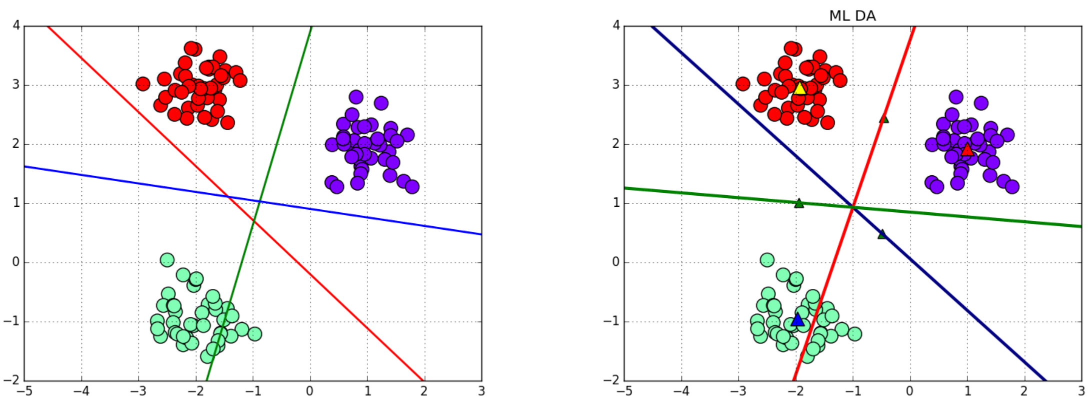

Figure 4: Three separable classes: SVM versus ML

b)

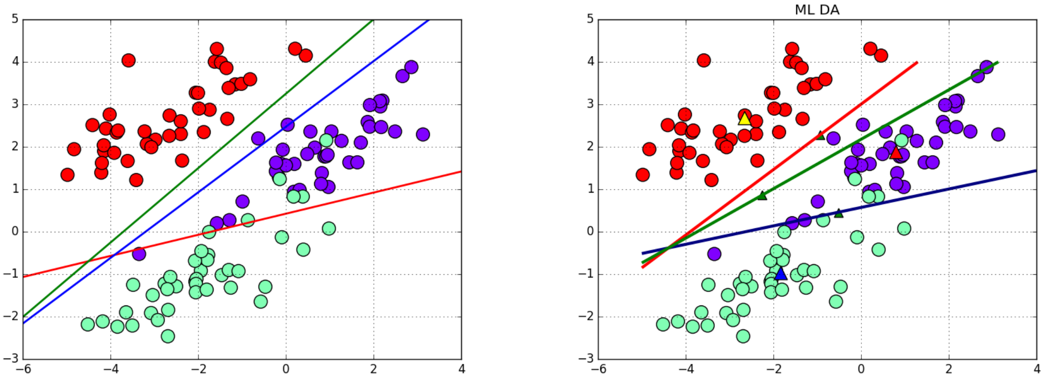

Figure 5: Three non-separable classes: SVM versus ML

Solution of Problem 3

a) The first step is to formulate the Lagrangian function:

$$\Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{a}\|^2 + c\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\lambda_i[y_i(\mathbf{a}^T\mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^{n}\alpha_i\xi_i,$$

where $\lambda_i, \alpha_i \geq 0$ and we know that $y_i = 1$ or -1.

Next step is to find Lagrange dual function:

$$g(\boldsymbol{\lambda}, \boldsymbol{\alpha}) = \inf_{\boldsymbol{\xi}, \mathbf{a}\in\mathbb{R}^n, b\in\mathbb{R}} \Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$$

Taking the derivation with respect to these parameters lead to:

$$\frac{\partial}{\partial \mathbf{a}}\Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = 0 \implies \mathbf{a} = \sum_{i=1}^{n}\lambda_i y_i \mathbf{x}_i$$

$$\frac{\partial}{\partial b}\Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = 0 \implies 0 = \sum_{i=1}^{n}\lambda_i y_i$$

$$\frac{\partial}{\partial \boldsymbol{\xi}}\Lambda(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = 0 \implies c = \lambda_i + \alpha_i \quad i = 1, \ldots, n$$

Using these assumptions, the dual problem can be written as:

$$\begin{aligned} \max \quad & \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\sum_{i,j} y_i y_j \lambda_i \lambda_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t.} \quad & 0 \leq \lambda_i \leq c \\ & \sum_{i=1}^{n}\lambda_i y_i = 0. \end{aligned}$$

b) Take two support vector $\mathbf{x}_k$ and $\mathbf{x}_l$ with $y_k = 1$ and $y_l = -1$. Those support vectors are those with $0 < \lambda < c$. Then since for support vectors we have $y_i(\mathbf{a}^T\mathbf{x}_i + b) = 1$, we have:

$$(\mathbf{a}^{\star T}\mathbf{x}_k + b^\star) = -(\mathbf{a}^{\star T}\mathbf{x}_l + b^\star) \implies b^\star = \frac{-1}{2}\mathbf{a}^{\star T}(\mathbf{x}_k + \mathbf{x}_l) \qquad (1)$$

In litterature, somethimes an average is calculated over all support vectors for better robustness.