

CO-496 Coursework #4

Zhenhao Li

November 29, 2018

Part I

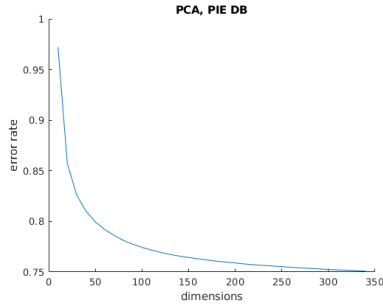


Figure 1: PCA

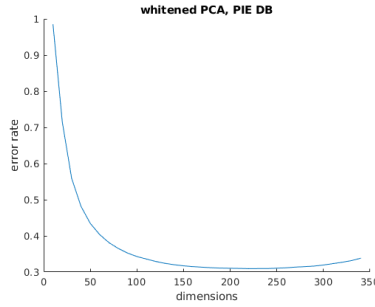


Figure 2: wPCA

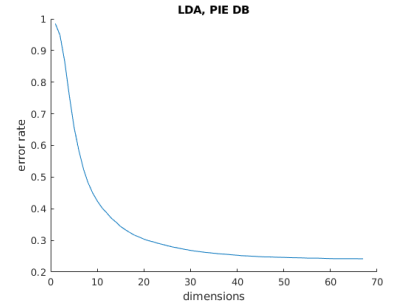


Figure 3: LDA

- From the three plots, LDA achieves the best performance and the error rate reduces to below 0.3 .
- The wPCA method performs better than the PCA method because it normalizes the data, setting the covariance matrix of the data equal to identity matrix. However, when the number of dimensions is over 250, the error rates of wPCA starts to raise back. This is possibly because the features of trivial eigenvalues are taken into consideration and they plays few help in representing the data. Since wPCA normalizes features on every dimension, these features are of same importance as those really "important" features.
- The input labels are important in reducing the error rate of LDA. PCA or wPCA are unsupervised methods while LDA is a supervised method. Therefore, with the same dataset, LDA can utilize more information than the former two, which results in lower error rates.

Part II

1. Formulate the Lagrangian of the above optimisation problem and derive its dual. Attach your answers in the report. Show the steps that you followed, do not only report the final results.

The Lagrangian is

$$L(R, \mathbf{a}, \xi_i, t_i, r_i) = R^2 + C \sum_{i=1}^n \xi_i + \sum_{i=1}^n t_i ((\mathbf{x}_i - \mathbf{a})^T (\mathbf{x}_i - \mathbf{a}) - R^2 - \xi_i) - \sum_{i=1}^n r_i \xi_i \quad (1)$$

with Lagrangian multiplier $t_i \geq 0, r_i \geq 0$. Computing the derivatives

$$\begin{aligned}\frac{\partial L}{\partial R} &= 2R - 2R \sum_{i=1}^n t_i = 0 \\ \frac{\partial L}{\partial \mathbf{a}} &= \sum_{i=1}^n t_i (2\mathbf{a} - 2\mathbf{x}_i) = 0 \\ \frac{\partial L}{\partial \xi_i} &= C - t_i - r_i = 0\end{aligned}\tag{2}$$

Substituting (2) back into (1) we get the dual optimisation problem

$$\begin{aligned}\max_{\mathbf{t}} L(\mathbf{t}) &= \mathbf{k}^T \mathbf{t} - \mathbf{t}^T \mathbf{X}^T \mathbf{X} \mathbf{t} \\ \text{subject to } \mathbf{t}^T \mathbf{1} &= 1, 0 \leq t_i \leq C \\ \text{where } \mathbf{k} &= [\|x_1\|_2^2, \dots, \|x_n\|_2^2]^T\end{aligned}\tag{3}$$

2. Perform the above when using arbitrary positive definite kernels (i.e., $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$), i.e. find the dual of the following optimisation problem

$$\begin{aligned}\min_{R, \mathbf{a}, \xi_i} R^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to } (\phi(\mathbf{x}_i) - \mathbf{a})^T (\phi(\mathbf{x}_i) - \mathbf{a}) &\leq R^2 + \xi_i, \forall \xi_i \geq 0\end{aligned}$$

Attach your answers in the report. Show the steps that you followed, do not only report the final results.

The Lagrangian is

$$L(R, \mathbf{a}, \xi_i, t_i, r_i) = R^2 + C \sum_{i=1}^n \xi_i + \sum_{i=1}^n t_i ((\phi(\mathbf{x}_i) - \mathbf{a})^T (\phi(\mathbf{x}_i) - \mathbf{a}) - R^2 - \xi_i) - \sum_{i=1}^n r_i \xi_i\tag{4}$$

with Lagrangian multiplier $t_i \geq 0, r_i \geq 0$. Computing the derivatives

$$\begin{aligned}\frac{\partial L}{\partial R} &= 2R - 2R \sum_{i=1}^n t_i = 0 \\ \frac{\partial L}{\partial \mathbf{a}} &= \sum_{i=1}^n t_i (2\mathbf{a} - 2\phi(\mathbf{x}_i)) = 0 \\ \frac{\partial L}{\partial \xi_i} &= C - t_i - r_i = 0\end{aligned}\tag{5}$$

Substituting (5) back into (4) we get the dual optimisation problem

$$\begin{aligned}\max_{\mathbf{t}} L(\mathbf{t}) &= \mathbf{k}^T \mathbf{t} - \mathbf{t}^T \Phi^T \Phi \mathbf{t} \\ \text{subject to } \mathbf{t}^T \mathbf{1} &= 1, 0 \leq t_i \leq C \\ \text{where } \mathbf{k} &= [k(\mathbf{x}_1, \mathbf{x}_1), \dots, k(\mathbf{x}_n, \mathbf{x}_n)]^T\end{aligned}\tag{6}$$

3.

The centre and radius for the two classes are

$$\begin{aligned}\mathbf{a}_{red} &= (0.0037, -0.0343) \\ R_{red} &= 0.9866 \\ \mathbf{a}_{blue} &= (-0.0036, 0.0015) \\ R_{blue} &= 1.9835\end{aligned}$$

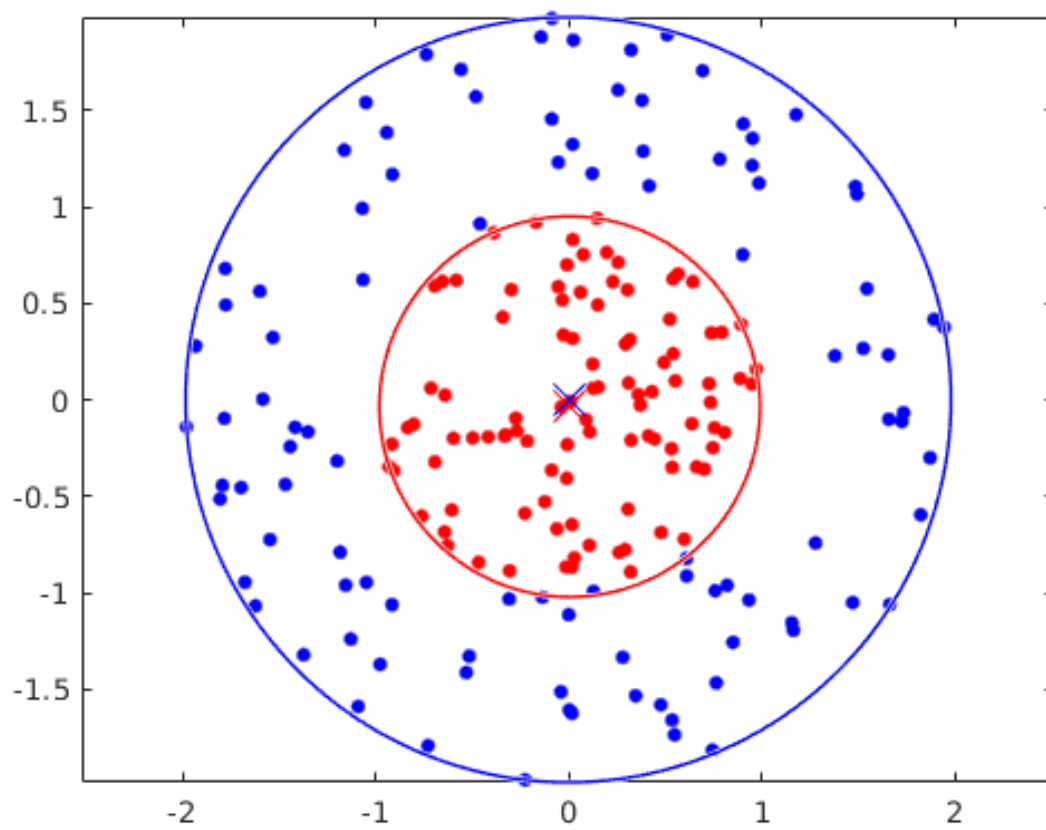


Figure 4: SVM