CO496 Coursework 4

Jinsung Ha

November 30, 2018

Part I:

Plots of recognition error versus the number of components kept are as shown below for each reduction techniques used.

I. Principal Component Analysis

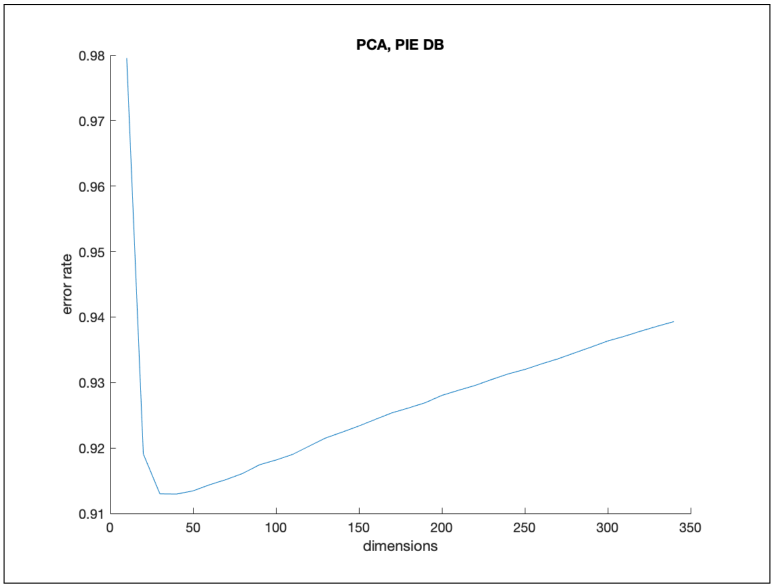

Figure 1: graph of error rate versus dimensions using PCA

II. Whitening Principal Component Analysis

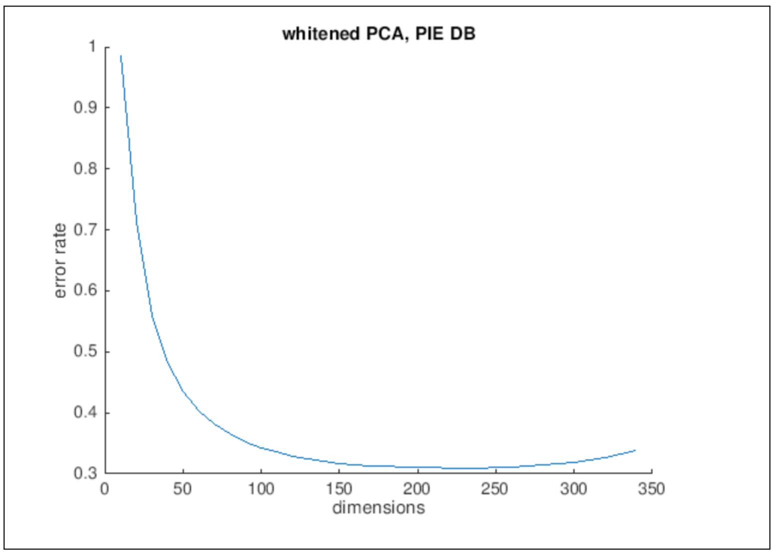

Figure 2: graph of error rate versus dimensions using WPCA

III. Linear Discriminant Analysis

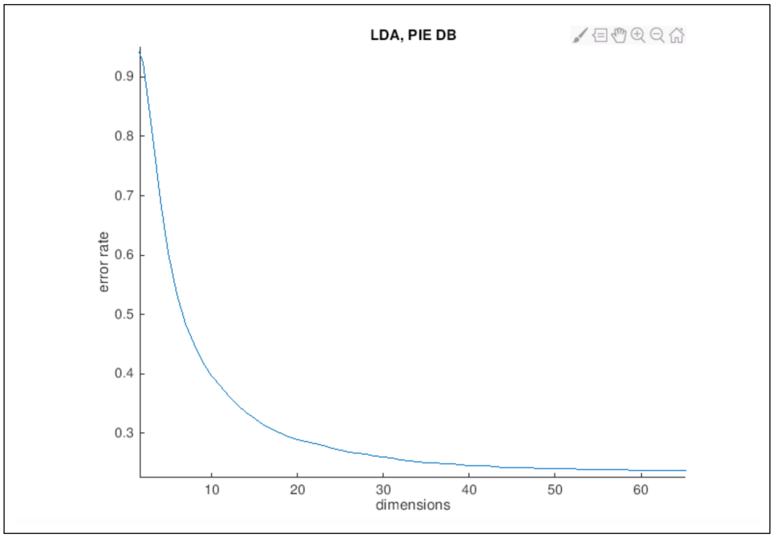

Figure 3: graph of error rate versus dimensions using LDA

From the above 3 plots, we can observe that error rates in all 3 graphs show initial steep decreases. However, for PCA, the error rate decreases very slowly as it reaches around 0.8, whereas in WPCA and LDA that mentioned point is around 0.4. Hence we know that the error rate of PCA is relatively higher than the other two.

Now comparing the WPCA and LDA, we can observe that graph of LDA reaches the relatively stable error rate of 0.4 with just dimension = 10, whereas for WPCA it took 50 dimensions. This means that LDA can have relatively similar precision to WPCA with much lesser dimensions, and hence we can conclude that the LDA technique is the best among the 3 dimensionality reduction techniques.

Part II:

1. Given constrained optimization problem

$$\begin{aligned} \min_{R,\mathbf{a},\zeta_i} \quad & R^2 + C\sum_{i=1}^{n}\zeta_i \\ \text{subject to} \quad & (\mathbf{x}_i-\mathbf{a})^T(\mathbf{x}_i-\mathbf{a}) \le R^2+\zeta_i \\ & \forall \zeta_i \ge 0. \end{aligned}$$

Then the Lagrangian can be formulated as such.

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

with Lagrangian multipliers $b_i \ge 0, r_i \ge 0$. Computing the derivatives

$$\begin{aligned} \frac{\partial L}{\partial R} &= 2R - 2R\sum_{i=1}^{n} b_i = 0 && \Rightarrow \sum_{i=1}^{n} b_i = 1 \\ \frac{\partial L}{\partial \mathbf{a}} &= \sum_{i=1}^{n} b_i(-2)(\mathbf{x}_i-\mathbf{a})^T = \mathbf{0} && \Rightarrow \mathbf{a}^T = \frac{1}{\sum_{i=1}^{n} b_i}\sum_{i=1}^{n} b_i\mathbf{x}_i^T = \frac{1}{1}\mathbf{b}^T\mathbf{X}^T \\ \frac{\partial L}{\partial \zeta_i} &= C - b_i - r_i = 0 \end{aligned}$$

Substituting above equations back to the dual expression,

$$\begin{aligned} L(\mathbf{b}) &= \sum_{i=1}^{n} b_i\Big((\mathbf{x}_i-\mathbf{a})^T(\mathbf{x}_i-\mathbf{a})\Big) \\ &= \sum_{i=1}^{n} b_i\Big(\mathbf{x}_i^T\mathbf{x}_i - \mathbf{x}_i^T\mathbf{a} - \mathbf{a}^T\mathbf{x}_i + \mathbf{a}^T\mathbf{a}\Big) \\ &= \sum_{i=1}^{n} b_i\mathbf{x}_i^T\mathbf{x}_i - \sum_{i=1}^{n}(b_i\mathbf{x}_i^T)\mathbf{a} - \sum_{i=1}^{n} b_i\mathbf{a}^T\mathbf{x}_i + \sum_{i=1}^{n}(b_i\mathbf{a}^T)\mathbf{a} \\ &= \sum_{i=1}^{n} b_i\mathbf{x}_i^T\mathbf{x}_i - \sum_{i=1}^{n} b_i\mathbf{a}^T\mathbf{x}_i \\ &= \sum_{i=1}^{n} b_i\mathbf{x}_i^T\mathbf{x}_i - \sum_{j=1}^{n}\sum_{i=1}^{n} b_ib_j\mathbf{x}_i^T\mathbf{x}_j \\ &= (\mathbf{K}')^T\mathbf{b} - \mathbf{b}^T\mathbf{K}\mathbf{b} \end{aligned}$$

Finally, we are left with a function of a what we wish to maximise. Putting this together with the constraints, we obtain the following dual optimization problem

$$\begin{aligned} \max_{\mathbf{b}} \quad & (\mathbf{K}')^T\mathbf{b} - \mathbf{b}^T\mathbf{K}\mathbf{b} \\ \text{subject to} \quad & 0 \le b_i \le C \quad i = 1,...,n \\ & \mathbf{1}^T\mathbf{b} = 1, \end{aligned}$$

where $\mathbf{b} = [b_1,...,b_n]^T, \mathbf{1} = [1,...,1]^T, \mathbf{K} = [\mathbf{x}_i^T\mathbf{x}_i]$

2. Now the constrained optimization problem is given as,

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

Then the Lagrangian can be formulated as such.

$$L(R,\mathbf{a},\zeta_i,b_i,r_i) = R^2 + C\sum_{i=1}^{n}\zeta_i + \sum_{i=1}^{n} b_i\Big((\phi(\mathbf{x}_i)-\mathbf{a})^T(\phi(\mathbf{x}_i)-\mathbf{a}) - R^2 - \zeta_i\Big) - \sum_{i=1}^{n} r_i\zeta_i$$

with Lagrangian multipliers $b_i \ge 0, r_i \ge 0$. Computing the derivatives

$$\begin{aligned}
\frac{\partial L}{\partial R} &= 2R - 2R\sum_{i=1}^{n} b_i = 0 && \Rightarrow \sum_{i=1}^{n} b_i = 1 \\
\frac{\partial L}{\partial \mathbf{a}} &= \sum_{i=1}^{n} b_i(-2)(\phi(\mathbf{x}_i)-\mathbf{a})^T = \mathbf{0} && \Rightarrow \mathbf{a}^T = \frac{1}{\sum_{i=1}^{n} b_i}\sum_{i=1}^{n} b_i\phi(\mathbf{x}_i)^T = \frac{1}{1}\mathbf{b}^T\,\phi(\mathbf{X})^T \\
\frac{\partial L}{\partial \zeta_i} &= C - b_i - r_i = 0
\end{aligned}$$

Substituting above equations back to the dual expression,

$$\begin{aligned}
L(\mathbf{b}) &= \sum_{i=1}^{n} b_i\Big((\phi(\mathbf{x}_i)-\mathbf{a})^T(\phi(\mathbf{x}_i)-\mathbf{a})\Big) \\
&= \sum_{i=1}^{n} b_i\Big(\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_i) - \phi(\mathbf{x}_i)^T\mathbf{a} - \mathbf{a}^T\phi(\mathbf{x}_i) + \mathbf{a}^T\mathbf{a}\Big) \\
&= \sum_{i=1}^{n} b_i\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_i) - \sum_{i=1}^{n}(b_i\phi(\mathbf{x}_i)^T)\mathbf{a} - \sum_{i=1}^{n} b_i\mathbf{a}^T\phi(\mathbf{x}_i) + \sum_{i=1}^{n}(b_i\mathbf{a}^T)\mathbf{a} \\
&= \sum_{i=1}^{n} b_i\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_i) - \sum_{i=1}^{n} b_i\mathbf{a}^T\phi(\mathbf{x}_i) \\
&= \sum_{i=1}^{n} b_i\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_i) - \sum_{j=1}^{n}\sum_{i=1}^{n} b_ib_j\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) \\
&= (\mathbf{K}^{'})^T\mathbf{b} - \mathbf{b}^T\mathbf{K}\mathbf{b}
\end{aligned}$$

Finally, we are left with a function of a what we wish to maximise. Putting this together with the constraints, we obtain the following dual optimization problem

$$\begin{aligned}
\max_{\mathbf{b}} \quad & (\mathbf{K}^{'})^T\mathbf{b} - \mathbf{b}^T\mathbf{K}\mathbf{b} \\
\text{subject to} \quad & 0 \le b_i \le C, \quad i = 1, ..., n \\
& \mathbf{1}^T\mathbf{b} = 0,
\end{aligned}$$

where $\mathbf{b} = [b_1, ..., b_n]^T, \mathbf{1} = [1, ..., 1]^T, \mathbf{K} \equiv [k(\mathbf{x}_i, \mathbf{x}_i)],$ and $k(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$ is the kernel.

3. Since quadprog function solves generic quadratic programming optimization problems of the form:

$$\begin{aligned} \min_{\mathbf{b}} \quad & \mathbf{f}^T\mathbf{g} + \frac{1}{2}\mathbf{g}^T\mathbf{H}\mathbf{g} \\ \text{subject to} \quad & \mathbf{A}\mathbf{g} \leq \mathbf{c}, \mathbf{A}_e\mathbf{g} = c_e,\, \mathbf{g}_l \leq \mathbf{g} \leq \mathbf{g}_u \end{aligned}$$

We rewrite the dual optimization problem found in 2), by

I. changing maximization to minimization by reversing the sign of the cost function

II. setting $\mathbf{g} = \mathbf{b}$

III. setting $\mathbf{H} = [\mathrm{k}(\mathbf{x}_i, \mathbf{x}_i)]$

IV. setting $\mathbf{f}$ = - $(\mathbf{K}^{'})^T$, $\mathbf{A} = \mathbf{0}$ and $\mathbf{c} = \mathbf{0}$ (a dummy inequality constraint), $\mathbf{A}_e = \mathbf{1}^T$ and $c_e = 0$, $\mathbf{g}_l = [0, ..., 0]^T$, and finally $\mathbf{g}_u = [C_1, ..., C_n]^T$.

i.e.

$$\begin{aligned} \min_{\mathbf{b}} \quad & -(\mathbf{K}^{'})^T\mathbf{b} + \mathbf{b}^T\mathbf{K}\mathbf{b} \\ \text{subject to} \quad & 0 \leq b_i \leq C\; i = 1, ..., n \\ & \mathbf{1}^T\mathbf{b} = 0, \end{aligned}$$

by using quadprog with appropriate elements to obtain the minimizing $\mathbf{b}$ value, we could complete the function calcRandCentre to obtain the plot of optimal enclosing hyper-sphere, and its radii and centres.

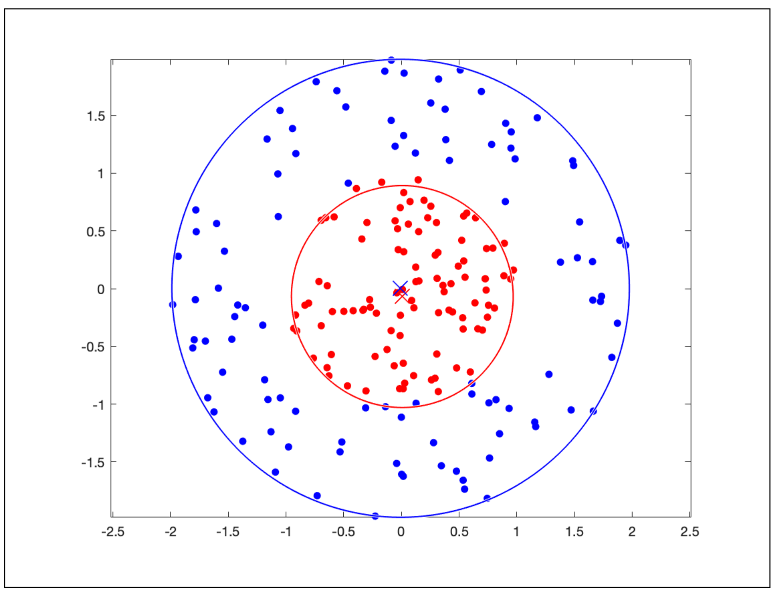

Figure 4: optimal enclosing hyper-sphere

$a_{red} = [0.0074, -0.0686]$
$a_{blue} = [-0.0073, 0.0030]$
$r_{red} = 0.9609$
$r_{blue} = 1.9827$