Lehrstuhl für STEUERUNGS- UND REGELUNGSTECHNIK Technische Universität München Prof. Dongheui Lee	MACHINE LEARNING IN ROBOTICS Exercises 5: PCA and LDA

Exercise 1

We have a mobile robot which has collected images of objects in its surrounding. The robot will use a clustering algorithm for performing groupng of similar images. Since the image data is high dimensional, a useful preprocessing step is to first project it into a lower dimension space before clustering. Now we have $d \times n$ dimensional data X which has n samples of $d \times 1$ dimensional vectors. These vectors correspond to image data and the dimension $d \gg n$. Now we are interested in calculating the principal components of this data. The principal components correpond to eigenvector of covariance matrix of C

$$C = \frac{1}{n-1} * X_c * {X_c}^\top \tag{1}$$

where X_c is obtained by subtracting the mean vector from X. Since for large values of d (for grayscale image of size 640×480, $d = 307200$), this computation can easily hang the onboard system.

A useful trick is to calculate the eigenvector of $C1$ which is $n \times n$

$$C1 = \frac{1}{n-1} * {X_c}^\top * X_c \tag{2}$$

and now if v is an eigenvector of $C1$ with corresponding eigenvalue λ then $v1 = X_c * v$ is an eigenvector of C with corresponding eigenvalue λ. Show that this claim is true.

Solution Exercise 1

If v is an eigenvector of $C1$ then

$C1v = \lambda v \Rightarrow \lambda v = \frac{1}{n-1} * {X_c}^\top * X_c v$

Multiplying bothsides by X_c we get

$\lambda X_c v = \frac{1}{n-1} X_c * {X_c}^\top * X_c v$

$\lambda v1 = \frac{1}{n-1} X_c * {X_c}^\top v1$

$\lambda v1 = C v1$

Exercise 2

Use proof by induction to show that the linear projection onto an M-dimensional subspace that maximizes the variance of the projected data is defined by the M eigenvectors of the data covariance matrix S, corresponding to the M largest eigenvalues.

Solution Exercise 2

Formula for covariance matrix is $\Sigma = E\left[(X - E[X])(X - E[X])^\top\right] = \frac{\sum_{i=1}^{N} (X^{(i)} - E[X])(X^{(i)} - E[X])^\top}{n-1}$

For induction we have first to proof for $M = 1$ (dimensionality reduction to 1 dimension), then assuming that for M (dimensionality reduction to M dimensions) the condition holds, we proove it for $M + 1$.
For $M = 1$:
Our projection on a given vector e_1 is defined by $y^{(i)} = e_1^T x^{(i)}$. Then our projected mean and covariance

would be:

$$
\begin{aligned}
Projected\ mean\ &= \frac{\sum e_1^T x^{(i)}}{N} = e_1^T \overline{x} && (3)\\
Projected\ variance\ &= \frac{\sum \{e_1^T x^{(i)} - e_1^T \overline{x}\}^2}{n-1} = \frac{\sum \{e_1^T (x^{(i)} - \overline{x})\}^2}{n-1}\\
&= \frac{\sum \left[e_1^T (x^{(i)} - \overline{x})\right] \left[e_1^T (x^{(i)} - \overline{x})\right]^\top}{n-1}\\
&= \frac{\sum e_1^T (x^{(i)} - \overline{x})(x^{(i)} - \overline{x})^\top e_1}{n-1}\\
&= e_1^T \frac{\sum \left[(x^{(i)} - \overline{x})(x^{(i)} - \overline{x})^\top\right]}{n-1} e_1 = e_1^T S e_1 && (4)
\end{aligned}
$$

S is the covariance matrix and we need to maximize the variance. Therefore, adding a Lagrange multiplier to constraint our vectors norm, we can maximize the variance with:

$$u = e_1^T S e_1 - \lambda_1 (e_1^T e_1 - 1) \tag{5}$$

$$\frac{\partial u}{\partial e_1} = 2Se_1 - 2\lambda_1 e_1 = 0 \tag{6}$$

$$Se_1 = \lambda_1 e_1 \tag{7}$$

The vector that maximizes this expression ($e_1^T S e_1 = \lambda e_1^T e_1$) is the eigenvector of S that has the highest eigenvalue, and is called the first principal component.

Now we suppose that this condition holds for M and we try to proof the same for $M+1$.
The variance in the $M+1$ direction is given by:

$$Projected\ variance\ = e_{M+1}^T S e_{M+1}^T \tag{8}$$

To maximize it, we just need to add the same constraint as before, and an orthogonality constraint for each one of the previous M orthogonal components. Therefore, we also add M Lagrange multipliers $\eta_1..\eta_M$ for each orthogonal vector:

$$u = e_{M+1}^T S e_{M+1} - \lambda_{M+1}(e_{M+1}^T e_{M+1} - 1) + \sum_{i=1}^{M} \eta_i e_{M+1}^T e_i \tag{9}$$

$$\frac{\partial u}{\partial e_{M+1}} = 2Se_{M+1} - 2\lambda_{M+1} e_{M+1} + \sum_{i=1}^{M} \eta_i e_i = 0 \tag{10}$$

If we now left-multiply this result with one of the previous e_j^T, we get:

$$e_j^T 2Se_{M+1} - e_j^T 2\lambda_{M+1} e_{M+1} + \sum_{i=1}^{M} e_j^T \eta_i e_i = 0 \tag{11}$$

$e_j^T 2Se_{M+1} = (e_j^T 2Se_{M+1})^\top = 2e_{M+1}^T S^\top e_j = 2e_{M+1}^T S e_j = 2\lambda_j e_{M+1}^T e_j = 0$

But as our new vector e_{M+1} must be orthogonal to any of the previous one, which are already orthogonal between themselves, we get:

$$e_j^T \eta_j e_j = 0 \tag{12}$$

Applying this for every previous orthogonal vector, we realize that $\eta_j = 0$ for every $j = 1..M$. Knowing that our maximization equation is:

$$Se_{M+1} = \lambda_{M+1} e_{M+1} \tag{13}$$

And to maximize this equation e must be the eigenvector with the highest eigenvalue, different from the previous M.

Exercise 3

Calculate the LDA projection for the following dataset:
Class1 $= \{(4,1),(2,4),(2,3),(3,6),(4,4)\}$ and Class2 $= \{(9,10),(6,8),(9,5),(8,7),(10,8)\}$ and using *1-nerest neighbour classifier* on the projected data, classify the new point $(7,4)$.

Solution Exercise 3

In LDA the projection plane is defined as

$w = S_W^{-1}(\mu_1 - \mu_2)$

$w = (S_1 + S_2)^{-1}(\mu_1 - \mu_2)$

wher $S_i = \sum_{n \in C_i} (x_n - \mu_i)(x_n - \mu_i)^T$ is the scatter matrix (normalization not important when finding the projection directions)

Now $\mu_1 = \frac{\{(4,1)+(2,4)+(2,3)+(3,6)+(4,4)\}}{5} = (3, 3.6)$ and $\mu_2 = \frac{\{(9,10)+(6,8)+(9,5)+(8,7)+(10,8)\}}{5} = (8.4, 7.6)$

Also $S_1 = \begin{pmatrix} 4 & -2 \\ -2 & 13.2 \end{pmatrix}$ and $S_2 = \begin{pmatrix} 9.2 & -0.2 \\ -0.2 & 13.2 \end{pmatrix}$

Then $S_W = \begin{pmatrix} 13.2 & -2.2 \\ -2.2 & 26.4 \end{pmatrix}$ and ${S_W}^{-1} = \begin{pmatrix} 0.0768 & 0.0064 \\ 0.0064 & 0.0384 \end{pmatrix}$

Now $w = \begin{pmatrix} 0.0768 & 0.0064 \\ 0.0064 & 0.0384 \end{pmatrix} \begin{pmatrix} -5.4 \\ -4 \end{pmatrix} = \begin{pmatrix} -0.4405 \\ -0.1882 \end{pmatrix}$

After projection on the discriminating plane

Class1 $=$ $\{-1.95, -1.6338, -1.44, -2.45, -2.5147\}$ and Class2 $=$ $\{-5.84, -4.1485, -4.90, -4.8412, -5.91\}$ and the new point $(7,4)$ is projected as -3.8361. Since it is closest to the second point of Class2, it is classified as belonging to Class2.