Fisher Discriminant Analysis

— Exercises with Solutions —

Laurenz Wiskott
Institut für Neuroinformatik
Ruhr-Universität Bochum, Germany, EU

4 February 2017

Contents

1 Fisher discriminant analysis **1**
1.1 Exercises 1
1.1.1 Exercise: Combining two data distributions 1
1.1.2 Exercise: Calculate Fisher discriminant from first and second moments 2

1 Fisher discriminant analysis

1.1 Exercises

1.1.1 Exercise: Combining two data distributions

Assume two data distributions are given in two dimensions with the following mean and second-moment matrix

$$\bar{\boldsymbol{x}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{1}$$

$$\boldsymbol{C}_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2}$$

$$\bar{\boldsymbol{x}}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \tag{3}$$

$$\boldsymbol{C}_2 = \begin{pmatrix} 9 & -4 \\ -4 & 12 \end{pmatrix}. \tag{4}$$

Furthermore, the first cluster has twice as many data points as the second one.

© 2017 Laurenz Wiskott (ORCID http://orcid.org/0000-0001-6237-740X, homepage https://www.ini.rub.de/PEOPLE/wiskott/). This work (except for all figures from other sources, if present) is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License, see http://creativecommons.org/licenses/by-sa/4.0/.
These exercises complement my corresponding lecture notes, and there is a version with and one without solutions. The table of contents of the lecture notes is reproduced here to give an orientation when the exercises can be reasonably solved. For best learning effect I recommend to first seriously try to solve the exercises yourself before looking into the solutions.
More teaching material is available at https://www.ini.rub.de/PEOPLE/wiskott/Teaching/Material/.

1. Calculate the mean of the two clusters together.

   Solution: Due to the linearity of the averaging process the combined mean is

$$\begin{aligned}\bar{\boldsymbol{x}} &= \frac{2}{3}\,\bar{\boldsymbol{x}}_1 + \frac{1}{3}\,\bar{\boldsymbol{x}}_2 && (5)\\ &= \frac{2}{3}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 1 \\ 2 \end{pmatrix} && (6)\\ &= \frac{1}{3}\begin{pmatrix} 3 \\ 2 \end{pmatrix}. && (7)\end{aligned}$$

2. Calculate the second-moment matrix of the two clusters together.

   Solution: Like for the mean the combined second-moment matrix is

$$\begin{aligned}\boldsymbol{C} &= \frac{2}{3}\,\boldsymbol{C}_1 + \frac{1}{3}\,\boldsymbol{C}_2 && (8)\\ &= \frac{2}{3}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 9 & -4 \\ -4 & 12 \end{pmatrix} && (9)\\ &= \frac{1}{3}\begin{pmatrix} 13 & -4 \\ -4 & 14 \end{pmatrix}. && (10)\end{aligned}$$

3. Calculate the covariance matrix

$$\boldsymbol{\Sigma} := \langle (\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \rangle\,. \tag{11}$$

   of the two clusters together.

   Solution: As we know from exercise (RefCovarianceFromMeanAndSecondMoment)

$$\begin{aligned}\boldsymbol{\Sigma} &:= \langle (\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \rangle && (12)\\ &= \langle \boldsymbol{x}\boldsymbol{x}^T \rangle - \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T\,. && (13)\end{aligned}$$

   Given the mean and the second-moment matrix of the combined data above we get

$$\begin{aligned}\Sigma &= \langle \boldsymbol{x}\boldsymbol{x}^T \rangle - \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T && (14)\\ &\overset{(10,7)}{=} \frac{1}{3}\begin{pmatrix} 13 & -4 \\ -4 & 14 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} 3 \\ 2 \end{pmatrix}\frac{1}{3}\begin{pmatrix} 3 & 2 \end{pmatrix} && (15)\\ &= \frac{1}{9}\begin{pmatrix} 39 & -12 \\ -12 & 42 \end{pmatrix} - \frac{1}{9}\begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix} && (16)\\ &= \frac{1}{9}\begin{pmatrix} 30 & -18 \\ -18 & 38 \end{pmatrix}. && (17)\end{aligned}$$

1.1.2 Exercise: Calculate Fisher discriminant from first and second moments

Assume two data distributions are given in two dimensions with the following mean and second-moment matrix

$$\begin{aligned}\bar{\boldsymbol{x}}_1 &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, && (1)\\ \boldsymbol{C}_1 &= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, && (2)\\ \bar{\boldsymbol{x}}_2 &= \begin{pmatrix} 1 \\ 2 \end{pmatrix}, && (3)\\ \boldsymbol{C}_2 &= \begin{pmatrix} 9 & -4 \\ -4 & 12 \end{pmatrix}. && (4)\end{aligned}$$

Furthermore, the first cluster has twice as many data points as the second one.

Calculate the first Fisher discriminant for the two clusters and illustrate the data and your result graphically.

Solution: Conceptually the first step is to remove the individual cluster means from the data points of the two clusters and then sphere the combined data. For the sphering we need the second-moment matrix of the cluster-mean corrected data. For each cluster individually this second-moment matrix is the same as the covariance matrix of the uncorrected data. Thus, we first calculate the covariance matrix

$$\boldsymbol{\Sigma} := \langle (\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \rangle \,. \tag{5}$$

of the two clusters individually.

As we know from exercise (RefCovarianceFromMeanAndSecondMoment)

$$\begin{aligned} \boldsymbol{\Sigma} &:= \langle (\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \rangle & (6)\\ &= \langle \boldsymbol{x}\boldsymbol{x}^T \rangle - \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^T \,. & (7) \end{aligned}$$

Given the means and the second-moment matrices above we get

$$\begin{aligned} \Sigma_1 &= \langle \boldsymbol{x}_1 \boldsymbol{x}_1^T \rangle - \bar{\boldsymbol{x}}_1 \bar{\boldsymbol{x}}_1^T & (8)\\ &\overset{(2,1)}{=} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} & (9)\\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & (10)\\ \Sigma_2 &= \langle \boldsymbol{x}_2 \boldsymbol{x}_2^T \rangle - \bar{\boldsymbol{x}}_2 \bar{\boldsymbol{x}}_2^T & (11)\\ &\overset{(4,3)}{=} \begin{pmatrix} 9 & -4 \\ -4 & 12 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix} & (12)\\ &= \begin{pmatrix} 8 & -6 \\ -6 & 8 \end{pmatrix}. & (13) \end{aligned}$$

Since calculating the covariance matrices implies removal of the mean, we can directly combine the individual covariance matrices to get the second-moment matrix of the cluster-mean corrected data.

$$\begin{aligned} \tilde{\boldsymbol{C}} &:= \frac{2}{3}\Sigma_1 + \frac{1}{3}\Sigma_1 & (14)\\ &= \frac{2}{3}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 8 & -6 \\ -6 & 8 \end{pmatrix} & (15)\\ &= \frac{1}{3}\begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}. & (16) \end{aligned}$$

In order to sphere the cluster-mean corrected data, we have to find the eigenvalues and normalized egenvectors of this matrix. Due to the symmetry of the second-moment matrix it is easy to guess that $(1,1)^T/\sqrt{2}$ is an eigenvector, and then $(1,-1)^T/\sqrt{2}$ is obviously a possible other one, orthogonal to the first. With a little bit of extra thought one can also guess that the second one has a larger eigenvalue, since, as one can see from $\tilde{\boldsymbol{C}}$, the two components of the centered data are negatively correlated. Thus, we assume that the orthogonal eigenvector matrix is

$$\tilde{\boldsymbol{U}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \tag{17}$$

as we can verify by multiplying the two matrices,

$$\begin{aligned}
\tilde{\boldsymbol{C}}\tilde{\boldsymbol{U}} &= \frac{1}{3}\begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} && (18)\\
&= \frac{1}{3\sqrt{2}}\begin{pmatrix} 16 & 4 \\ -16 & 4 \end{pmatrix} && (19)\\
&= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\underbrace{\frac{1}{3}\begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}}_{=:\,\tilde{\boldsymbol{\Lambda}}} && (20)\\
&= \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Lambda}}\,, && (21)
\end{aligned}$$

which also shows that the corresponding eigenvalues are

$$\lambda_1 \,=\, 16/3\,, \quad \lambda_2 \,=\, 4/3\,. \tag{22}$$

The sphering matrix then is

$$\begin{aligned}
\tilde{\boldsymbol{W}} &:= \tilde{\boldsymbol{U}}\,\mathrm{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}\right)\tilde{\boldsymbol{U}}^T && (23)\\
&\overset{(17,22)}{=} \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{16/3}} & 0 \\ 0 & \frac{1}{\sqrt{4/3}} \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}^T && (24)\\
&= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\frac{\sqrt{3}}{4}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} && (25)\\
&= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\frac{\sqrt{3}}{4\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 2 & 2 \end{pmatrix} && (26)\\
&= \frac{\sqrt{3}}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}. && (27)
\end{aligned}$$

The sphered cluster-mean corrected data becomes

$$\hat{\boldsymbol{x}} \;:=\; \tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}} \tag{28}$$

and we verify that it has a unit second-moment matrix indeed,

$$\begin{aligned}
\hat{\boldsymbol{C}} &= \langle\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T\rangle && (29)\\
&\overset{(28)}{=} \langle\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{W}}^T\rangle && (30)\\
&= \tilde{\boldsymbol{W}}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{W}}^T && (31)\\
&\overset{(27,16)}{=} \frac{\sqrt{3}}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\frac{1}{3}\begin{pmatrix} 10 & -6 \\ -6 & 10 \end{pmatrix}\frac{\sqrt{3}}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}^T && (32)\\
&= \frac{\sqrt{3}}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\frac{\sqrt{3}}{3\cdot 8}\begin{pmatrix} 24 & -8 \\ -8 & 24 \end{pmatrix} && (33)\\
&= \frac{1}{8\cdot 8}\begin{pmatrix} 64 & 0 \\ 0 & 64 \end{pmatrix} && (34)\\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. && (35)
\end{aligned}$$

In this sphered coordinate system we now have to calculate the first eigenvector of the covariance matrix of the cluster means. Since there are only two clusters this is simply the normalized difference vector between the two means, i.e.

$$\hat{\boldsymbol{d}} = \frac{\tilde{\boldsymbol{W}}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}{\|\tilde{\boldsymbol{W}}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)\|} \tag{36}$$

$$\overset{(27,1,3)}{=} \frac{\frac{\sqrt{3}}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)}{\text{numerator}} \tag{37}$$

$$= \frac{\frac{\sqrt{3}}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 0 \\ -2 \end{pmatrix}}{\text{numerator}} \tag{38}$$

$$= \frac{1}{\sqrt{10}}\begin{pmatrix} -1 \\ -3 \end{pmatrix}. \tag{39}$$

The decision whether a data point $\boldsymbol{x}$ belongs to one class or the other can now be based on whether the inner product $\hat{\boldsymbol{d}}^T\hat{\boldsymbol{x}}$ is greater or less than some suitable threshold θ. Translating this back into the original coordinate system yields

$$\theta < \hat{\boldsymbol{d}}^T\hat{\boldsymbol{x}} \tag{40}$$

$$\overset{(28)}{=} \underbrace{\hat{\boldsymbol{d}}^T\tilde{\boldsymbol{W}}}_{=:\boldsymbol{d}^T}\boldsymbol{x} \tag{41}$$

$$\overset{(28)}{=} \boldsymbol{d}^T\boldsymbol{x} \tag{42}$$

$$\text{with} \quad \boldsymbol{d} := \tilde{\boldsymbol{W}}^T\hat{\boldsymbol{d}} \tag{43}$$

$$= \frac{\sqrt{3}}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}^T\frac{1}{\sqrt{10}}\begin{pmatrix} -1 \\ -3 \end{pmatrix} \tag{44}$$

$$= \frac{\sqrt{3}}{8\cdot\sqrt{10}}\begin{pmatrix} -6 \\ -10 \end{pmatrix}. \tag{45}$$

Since the normalization of $\boldsymbol{d}$ is pretty arbitrary and can always be incorporated in the threshold, one can choose more conveniently

$$\boldsymbol{d} := \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \tag{46}$$

which points in the same direction.