Data Mining
Corrections of the exercises

François Bachoc
Université Paul Sabatier

The lecture notes can be found here: `https://www.math.univ-toulouse.fr/~xgendre/ens/m2se/DataMining.pdf`.

Exercise 1

1) *Consider 10 cars that are for sale with prices (in k euros)* 10, 6, 7, 6, 22, 43, 33, 7, 8, 11. *Consider the case of uniform weights. Compute the mean price. Compute the median price.*

Correction: The mean price is

$$\frac{10+6+7+6+22+43+33+7+8+11}{10} = \frac{153}{10} \approx 15.3.$$

The sorted data are

$$6, 6, 7, 7, 8, 10, 11, 22, 33, 43.$$

The proportion of the data smaller or equal to 8 is $5/10 = 0.5$. For any $t < 8$, the proportion of the data smaller or equal to t is less than $4/10$. Thus the median is 8.

2) *Consider the following regions with population (in millions) and unemployment rates (in percent) given by the pairs* $(4,8),(4,4),(6,10),(3,9),(2,6),(7,21),(6,11),(4,7),(5,2),(8,8)$ *of the form* $(\text{population}, \text{unemployment})$. *Each region is weighted by its population. Compute the normalized weights* $w_1, \ldots, w_n$ *for these data. With these normalized weights, compute the mean and the standard deviation of the unemployment rate. You can use the formula* $\sigma^2(x) = \overline{x^2} - \bar{x}^2$. *Compute the quantile* $q_{0.25}(x)$ *for these data* x. *You can use a calculator.*

Correction: The total of the population (the sum of the unnormalized weights) is

$$4+4+6+3+2+7+6+4+5+8 = 49.$$

Hence the normalized weights are

$$w_1 = \frac{4}{49}, w_2 = \frac{4}{49}, w_3 = \frac{6}{49}, w_4 = \frac{3}{49}, w_5 = \frac{2}{49}, w_6 = \frac{7}{49}, w_7 = \frac{6}{49}, w_8 = \frac{4}{49}, w_9 = \frac{5}{49}, w_{10} = \frac{8}{49}.$$

The mean is

$$\frac{4 \times 8 + 4 \times 4 + 6 \times 10 + 3 \times 9 + 2 \times 6 + 7 \times 21 + 6 \times 11 + 4 \times 7 + 5 \times 2 + 8 \times 8}{49} = \frac{462}{49} \approx 9.43.$$

The variance is approximately

$$\frac{4}{49}8^2 + \frac{4}{49}4^2 + \frac{6}{49}10^2 + \frac{3}{49}9^2 + \frac{2}{49}6^2 + \frac{7}{49}21^2 + \frac{6}{49}11^2 + \frac{4}{49}7^2 + \frac{5}{49}2^2 + \frac{8}{49}8^2 - 9.43^2 \approx 28.97.$$

The standard deviation is approximately

$$\sqrt{28.97} \approx 5.38.$$

Let us write the ordered data $x_{(1)} \leq \cdots \leq x_{(10)}$ (ordering $x_1, \ldots, x_{10}$) and the corresponding (normalized) weights $w_{x_{(1)}}, \ldots, w_{x_{(10)}}$ (keeping the correspondence between an individual x_i and its weight). We obtain

$$(w_{x_{(1)}}, x_{(1)}) = (\frac{5}{49}, 2), (w_{x_{(2)}}, x_{(2)}) = (\frac{4}{49}, 4), (w_{x_{(3)}}, x_{(3)}) = (\frac{2}{49}, 6), (w_{x_{(4)}}, x_{(4)}) = (\frac{4}{49}, 7), (w_{x_{(5)}}, x_{(5)}) = (\frac{8}{49}, 8),$$
$$(w_{x_{(6)}}, x_{(6)}) = (\frac{4}{49}, 8), (w_{x_{(7)}}, x_{(7)}) = (\frac{3}{49}, 9), (w_{x_{(8)}}, x_{(8)}) = (\frac{6}{49}, 10), (w_{x_{(9)}}, x_{(9)}) = (\frac{6}{49}, 11), (w_{x_{(10)}}, x_{(10)}) = (\frac{7}{49}, 21).$$

Let us compute some cumulated weights

$$w_{x_{(1)}} + w_{x_{(2)}} + w_{x_{(3)}} = \frac{5+4+2}{49} \approx 0.224,$$

$$w_{x_{(1)}} + w_{x_{(2)}} + w_{x_{(3)}} + w_{x_{(4)}} = \frac{5+4+2+4}{49} \approx 0.306.$$

Hence the quantile $q_{0.25}(x)$ is 7, because more than a fraction 0.25 of the (weighted) data are smaller or equal to $x_{(4)} = 7$ but for any $t < 7$, strictly less than a fraction 0.25 of the data are smaller or equal to t.

3) *Consider data $x_1, \ldots, x_n$ with $n = 100$ and $x_i = i$ for $i = 1, \ldots, 100$. Consider uniform weights. Compute the mean and median of x. Consider that a data point $x_{n+1} = 10^6$ is added. Compute the new mean and median (still with uniform weights). Interpret the results.*

Correction: The first mean is

$$\frac{\sum_{i=1}^{100} i}{100} = \frac{\frac{100 \times 101}{2}}{100} = 50.5.$$

The first median is 50, because 50/100 of the data are smaller or equal to 50 but for $t < 50$, at most $49/100 < 0.5$ of the data are smaller or equal to t.

The second mean is

$$\frac{10^6 + \sum_{i=1}^{100} i}{101} = \frac{10^6}{101} + \frac{\frac{100 \times 101}{2}}{101} \approx 9951.$$

The second median is 51 because $51/101 \geq 0.5$ of the data are smaller or equal to 51 but for $t < 51$ at most $50/101 < 0.5$ of the data are smaller or equal to t.

The interpretation is that the mean is much more sensitive than the median to outlier/extreme individual data points.

Exercise 2

Prove that the histogram function $h_{x,\lambda} : \mathbb{R} \to \mathbb{R}$ on page 4 of the lecture notes has integral 1.

Correction:
We have

$$\begin{aligned}
\int_{-\infty}^{\infty} h_{x,\lambda}(t) dt &= \sum_{i=1}^{n} w_i \frac{1}{\lambda} \int_{-\infty}^{\infty} K\left(\frac{t - x_i}{\lambda}\right) dt \\
\text{(change of variable } u = (t - x_i)/\lambda :) &= \sum_{i=1}^{n} w_i \int_{-\infty}^{\infty} K(u) du \\
&= \sum_{i=1}^{n} w_i \\
&= 1.
\end{aligned}$$

Exercise 3

Prove that $|\rho(x, y)| = 1$ if and only if $(x_1, y_1), \ldots, (x_n, y_n)$ are all distributed on a straight line (as written on page 6 of the lecture notes).

Correction: Let $\tilde{x}_w = (w_1^{1/2}(x_1 - \bar{x}), \dots, w_n^{1/2}(x_n - \bar{x}))$ and $\tilde{y}_w = (w_1^{1/2}(y_1 - \bar{y}), \dots, w_n^{1/2}(y_n - \bar{y}))$. Recall the notation $\langle \cdot, \cdot \rangle$ for the standard inner product on $\mathbb{R}^n$. Then, we have

$$\sigma(x, y) = \sum_{i=1}^n w_i(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n w_i^{1/2}(x_i - \bar{x}) w_i^{1/2}(y_i - \bar{y}) = \sum_{i=1}^n (\tilde{x}_w)_i (\tilde{y}_w)_i = \langle \tilde{x}_w, \tilde{y}_w \rangle.$$

Similarly $\sigma^2(x) = \langle \tilde{x}_w, \tilde{x}_w \rangle$ and $\sigma^2(y) = \langle \tilde{y}_w, \tilde{y}_w \rangle$. Hence, $|\rho(x, y)| = 1$ is equivalent to

$$|\langle \tilde{x}_w, \tilde{y}_w \rangle| = \sqrt{\langle \tilde{x}_w, \tilde{x}_w \rangle \times \langle \tilde{y}_w, \tilde{y}_w \rangle}$$

which is the equality case in Cauchy-Shwarz inequality. Hence $|\rho(x, y)| = 1$ if and only if there exists $a \in \mathbb{R}$ such that for $i = 1, \dots, n$, $w_i^{1/2}(y_i - \bar{y}) = a w_i^{1/2}(x_i - \bar{x})$, that is, for $i = 1, \dots, n$, $(y_i - \bar{y}) = a(x_i - \bar{x})$. Hence $(x_1 - \bar{x}, y_1 - \bar{y}), \dots, (x_n - \bar{x}, y_n - \bar{y})$ are all distributed on a straight line. Hence (by translation in $\mathbb{R}^2$ with shift $(\bar{x}, \bar{y})$), also $(x_1, y_1), \dots, (x_n, y_n)$ are all distributed on a straight line.

Exercise 4

Consider 7 elementary school pupils with ages and weights given by the pairs

$$(4, 25), (5, 28), (6, 31), (7, 33), (8, 32), (9, 39), (10, 43)$$

(of the form (age, weight)). Compute the covariance between age and weight (uniform weights). You can use the formula $\sigma(x, y) = \overline{xy} - \bar{x}\bar{y}$ Interpret the result.

Correction: The mean of the age is

$$\frac{4 + 5 + 6 + 7 + 8 + 9 + 10}{7} = \frac{49}{7} = 7.$$

The mean of the weight is

$$\frac{25 + 28 + 31 + 33 + 32 + 39 + 43}{7} = \frac{231}{7} = 33.$$

The covariance is then

$$\begin{aligned} &\frac{1}{7}\Big(4 \times 25 + 5 \times 28 + 6 \times 31 + 7 \times 33 + 8 \times 32 + \\ &9 \times 39 + 10 \times 43\Big) - 7 \times 33 \\ &= 11. \end{aligned}$$

The interpretation is that the covariance is positive, which is in agreement with the fact, that, in general, the weight increases when the age increases.

Exercise 5

Prove the expression of $\hat{a}$ and $\hat{b}$ on page 6 of the lecture notes. You can use the formulas

$$\sum_{i=1}^n w_i x_i y_i = \sigma(x, y) + \bar{x}\bar{y}, \quad \sum_{i=1}^n w_i x_i^2 = \sigma(x)^2 + \bar{x}^2, \quad \sum_{i=1}^n w_i y_i^2 = \sigma(y)^2 + \bar{y}^2.$$

Correction: The derivatives of the least square criterion are

$$\begin{aligned} \frac{\partial}{\partial a} \sum_{i=1}^n w_i(y_i - a x_i - b)^2 &= -2 \sum_{i=1}^n w_i(y_i - a x_i - b) x_i \\ &= -2 \left(\sigma(x, y) + \bar{x}\bar{y} - a(\sigma(x)^2 + \bar{x}^2) - b\bar{x}\right) \end{aligned} \tag{1}$$

and

$$\begin{aligned} \frac{\partial}{\partial b} \sum_{i=1}^n w_i(y_i - a x_i - b)^2 &= -2 \sum_{i=1}^n w_i(y_i - a x_i - b) \\ &= -2 \left(\bar{y} - a\bar{x} - b\right). \end{aligned} \tag{2}$$

The optimal coefficients $\hat{a}$ and $\hat{b}$ make these two partial derivatives equal to zero, which yields, from (2),

$$\hat{b} = \bar{y} - \hat{a}\bar{x}.$$

Then from (1), we obtain

$$0 = \sigma(x, y) + \bar{x}\bar{y} - \hat{a}\sigma(x)^2 - \hat{a}\bar{x}^2 - (\bar{y} - \hat{a}\bar{x})\bar{x}$$

which gives

$$0 = \sigma(x, y) - \hat{a}\sigma(x)^2$$

and thus

$$\hat{a} = \frac{\sigma(x, y)}{\sigma(x)^2}.$$

Exercise 6

We consider 6 companies and their numbers of employees (in thousands), annual growth (percent) and age (years). The data for these companies, of the form (employees,growth,age) are

$$(4, 5, 22), (6, 7, 11), (6, 8, 2), (3, 8, 54), (8, 2, 34), (4, 5, 5).$$

Construct the data matrix X and the centered data matrix $\bar{X}$ associated to these data (uniform weights).

Correction: We have

$$X = \begin{pmatrix} 4 & 5 & 22 \\ 6 & 7 & 11 \\ 6 & 8 & 2 \\ 3 & 8 & 54 \\ 8 & 2 & 34 \\ 4 & 5 & 5 \end{pmatrix}.$$

The center of gravity is

$$g(x) = \frac{1}{6} \begin{pmatrix} 4+6+6+3+8+4 \\ 5+7+8+8+2+5 \\ 22+11+2+54+34+5 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 31 \\ 35 \\ 128 \end{pmatrix} \approx \begin{pmatrix} 5.17 \\ 5.83 \\ 21.33 \end{pmatrix}.$$

The centered data matrix is then

$$\bar{X} = \begin{pmatrix} -1.17 & -0.83 & 0.66 \\ 0.83 & 1.17 & -10.33 \\ 0.83 & 2.17 & -19.33 \\ -2.17 & 2.17 & 32.66 \\ 2.83 & -3.83 & 12.66 \\ -1.17 & -0.83 & -16.33 \end{pmatrix}.$$

Exercise 7

Show that ${}^t uMu = {}^t u[(M + {}^t M)/2]u$ as stated on page 8 of the lecture notes.

Correction: We have

$${}^t u[(M + {}^t M)/2]u = \frac{1}{2}{}^t uMu + \frac{1}{2}{}^t u\,{}^t Mu.$$

Since ${}^t u\,{}^t Mu$ is a scalar, it is equal to its transpose which is ${}^t u^t({}^t M)u = {}^t uMu$. This concludes the proof.

Exercise 8

Consider the two-dimensional linear subspace E of $\mathbb{R}^3$ spanned by the two vectors $(1, 1, 1)$ and $(1, 0, 0)$. Consider the diagonal matrix M with diagonal elements $(1, 1/2, 1/2)$ to define the inner product $\langle\cdot,\cdot\rangle_M$.

1) *Show that $(1, 0, 0)$ and $(0, 1, 1)$ constitute an M-orthonormal basis of E.*

Correction:

The vector $(1,0,0)$ belongs to E. The vector $(0,1,1)=(1,1,1)-(1,0,0)$ also belongs to E. We have

$$\langle (1,0,0),(0,1,1)\rangle_M = 1.1.0 + 0.(1/2).1 + 0.(1/2).1 = 0,$$

$$\langle (1,0,0),(1,0,0)\rangle_M = 1.1.1 + 0.(1/2).0 + 0.(1/2).0 = 1$$

and

$$\langle (0,1,1),(0,1,1)\rangle_M = 0.1.0 + 1.(1/2).1 + 1.(1/2).1 = 1$$

hence $(1,0,0)$ and $(0,1,1)$ are M-orthonormal and thus are an M-orthonormal basis of E which is of dimension 2.

2) *Compute the M-orthogonal projection of $(1,1,-1)$ on E.*

Correction:

The M-orthogonal projection is

$$\begin{aligned}
&\langle (1,1,-1),(1,0,0)\rangle_M (1,0,0) + \langle (1,1,-1),(0,1,1)\rangle_M (0,1,1)\\
&= \Big(1.1.1 + 1.(1/2).0 + (-1).(1/2).0\Big)(1,0,0) + \Big(1.1.0 + 1.(1/2).1 + (-1).(1/2).1\Big)(0,1,1)\\
&= 1.(1,0,0) + 0(0,1,1)\\
&= (1,0,0).
\end{aligned}$$

Exercise 9

The goal is to prove that taking the d first eigenvectors of ΣM (that are M-orthonormal) maximizes the inertia (page 10 of the lecture notes). We let the p eigenvalues of ΣM be $\lambda_1 > \cdots > \lambda_p > 0$. For any M-orthonormal vectors $z^1,\dots,z^d$ in $\mathbb{R}^p$ the intertia is $\sum_{j=1}^d \langle \Sigma M z^j, z^j\rangle_M$ (page 10 of the lecture notes). Assume $d < p$.

1) *Consider M-orthonormal vectors $z^1,\dots,z^d$ in $\mathbb{R}^p$ and write for $j=1,\dots,d$ $z^j = \sum_{i=1}^p A_{i,j} v^i$, where $v^1,\dots,v^p$ are p M-orthonormal eigenvectors of ΣM associated to the eigenvalues $\lambda_1,\dots,\lambda_p$ and A is a $p\times d$ matrix. Show that ${}^tAA = I_d$.*

Correction:
We have, for $k,l = 1,\dots,d$,

$$({}^tAA)_{k,l} = \sum_{i=1}^p A_{i,k}A_{i,l} = \sum_{i=1}^p\sum_{j=1}^p A_{i,k}A_{j,l}\langle v^i, v^j\rangle_M = \langle \sum_{i=1}^p A_{i,k}v^i, \sum_{j=1}^p A_{j,l}v^j\rangle_M = \langle z^k, z^l\rangle_M = 1_{k=l},$$

where we recall the notation $1_{k=l}$ which is 1 if $k=l$ and 0 otherwise.

2) *For the same vectors $z^1,\dots,z^d$ as before, show that*

$$\sum_{j=1}^d \langle \Sigma M z^j, z^j\rangle_M = \sum_{j=1}^d\sum_{a=1}^p \lambda_a A_{a,j}^2.$$

Correction:

We have

$$\begin{aligned}
\sum_{j=1}^d \langle \Sigma M z^j, z^j\rangle_M &= \sum_{j=1}^d \langle \sum_{a=1}^p A_{a,j}\Sigma M v^a, \sum_{b=1}^p A_{b,j} v^b\rangle_M = \sum_{j=1}^d\sum_{a=1}^p\sum_{b=1}^p A_{a,j}A_{b,j}\langle \Sigma M v^a, v^b\rangle_M\\
&= \sum_{j=1}^d\sum_{a=1}^p\sum_{b=1}^p A_{a,j}A_{b,j}\lambda_a\langle v^a, v^b\rangle_M = \sum_{j=1}^d\sum_{a=1}^p\sum_{b=1}^p A_{a,j}A_{b,j}\lambda_a 1_{a=b} = \sum_{j=1}^d\sum_{a=1}^p \lambda_a A_{a,j}^2.
\end{aligned}$$

3) *For $a=1,\dots,p$, let $\beta_a = \sum_{j=1}^d A_{a,j}^2$. Show that $\beta_a \leq 1$. Hint: you can add columns to A to obtain a $p\times p$ orthonormal matrix $\tilde{A}$ such that ${}^t\tilde{A}\tilde{A} = I_p$ (this is always possible since the d columns of A are I_p-orthonormal).*

Correction: With $\tilde{A}$ as in the hint, β_a is smaller than the square Euclidean norm of the row a of $\tilde{A}$. We have $\tilde{A}{}^t\tilde{A} = I_p$ since $\tilde{A}$ is square and ${}^t\tilde{A}\tilde{A} = I_p$. Hence, the rows of $\tilde{A}$ are orthonormal hence they have square

Euclidean norms 1.

4) *Show that* $\sum_{a=1}^{p} \beta_a = d$.

Correction: The quantity $\sum_{a=1}^{p} \beta_a = \sum_{a=1}^{p} \sum_{j=1}^{d} A_{a,j}^2$ is the sum of the square norms of the d columns of A which are orthonormal since ${}^tAA = I_d$. Hence these d square norms are 1 and thus their sum is d.

5) *Show that the maximum of* $\sum_{a=1}^{p} \lambda_a \beta_a$ *under the constraints* $0 \leq \beta_a \leq 1$ *for* $a = 1, \dots, p$ *and* $\sum_{a=1}^{p} \beta_a = d$ *is* $\sum_{a=1}^{d} \lambda_a$.

Correction: There exists a maximizer by continuity and compacity. Let $\hat{\beta}_1, \dots, \hat{\beta}_p$ maximize $\sum_{a=1}^{p} \lambda_a \beta_a$ under the constraints above and assume $\hat{\beta}_q > 0$ for some $q \in \{d+1, \dots, p\}$. Then there is some $i \in \{1, \dots, d\}$ such that $\hat{\beta}_i < 1$ since $\sum_{a=1}^{p} \hat{\beta}_a = d$. Let $\epsilon > 0$ be such that $\hat{\beta}_q \geq \epsilon$ and $\hat{\beta}_i \leq 1 - \epsilon$. Then, if we decrease $\hat{\beta}_q$ by ϵ and increase $\hat{\beta}_i$ by ϵ, the constraints on $\hat{\beta}_1, \dots, \hat{\beta}_p$ are still satisfied and the criterion is increased by $\epsilon(\lambda_i - \lambda_q) > 0$. This shows that the only possible maximizer is $\hat{\beta}_1 = 1, \dots, \hat{\beta}_d = 1, \hat{\beta}_{d+1} = 0, \dots, \hat{\beta}_p = 0$. This yields $\sum_{a=1}^{p} \lambda_a \hat{\beta}_a = \sum_{a=1}^{d} \lambda_a$.

6) *Conclude by showing that the d first eigenvectors of* ΣM *maximize the inertia.*

Correction: From the previous questions, for any M-orthonormal vectors $z^1, \dots, z^d$ in $\mathbb{R}^d$, the intertia is $\sum_{a=1}^{p} \lambda_a \beta_a$ and is smaller than $\sum_{a=1}^{d} \lambda_a$. With the d first eigenvectors $v^1, \dots, v^d$ of M, the intertia becomes

$$\sum_{j=1}^{d} \langle \Sigma M v^j, v^j \rangle_M = \sum_{j=1}^{d} \langle \lambda_j v^j, v^j \rangle_M = \sum_{j=1}^{d} \lambda_j,$$

since $v^1, \dots, v^d$ are M-orthonormal. Hence, $v^1, \dots, v^d$ indeed maximize the intertia.

Exercise 10

Prove the equation ${}^tV(M\Sigma M)V = \Lambda$ *on page 16 of the lecture notes.*

Correction:
We have

$${}^tV(M\Sigma M)V = \begin{pmatrix} {}^tv^1 \\ \vdots \\ {}^tv^p \end{pmatrix} M(\Sigma M) \begin{pmatrix} v^1 & \dots & v^p \end{pmatrix} = \begin{pmatrix} {}^tv^1 M \\ \vdots \\ {}^tv^p M \end{pmatrix} \begin{pmatrix} (\Sigma M)v^1 & \dots & (\Sigma M)v^p \end{pmatrix} = \begin{pmatrix} {}^tv^1 M \\ \vdots \\ {}^tv^p M \end{pmatrix} \begin{pmatrix} \lambda_1 v^1 & \dots & \lambda_p v^p \end{pmatrix}.$$

The right-most matrix above is $p \times p$ and, for $i, j = 1, \dots, p$, its element i, j is ${}^tv^i M \lambda_j v^j = \lambda_j \langle v^i, v^j \rangle_M = \lambda_j 1_{i=j}$ which concludes the proof.

Exercise 11

Show that if a d-dimensional subspace E_d *maximizes the intertia of the projected observations* $I_M(x, E_d)$ *(page 10 of the lecture notes) then it minimizes the intertia of the projection errors* $I_M(x - \pi_{E_d}(x)) = \sum_{i=1}^{n} w_i ||\tilde{x}_i - \pi_{E_d}(\tilde{x}_i)||_M^2$.

Correction:
For $i = 1, \dots, n$, by definition of the M-orthogonal projection, we have $||\tilde{x}_i||_M^2 = ||\pi_{E_d}(\tilde{x}_i)||_M^2 + ||\tilde{x}_i - \pi_{E_d}(\tilde{x}_i)||_M^2$. Hence we obtain

$$\sum_{i=1}^{n} w_i ||\tilde{x}_i||_M^2 = \sum_{i=1}^{n} w_i ||\pi_{E_d}(\tilde{x}_i)||_M^2 + \sum_{i=1}^{n} w_i ||\tilde{x}_i - \pi_{E_d}(\tilde{x}_i)||_M^2 = I_M(x, E_d) + I_M(x - \pi_{E_d}(x)).$$

The quantity on the left-hand side above does not depend on E_d. Hence, if E_d maximizes $I_M(x, E_d)$, it maximizes $\sum_{i=1}^{n} w_i ||\tilde{x}_i||_M^2 - I_M(x - \pi_{E_d}(x))$ and thus is minimizes $I_M(x - \pi_{E_d}(x))$.

Exercise 12

1) *In the context of page 20 of the lecture notes, show that for $i, j \in \{1, \dots, p\}$,*

$$\langle \tilde{x}^j, u^i \rangle_W = \sqrt{\lambda_i} v_j^i.$$

Correction:
We have

$$\langle \tilde{x}^j, u^i \rangle_W = {}^t\tilde{x}^j W u^i = \frac{1}{\sqrt{\lambda_i}} {}^t\tilde{x}^j W c^i = \frac{1}{\sqrt{\lambda_i}} [{}^t\bar{X} W C]_{j,i} = \frac{1}{\sqrt{\lambda_i}} [{}^t\bar{X} W \bar{X} M V]_{j,i} = \frac{1}{\sqrt{\lambda_i}} [\Sigma M V]_{j,i}$$
$$= \frac{1}{\sqrt{\lambda_i}} [\Sigma M v^i]_j = \sqrt{\lambda_i} (v^i)_j = \sqrt{\lambda_i} v_j^i.$$

2) *In the context of page 20 of the lecture notes, show that, for $j \in \{1, \dots, p\}$,*

$$\sum_{k=1}^{p} \rho(x^j, c^k)^2 = 1.$$

Correction: We have

$$\sum_{k=1}^{p} \rho(x^j, c^k)^2 = \sum_{k=1}^{p} \frac{\lambda_k}{\sigma^2(x^j)} (v_j^k)^2 = \sum_{k=1}^{p} \frac{\langle \tilde{x}^j, u^k \rangle_W^2}{\sigma^2(x^j)} = \frac{||\tilde{x}^j||_W^2}{\sigma^2(x^j)} = 1.$$

Above, in the first equality, we have used a formula on page 20 of the lecture notes, in the second equality we have used **1)** and in the third inequality, we have used that $u^1, \dots, u^p$ is a W-orthonormal basis of $\mathbb{R}^n$.

Exercise 13

1) *Show that for $j = 1, \dots, p$, the new vector of observations c^j has mean zero (context of page 15 of the lecture notes).*

Correction: We have $c^j = \bar{X} M v^j$. The mean of c^j is $\sum_{i=1}^{n} w_i c_i^j = {}^t1_n W c^j$ with 1_n the $n \times 1$ vector composed of ones. Then the mean of c^j is

$${}^t1_n W \bar{X} M v^j = ({}^t1_n W \bar{X}) M v^j.$$

For $k = 1, \dots, p$, the column k of ${}^t1_n W \bar{X}$ is ${}^t1_n W \tilde{x}^k$ which is the mean of $\tilde{x}^k$ which is zero. Hence ${}^t1_n W \bar{X} M v^j = 0$ which concludes the proof.

2) *For $d = 1, \dots, p$, let E_d be the two-dimensional subset of $\mathbb{R}^n$ spanned by the d first W-orthonormal eigenvectors $u^1, \dots, u^d$ of $\bar{X} M {}^t\bar{X} W$ (context of page 20 of the lecture notes). Show that, for $j = 1, \dots, p$,*

$$\frac{||\pi_{E_d}(x^j)||_W^2}{||x^j||_W^2} = \sum_{k=1}^{d} \rho(x^j, c^k)^2.$$

Correction:
Using that $u^1, \dots, u^p$ is a W-orthonormal basis, we have

$$\frac{||\pi_{E_d}(x^j)||_W^2}{||x^j||_W^2} = \frac{\sum_{k=1}^{d} \langle x^j, u^k \rangle_W^2}{\sum_{k=1}^{p} \langle x^j, u^k \rangle_W^2} = \frac{\sigma(x^j)^2 \sum_{k=1}^{d} \rho(x^j, c^k)^2}{\sigma(x^j)^2 \sum_{k=1}^{p} \rho(x^j, c^k)^2} = \frac{\sum_{k=1}^{d} \rho(x^j, c^k)^2}{\sum_{k=1}^{p} \rho(x^j, c^k)^2} = \sum_{k=1}^{d} \rho(x^j, c^k)^2.$$

Above, in the second equality we have used a formula on page 20 of the lecture notes. For the last equality, we have used Exercize 12.

Exercise 14

The goal is to carry out the computations of PCA on simple simulated data. Note that you are not expected to interpret the results, since the data are simulated arbitrarily and do not come from a real data set. Consider the data matrix

$$X = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 \\ 0 & -4 & -4 & 0 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{pmatrix}.$$

1) *Consider uniform weights and the matrix $M = I_4$ to compute the distances on the space of individuals. Compute the covariance matrix.*

Correction: We remark that the matrix X is centered, $X = \bar{X}$.
The covariance matrix is

$$\Sigma = \frac{1}{5}{}^t\bar{X}\bar{X} = \frac{1}{5}\begin{pmatrix} 4 & 0 & 0 & 4 \\ 0 & 22 & 22 & 0 \\ 0 & 22 & 22 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}.$$

2) *Show that two eigenvalues of Σ are 0 and that the two first eigenvectors are*

$$v^1 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} \quad \text{and} \quad v^2 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{pmatrix}$$

with eigenvalues $44/5$ *and* $8/5$.

Correction:
The lines 1 and 4 of Σ are identical. The lines 2 and 3 of Σ are identical. The lines 1 and 2 of Σ are linearly independent. Hence Σ has rank 2 and thus its two smallest eigenvalues are 0. We have

$$\Sigma\begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 4 & 0 & 0 & 4 \\ 0 & 22 & 22 & 0 \\ 0 & 22 & 22 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}\begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 0 \\ \frac{22}{\sqrt{2}} + \frac{22}{\sqrt{2}} \\ \frac{22}{\sqrt{2}} + \frac{22}{\sqrt{2}} \\ 0 \end{pmatrix} = \frac{44}{5}\begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}.$$

Similarly

$$\Sigma\begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 4 & 0 & 0 & 4 \\ 0 & 22 & 22 & 0 \\ 0 & 22 & 22 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{pmatrix} = \frac{1}{5}\begin{pmatrix} \frac{4}{\sqrt{2}} + \frac{4}{\sqrt{2}} \\ 0 \\ 0 \\ \frac{4}{\sqrt{2}} + \frac{4}{\sqrt{2}} \end{pmatrix} = \frac{8}{5}\begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{pmatrix}.$$

Finally, v^1 and v^2 indeed have Euclidean norm 1.

3) *Compute $c^1, c^2, \rho(x^1, c^1), \rho(x^1, c^2), \dots, \rho(x^4, c^1), \rho(x^4, c^2)$.*

Correction:
We have

$$\begin{pmatrix} c^1 & c^2 \end{pmatrix} = \bar{X}\begin{pmatrix} v^1 & v^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 \\ 0 & -4 & -4 & 0 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 & \sqrt{2} \\ 2\sqrt{2} & \sqrt{2} \\ -4\sqrt{2} & 0 \\ \sqrt{2} & -\sqrt{2} \\ \sqrt{2} & -\sqrt{2} \end{pmatrix}.$$

Also, with a formula on page 20 of the lecture notes,

$$\rho(x^1, c^1) = \rho(x^4, c^1) = \frac{\sqrt{\lambda_1}}{\sigma(x^1)}v_1^1 = \frac{\sqrt{44/5}}{\sqrt{4/5}}0 = 0,$$

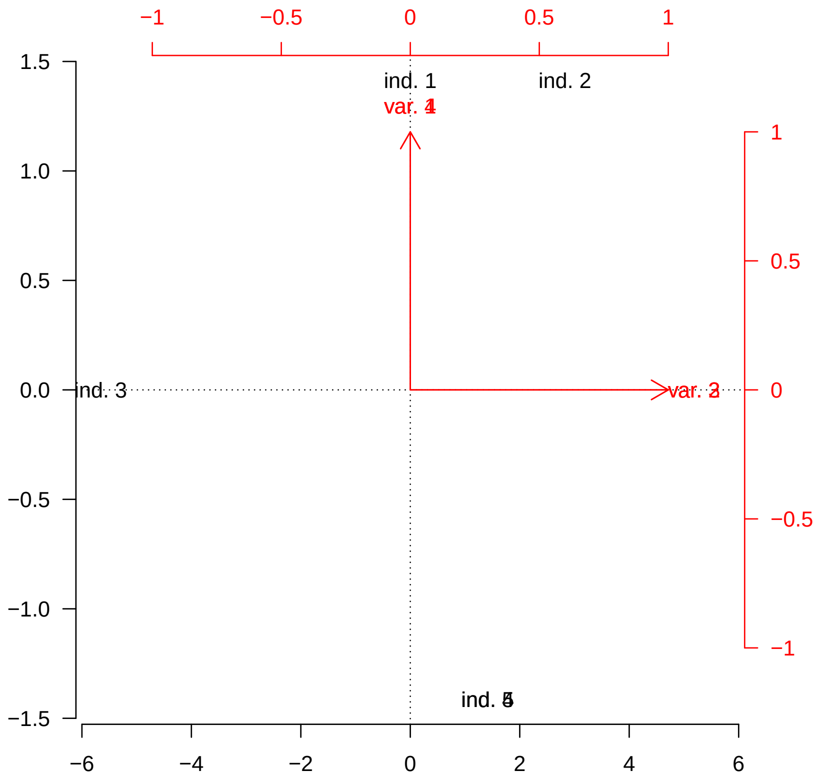
$$\rho(x^1, c^2) = \rho(x^4, c^2) = \frac{\sqrt{\lambda_2}}{\sigma(x^1)}v_1^2 = \frac{\sqrt{8/5}}{\sqrt{4/5}}\frac{1}{\sqrt{2}} = 1,$$

$$\rho(x^2, c^1) = \rho(x^3, c^1) = \frac{\sqrt{\lambda_1}}{\sigma(x^2)} v_2^1 = \frac{\sqrt{44/5}}{\sqrt{22/5}} \frac{1}{\sqrt{2}} = 1,$$

$$\rho(x^2, c^2) = \rho(x^3, c^2) = \frac{\sqrt{\lambda_2}}{\sigma(x^2)} v_2^2 = \frac{\sqrt{8/5}}{\sqrt{22/5}} 0 = 0.$$

4) *Draw a biplot (Section 1.2.4 of the lecture notes).*

Correction:



Remark that the plots of individuals 4 and 5 are overlapping. Remark that the plots of variables 2 and 3 are overlapping. Remark that the plots of variables 1 and 4 are overlapping.

Exercise 15

We consider n individuals, where each of them has the two qualitative variables employment and education. Employment takes the $p = 2$ values "employed" (E) and "unemployed" (U). Education takes the 3 values "Up to High school" (HS), "Undergraduate degree" (U) and "Graduate degree" (G). The n individuals are given by the data matrix

$$\begin{pmatrix} E & HS \\ E & U \\ U & HS \\ U & G \\ E & G \\ E & G \\ U & HS \\ U & HS \\ E & U \\ U & U \\ U & HS \end{pmatrix},$$

with column 1 for employment and column 2 for education.

1) *Construct the contingency table, with the marginal totals and grand total (page 26 of the lecture notes).*

Correction:

	HS	U	G	Total
E	$n_{11} = 1$	$n_{12} = 2$	$n_{13} = 2$	$n_{1.} = 5$
U	$n_{21} = 4$	$n_{22} = 1$	$n_{23} = 1$	$n_{2.} = 6$
Total	$n_{.1} = 5$	$n_{.2} = 3$	$n_{.3} = 3$	$n = 11$

2) *Compute the line profiles matrix P_1 and the corresponding center of gravity g_1*

Correction:
We have

$$P_1 = \begin{pmatrix} \frac{n_{11}}{n_{1.}} & \frac{n_{12}}{n_{1.}} & \frac{n_{13}}{n_{1.}} \\ \frac{n_{21}}{n_{2.}} & \frac{n_{22}}{n_{2.}} & \frac{n_{23}}{n_{2.}} \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \\ \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

The center of gravity is the average of the two transposed lines of P_1 (average of two individuals), with the weights given by $n_{1.}/n$ and $n_{2.}/n$:

$$g_1 = \frac{5}{11} \begin{pmatrix} \frac{1}{5} \\ \frac{2}{5} \\ \frac{2}{5} \end{pmatrix} + \frac{6}{11} \begin{pmatrix} \frac{4}{6} \\ \frac{1}{6} \\ \frac{1}{6} \end{pmatrix} = \begin{pmatrix} 0.45 \\ 0.27 \\ 0.27 \end{pmatrix}.$$

3) *Compute the column profiles matrix P_2 and the corresponding center of gravity g_2*

Correction:
We have

$$P_2 = \begin{pmatrix} \frac{n_{11}}{n_{.1}} & \frac{n_{21}}{n_{.1}} \\ \frac{n_{12}}{n_{.2}} & \frac{n_{22}}{n_{.2}} \\ \frac{n_{13}}{n_{.3}} & \frac{n_{23}}{n_{.3}} \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & \frac{4}{5} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

The center of gravity is the average of the three transposed lines of P_2 (average of three individuals), with the weights given by $n_{.1}/n$, $n_{.2}/n$ and $n_{.3}/n$:

$$g_2 = \frac{5}{11} \begin{pmatrix} \frac{1}{5} \\ \frac{4}{5} \end{pmatrix} + \frac{3}{11} \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \end{pmatrix} + \frac{3}{11} \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 0.45 \\ 0.54 \end{pmatrix}.$$

Exercise 16

1) *In the context of page 30 of the lecture notes, show that $||g_1||^2_{M_1} = 1$.*

Correction:
Remark that, using the expressions of g_1 and M_1 page 30 of the lecture notes,

$$||g_1||^2_{M_1} = {}^t g_1 M_1 g_1 = \frac{1}{n} {}^t\mathbf{1}_q D_2^{-1} n D_2 \frac{1}{n} D_2^{-1} \mathbf{1}_q = \frac{1}{n} {}^t\mathbf{1}_q D_2^{-1} \mathbf{1}_q = \frac{1}{n} \sum_{i=1}^{q} (D_2^{-1})_{ii} = \frac{1}{n} \sum_{i=1}^{q} n_{.i} = 1.$$

2) *In the context of page 30 of the lecture notes, show that $\Sigma_1 M_1$ and ${}^t T D_1 T D_2$ have the same eigenvalues apart from the one associated to g_1.*

Correction:
From the spectral theorem, there exist $q \times 1$ vectors $w_1, \dots, w_q$ that are eigenvectors of $\Sigma_1 M_1$, that are M_1-orthonormal and that are associated to the eigenvalues $\lambda_1 \geq \dots \geq \lambda_q$. We know from page 30 of the lecture notes that 0 is an eigenvalue of $\Sigma_1 M_1$, so $\lambda_q = 0$. Let $r \leq q - 1$ be defined by $\lambda_r > 0$ and $\lambda_{r+1} = 0$. From page 30 of the lecture notes we know that $\Sigma_1 M_1 g_1 = 0$. Hence $g_1 \in \text{span}(w_{r+1}, \dots, w_q)$. Hence, by a change of M_1-orthonormal basis, we can construct $w'_{r+1}, \dots, w'_{q-1}$ that are M_1-orthonormal and such that $\text{span}(w'_{r+1}, \dots, w'_{q-1}, g_1) = \text{span}(w_{r+1}, \dots, w_q)$. Hence, without loss of generality, we consider that $w_q = g_1$.

Consider $i \in \{1, \dots, q-1\}$. w_i and g_1 are M_1-orthogonal. Hence, using $M_1 = nD_2$ in the lecture notes,

$$n g_1 {}^t g_1 D_2 w_i = g_1 {}^t g_1 M_1 w_i = g_1 \langle g_1, w_i \rangle_{M_1} = 0.$$

Hence, from the decomposition of $\Sigma_1 M_1$ in the lecture notes

$$\lambda_i w_i = \Sigma_1 M_1 w_i = {}^t T D_1 T D_2 w_i - n g_1 {}^t g_1 D_2 w_i = {}^t T D_1 T D_2 w_i.$$

Hence, the eigenvalues of ${}^t T D_1 T D_2$, other than the one associated to g_1 are $\lambda_1 \geq \dots \geq \lambda_{q-1}$ which are the same as the eigenvalues of $\Sigma_1 M_1$ other than the one associated to g_1.

Exercise 17

1) *In the context of page 31 of the lecture notes, show that the data from the matrix $C^{(1)}$ have $\kappa \times 1$ mean vector 0, with the weight matrix W_1.*

Correction:
The $1 \times \kappa$ mean vector is, using the expressions of W_1, $C^{(1)}$ and g_1 on pages 30 and 31 of the lecture notes,

$${}^t \mathbf{1}_p W_1 C^{(1)} = {}^t \mathbf{1}_p \frac{1}{n} D_1^{-1} n D_1 T D_2 V_1 = {}^t \mathbf{1}_p T D_2 V_1 = n {}^t g_1 D_2 V_1 = {}^t g_1 M_1 V_1.$$

The component i of ${}^t g_1 M_1 V_1$ is $\langle g_1, v_i \rangle_{M_1}$ where v_i is one of the eigenvectors of ${}^t T D_1 T D_2$ that is M_1-orthogonal to g_1 by definition. Hence $\langle g_1, v_i \rangle_{M_1} = 0$ and thus the component i is zero and thus the $1 \times \kappa$ mean vector is zero.

Exercise 18

We consider $n = 8$ voters, where each of them has the two qualitative variables work and preference. Work takes the $p = 3$ values "Dentist" (Den), "Teacher" (T) and "Developper" (Dev). Preference takes the 3 values "Left" (L), "Center" (C) and "Right" (R). The n individuals are given by the data matrix

$$\begin{pmatrix} Den & R \\ T & L \\ T & L \\ Dev & C \\ Dev & C \\ Den & R \\ T & L \\ Dev & C \end{pmatrix},$$

with column 1 for work and column 2 for preference. In this exercize, you have the choice between giving exact expressions of the results (using fractions, square roots,...) or giving (approximate) numerical results.

1) *Construct the contingency table, with the marginal totals and grand total (page 26 of the lecture notes).*

Correction:

	L	C	R	Total
Den	$n_{11} = 0$	$n_{12} = 0$	$n_{13} = 2$	$n_{1.} = 2$
T	$n_{21} = 3$	$n_{22} = 0$	$n_{23} = 0$	$n_{2.} = 3$
Dev	$n_{31} = 0$	$n_{32} = 3$	$n_{33} = 0$	$n_{3.} = 3$
Total	$n_{.1} = 3$	$n_{.2} = 3$	$n_{.3} = 2$	$n = 8$

2) *Compute the lines profile matrix P_1, the center of gravity g_1, the matrix D_1, the matrix D_2, the weight matrix W_1 and the distance matrix M_1.*

Correction:

Using the formulas given in the lecture notes, we have

$$D_1 = \begin{pmatrix} \frac{1}{n_{1.}} & 0 & 0 \\ 0 & \frac{1}{n_{2.}} & 0 \\ 0 & 0 & \frac{1}{n_{3.}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix},$$

$$P_1 = D_1 T = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0 & 0 & 2 \\ 3 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$W_1 = \begin{pmatrix} \frac{n_{1.}}{n} & 0 & 0 \\ 0 & \frac{n_{2.}}{n} & 0 \\ 0 & 0 & \frac{n_{3.}}{n} \end{pmatrix} = \begin{pmatrix} \frac{2}{8} & 0 & 0 \\ 0 & \frac{3}{8} & 0 \\ 0 & 0 & \frac{3}{8} \end{pmatrix},$$

$$D_2 = \begin{pmatrix} \frac{1}{n_{.1}} & 0 & 0 \\ 0 & \frac{1}{n_{.2}} & 0 \\ 0 & 0 & \frac{1}{n_{.3}} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix},$$

$$M_1 = nD_2 = 8 \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{8}{3} & 0 & 0 \\ 0 & \frac{8}{3} & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

and

$$g_1 = \frac{1}{n} {}^t T \mathbf{1}_p = \frac{1}{8} \begin{pmatrix} 0 & 3 & 0 \\ 0 & 0 & 3 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{8} \\ \frac{3}{8} \\ \frac{2}{8} \end{pmatrix}.$$

3) *To perform the PCA on the line profiles, provide the matrix that has the* $\kappa = \min(p, q) - 1$ *non trivial eigenvalues* $\lambda_1 \geq \cdots \geq \lambda_\kappa \geq 0$*. Compute* $\kappa, \lambda_1, \ldots, \lambda_\kappa$*. Compute also the* $q \times \kappa$ *matrix* V_1 *which columns are the* M_1*-orthogonal eigenvectors corresponding to these non-trivial eigenvalues (the choice of these* κ *eigenvectors will not be unique, so you can take any choice you want that is valid).*

Correction:

Here $\kappa = \min(3,3) - 1 = 2$. From pages 30, 31 of the lecture notes, the matrix is ${}^t T D_1 T D_2$, which has the trivial eigenvalue 1 and the other eigenvalues λ_1, λ_2. We have

$${}^t T D_1 T D_2 = \begin{pmatrix} 0 & 3 & 0 \\ 0 & 0 & 3 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0 & 0 & 2 \\ 3 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Here the three eigenvalues of the matrix are 1. Hence $\lambda_1 = \lambda_2 = 1$. Here the trivial eigenvector is

$$g_1 = \begin{pmatrix} \frac{3}{8} \\ \frac{3}{8} \\ \frac{2}{8} \end{pmatrix},$$

so the other two eigenvectors must be M_1-orthogonal to g_1. Here any vector is an eigenvector of the identity matrix, so there is not an unique choice of the two eigenvectors. Let us find a first vector of the form ${}^t(a, b, c)$ that is M_1-orthogonal to g_1 and has M_1-norm 1. This yields

$$0 = \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} \frac{8}{3} & 0 & 0 \\ 0 & \frac{8}{3} & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} \frac{3}{8} \\ \frac{3}{8} \\ \frac{2}{8} \end{pmatrix} = a + b + c$$

and

$$1 = \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} \frac{8}{3} & 0 & 0 \\ 0 & \frac{8}{3} & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \frac{8a^2}{3} + \frac{8b^2}{3} + 4c^2.$$

Hence we take (arbitrarily) the first non-trivial eigenvector as

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{4} \\ -\frac{\sqrt{3}}{4} \\ 0 \end{pmatrix}$$

Let us find a second non-trivial eigenvector of the form ${}^t(c,d,e)$ that has M_1-norm 1 and is M_1-orthogonal to g_1 and ${}^t(a,b,c)$. From before, we thus have

$$c+d+e=0$$

and

$$\frac{8c^2}{3}+\frac{8d^2}{3}+4e^2=1.$$

We also have

$$0=\begin{pmatrix}c & d & e\end{pmatrix}\begin{pmatrix}\frac{8}{3} & 0 & 0\\ 0 & \frac{8}{3} & 0\\ 0 & 0 & 4\end{pmatrix}\begin{pmatrix}\frac{\sqrt{3}}{4}\\ -\frac{\sqrt{3}}{4}\\ 0\end{pmatrix}=\frac{8}{3}\frac{\sqrt{3}}{4}(c-d).$$

Hence we take (c,d,e) of the form t$(1,1,-2)$, with $t\geq 0$ such that

$$\frac{8t^2}{3}+\frac{8t^2}{3}+4(-2t)^2=1$$

and so

$$1=t^2\left(\frac{8}{3}+\frac{8}{3}+16\right)=t^2(\frac{64}{3}).$$

Hence we take $t=\sqrt{3}/\sqrt{64}$ and thus we take the second eigenvector as

$$\begin{pmatrix}c\\ d\\ e\end{pmatrix}=\begin{pmatrix}\frac{\sqrt{3}}{\sqrt{64}}\\ \frac{\sqrt{3}}{\sqrt{64}}\\ -2\frac{\sqrt{3}}{\sqrt{64}}\end{pmatrix}.$$

Hence eventually

$$V_1=\begin{pmatrix}\frac{\sqrt{3}}{4} & \frac{\sqrt{3}}{\sqrt{64}}\\ -\frac{\sqrt{3}}{4} & \frac{\sqrt{3}}{\sqrt{64}}\\ 0 & -2\frac{\sqrt{3}}{\sqrt{64}}\end{pmatrix}\approx\begin{pmatrix}0.43 & 0.22\\ -0.43 & 0.22\\ 0 & -0.44\end{pmatrix}.$$

4) *Compute the principal component matrix $C^{(1)}$.*

Correction:
We have

$$C^{(1)}=P_1M_1V_1=\begin{pmatrix}0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}\begin{pmatrix}\frac{8}{3} & 0 & 0\\ 0 & \frac{8}{3} & 0\\ 0 & 0 & 4\end{pmatrix}\begin{pmatrix}0.43 & 0.22\\ -0.43 & 0.22\\ 0 & -0.44\end{pmatrix}=\begin{pmatrix}0 & 0 & 4\\ \frac{8}{3} & 0 & 0\\ 0 & \frac{8}{3} & 0\end{pmatrix}\begin{pmatrix}0.43 & 0.22\\ -0.43 & 0.22\\ 0 & -0.44\end{pmatrix}\approx\begin{pmatrix}0 & -1.76\\ 1.15 & 0.59\\ -1.15 & 0.59\end{pmatrix}.$$

5) *Using the transition formulae, compute the the principal component matrix $C^{(2)}$.*

Correction:
The matrix of non-trivial eigenvalues is

$$\Lambda=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

Furthermore

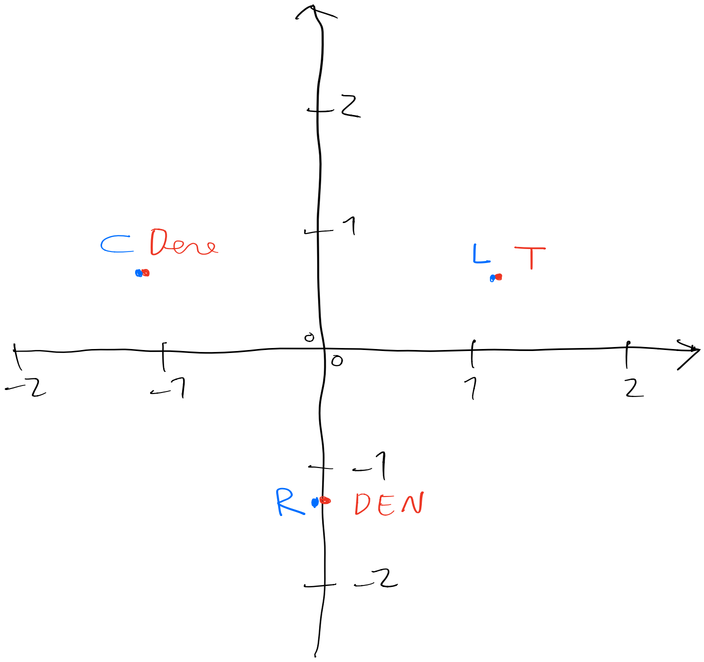
$$P_2=\begin{pmatrix}\frac{n_{11}}{n_{.1}} & \frac{n_{21}}{n_{.1}} & \frac{n_{31}}{n_{.1}}\\ \frac{n_{12}}{n_{.2}} & \frac{n_{22}}{n_{.2}} & \frac{n_{32}}{n_{.2}}\\ \frac{n_{13}}{n_{.3}} & \frac{n_{23}}{n_{.3}} & \frac{n_{33}}{n_{.3}}\end{pmatrix}=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{pmatrix}.$$

Hence, from the transition formulae, we have

$$C^{(2)}=P_2C^{(1)}\Lambda^{-1/2}=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{pmatrix}\begin{pmatrix}0 & -1.76\\ 1.15 & 0.59\\ -1.15 & 0.59\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}=\begin{pmatrix}1.15 & 0.59\\ -1.15 & 0.59\\ 0 & -1.76\end{pmatrix}.$$

6)
Plot the individuals of the two principal component matrix $C^{(1)}$ and $C^{(2)}$, with one color for each of the matrices, together with category name of each individual. Interpret the plot.

Correction:



We observe that the values of first variable are perfectly matched by the corresponding values of the second variable. More precisely:

- All voters who have the work variable equal to the value Den have the preference variable equal to the value R. In the plot we see that Den (in red) and R (in blue) coincide.
- All voters who have the work variable equal to the value Dev have the preference variable equal to the value C. In the plot we see that Dev (in red) and C (in blue) coincide.
- All voters who have the work variable equal to the value T have the preference variable equal to the value L. In the plot we see that T (in red) and L (in blue) coincide.

Exercise 19

1) *Prove that for $\ell, \ell' \in \{1, \ldots, m\}$, $(\bar{W})_{\ell,\ell'} = \bar{w}_\ell$ if $\ell = \ell'$ and $(\bar{W})_{\ell,\ell'} = 0$ if $\ell \neq \ell'$ (context of page 39 of the lecture notes).*

Correction: Let $\ell, \ell' \in \{1, \ldots, m\}$. We have

$$(\bar{W})_{\ell,\ell'} = ({}^tTWT)_{\ell,\ell'} = \sum_{a,b=1}^{n} ({}^tT)_{\ell,a} W_{a,b} T_{b,\ell'} = \sum_{a,b=1}^{n} T_{a,\ell} w_a \mathbf{1}_{a=b} T_{b,\ell'} = \sum_{a=1}^{n} w_a T_{a,\ell} T_{a,\ell'} = \sum_{a=1}^{n} w_a \mathbf{1}_{a \in \Omega_\ell} \mathbf{1}_{a \in \Omega_{\ell'}}.$$

Hence, since $\Omega_1, \dots, \Omega_m$ are a partition of $\{1, \dots, n\}$, if $\ell \neq \ell'$ then $\mathbf{1}_{a\in\Omega_\ell}\mathbf{1}_{a\in\Omega_{\ell'}} = 0$ for all $a \in \{1, \dots, n\}$ and thus $(\bar{W})_{\ell,\ell} = 0$. In the case $\ell = \ell'$, we obtain from the above computations

$$(\bar{W})_{\ell,\ell} = \sum_{a=1}^{n} w_a \mathbf{1}_{a\in\Omega_\ell} = \sum_{a\in\Omega_\ell} w_a = \bar{w}_\ell.$$

2) *Prove that*

$$\bar{W}^{-1}\, {}^t T W X = \begin{pmatrix} {}^t g_1 \\ \vdots \\ {}^t g_m \end{pmatrix}$$

(context of page 40 of the lecture notes).

Correction:
Let $\ell \in \{1, \dots, m\}$ and $j \in \{1, \dots, p\}$. We have

$$(\bar{W}^{-1}\, {}^t T W X)_{\ell,j} = \sum_{a=1}^{m} \sum_{b,c=1}^{n} (\bar{W}^{-1})_{\ell,a} ({}^t T)_{a,b} W_{b,c} X_{c,j} = \sum_{a=1}^{m} \sum_{b,c=1}^{n} (\bar{W}^{-1})_{\ell,a} T_{b,a} W_{b,c} x_c^j$$

$$= \sum_{a=1}^{m} \sum_{b,c=1}^{n} \mathbf{1}_{\ell=a} \frac{1}{\bar{w}_\ell} T_{b,\ell} \mathbf{1}_{b=c} w_b x_c^j = \sum_{c=1}^{n} \frac{1}{\bar{w}_\ell} T_{c,\ell} w_c x_c^j = \frac{1}{\bar{w}_\ell} \sum_{c=1}^{n} \mathbf{1}_{c\in\Omega_\ell} w_c x_c^j = \frac{1}{\bar{w}_\ell} \sum_{c\in\Omega_\ell} w_c x_c^j = (\frac{1}{\bar{w}_\ell} \sum_{c\in\Omega_\ell} w_c x_c)_j = (g_\ell)_j$$

$$= \begin{pmatrix} {}^t g_1 \\ \vdots \\ {}^t g_m \end{pmatrix}_{\ell,j} .$$

3) *Prove that* ${}^t\bar{G}\bar{W}\bar{G} = {}^t\bar{X}_b W \bar{X}_b$ *(context of page 41 of the lecture notes).*

Correction:
We have, for $j, k \in \{1, \dots, p\}$,

$$({}^t\bar{G}\bar{W}\bar{G})_{j,k} = \sum_{\ell,\ell'=1}^{m} ({}^t\bar{G})_{j,\ell} (\bar{W})_{\ell,\ell'} \bar{G}_{\ell',k} = \sum_{\ell,\ell'=1}^{m} \bar{G}_{\ell,j} \mathbf{1}_{\ell=\ell'} \bar{w}_\ell \bar{G}_{\ell',k} = \sum_{\ell=1}^{m} \bar{w}_\ell \bar{G}_{\ell,j} \bar{G}_{\ell,k} = \sum_{\ell=1}^{m} \left(\sum_{i\in\Omega_\ell} w_i \right) (g_\ell - g)_j (g_\ell - g)_k$$

$$= \sum_{\ell=1}^{m} \sum_{i\in\Omega_\ell} w_i (g_\ell - g)_j (g_\ell - g)_k.$$

We remark that for $i \in \{1, \dots, n\}$, $\ell \in \{1, \dots, m\}$, with $i \in \Omega_\ell$, we have that ${}^t(g_\ell - g)$ is the line i of $\bar{X}_b$. Hence we obtain

$$({}^t\bar{G}\bar{W}\bar{G})_{j,k} = \sum_{\ell=1}^{m} \sum_{i\in\Omega_\ell} w_i (\bar{X}_b)_{i,j} (\bar{X}_b)_{i,k} = \sum_{i=1}^{n} w_i (\bar{X}_b)_{i,j} (\bar{X}_b)_{i,k} = \left({}^t\bar{X}_b W \bar{X}_b\right)_{j,k} .$$

Exercise 20

1) *Here, you can use the concept of rank from linear algebra. For a matrix K of size $a \times b$, its rank* $\operatorname{rank}(K)$ *satisfies* $\operatorname{rank}(K) \leq \min(a, b)$. *Also, if ${}^t v K = 0$ for a non-zero $a \times 1$ vector v, then* $\operatorname{rank}(K) \leq a - 1$. *Also, for two (rectangular) matrices A, B we have* $\operatorname{rank}(AB) \leq \min(\operatorname{rank}(A), \operatorname{rank}(B))$. *Finally, the number of non-zero eigenvalues of a matrix is smaller than its rank.*

Assume that Σ is invertible. Consider the eigenvalues $\lambda_1 \geq \dots \geq \lambda_p$ of $\Sigma_b \Sigma^{-1}$ (context of page 41 of the lecture notes). Show that at most $\min(p, \kappa - 1)$ *elements of* $(\lambda_1, \dots, \lambda_p)$ *are non-zero.*

Correction:

If $\kappa \geq p+1$, this is trivially true because $\Sigma_b \Sigma^{-1}$ is $p \times p$. Assume now that $\kappa \leq p$. Let $\mathbf{1}_\kappa$ be the $\kappa \times 1$ vector composed on ones. We have $\Sigma_b = {}^t\bar{G}\bar{W}\bar{G}$, where $\bar{W}\bar{G}$ is $\kappa \times p$ and satisfies ${}^t\mathbf{1}_\kappa(\bar{W}\bar{G}) = 0$, because the element j of $\mathbf{1}_\kappa(\bar{W}\bar{G})$ is $\sum_{\ell=1}^{\kappa} \bar{w}_\ell((g_\ell)_j - g_j) = g_j - g_j = 0$. Hence, $\text{rank}(\bar{W}\bar{G}) \leq \kappa - 1$. Hence, $\text{rank}(\Sigma_b) = \text{rank}({}^t\bar{G}\bar{W}\bar{G}) \leq \min(\text{rank}({}^t\bar{G}), \text{rank}(\bar{W}\bar{G})) \leq \kappa - 1$. Hence $\text{rank}(\Sigma_b\Sigma^{-1}) \leq \min(\kappa - 1, \text{rank}(\Sigma^{-1})) \leq \kappa - 1$. Hence at most $\kappa - 1$ eigenvalues of $\Sigma_b\Sigma^{-1}$ are non-zero which concludes the proof.

2) *Here we will show that the eigenvalues of $\Sigma_b\Sigma^{-1}$ are in $[0,1]$.* **a)** *Show that the eigenvalues of $\Sigma_b\Sigma^{-1}$ are eigenvalues of $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}$.* **b)** *Show that these eigenvalues are positive.* **c)** *Then, you can use the following result from linear algebra: the largest eigenvalue of $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}$ is*

$$\max_{||x||=1} {}^tx\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}x.$$

Show that this largest eigenvalue is smaller than 1.

Correction: **a)** If x is non-zero and $\lambda \in \mathbb{R}$ is such that $\Sigma_b\Sigma^{-1}x = \lambda x$ then left-multiplying by $\Sigma^{-1/2}$ we obtain $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}(\Sigma^{-1/2}x) = \lambda(\Sigma^{-1/2}x)$ and so λ is an eigenvalue of $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}$.

b) For any $p \times 1$ vector x, we have ${}^tx\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}x = {}^t(\Sigma^{-1/2}x)\Sigma_b(\Sigma^{-1/2}x) \geq 0$ because Σ_b is a covariance matrix and thus it is non-negative definite. So $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}$ is also non-negative definite and thus its eigenvalues are non-negative.

c) Using $\Sigma_b = \Sigma - \Sigma_w$ we obtain for any $p \times 1$ vector x with $||x|| = 1$,

$${}^tx\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}x = {}^tx\Sigma^{-1/2}\Sigma\Sigma^{-1/2}x - {}^tx\Sigma^{-1/2}\Sigma_w\Sigma^{-1/2}x = ||x||^2 - {}^tx\Sigma^{-1/2}\Sigma_w\Sigma^{-1/2}x = 1 - {}^tx\Sigma^{-1/2}\Sigma_w\Sigma^{-1/2}x.$$

The quantity ${}^tx\Sigma^{-1/2}\Sigma_w\Sigma^{-1/2}x$ is non-negative from the same arguments as in b). More precisely, Σ_w is a weighted sum of covariance matrices with non-negative weights, so it is non-negative definite. Hence, we obtain ${}^tx\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}x \leq 1$. So from the linear algebra result of the exercise, the largest eigenvalue of $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}$ is smaller than 1.

3) *In the context of page 41 of the lecture notes, show that if $\lambda_1 = 0$, then the m centers of gravity of the m groups are equal.*

Correction:

We have shown in 2) that the non-zero eigenvalues of $\Sigma_b\Sigma^{-1}$ are between 0 and 1 and, by definition, λ_1 is the largest. Hence all the eigenvalues of $\Sigma_b\Sigma^{-1}$ are zero. If $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}x = \lambda x$ then, left multiplying by $\Sigma^{1/2}$, we obtain $\Sigma_b\Sigma^{-1}(\Sigma^{1/2}x) = \lambda(\Sigma^{1/2}x)$. Hence all the eigenvalues of $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}$ are zero. Hence the matrix $\Sigma^{-1/2}\Sigma_b\Sigma^{-1/2}$ is the zero matrix (this is seen by diagonalizing it). We can left- and right- multiply by $\Sigma^{1/2}$ and we obtain that Σ_b is the zero matrix. Hence the covariance matrix of the m centers of gravity $g_1, \ldots, g_m$ is zero, which means that all these centers of gravity are equal to their mean

$$\sum_{\ell=1}^{m} \bar{w}_\ell g_\ell.$$

Exercise 21

1) *Consider the setting of pages 50, 51 of the lecture notes. Consider $n = 20$ individuals with corresponding* 20 *values of a qualitative variable, taking values in $\{\tau_1, \tau_2, \tau_3\}$, given by*

$$t = {}^t\left(\tau_1, \tau_2, \tau_3, \tau_2, \tau_1, \tau_1, \tau_2, \tau_2, \tau_3, \tau_3, \tau_1, \tau_3, \tau_2, \tau_1, \tau_1, \tau_2, \tau_3, \tau_2, \tau_2, \tau_1\right).$$

Consider uniform weights. Assume that a partition of the input space has been obtained and that its first region R_1 contains the individuals $x_2, x_4, x_5, x_7, x_{10}, x_{12}, x_{15}, x_{16}, x_{20}$ and only these individuals. Compute the frequencies $\hat{p}_{11}, \hat{p}_{12}, \hat{p}_{13}$. Which value of the qualitative variable should be attributed to R_1, if a classifier is built from this partition of the input space? Compute the value of the Gini index $\mathcal{G}_1$ of R_1.

Correction: We have

$$\left(t_2, t_4, t_5, t_7, t_{10}, t_{12}, t_{15}, t_{16}, t_{20}\right) = \left(\tau_2, \tau_2, \tau_1, \tau_2, \tau_3, \tau_3, \tau_1, \tau_2, \tau_1\right).$$

Hence

$$\hat{p}_{11} = \frac{3}{9}, \quad \hat{p}_{12} = \frac{4}{9}, \quad \hat{p}_{13} = \frac{2}{9}.$$

R_1 should be attributed the most frequent value, that is τ_2. The value of the Gini index is

$$\mathcal{G}_1 = \sum_{\ell=1}^{3} \hat{p}_{1\ell}(1-\hat{p}_{1\ell}) = \frac{3}{9}\left(1-\frac{3}{9}\right)+\frac{4}{9}\left(1-\frac{4}{9}\right)+\frac{2}{9}\left(1-\frac{2}{9}\right) = \frac{3}{9}\times\frac{6}{9}+\frac{4}{9}\times\frac{5}{9}+\frac{2}{9}\times\frac{7}{9} = \frac{3\times 6+4\times 5+2\times 7}{81} = \frac{52}{81}.$$

2) *Consider the data matrix*

$$X = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$$

and the corresponding values of the qualitative variable

$$t = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_1 \\ \tau_2 \end{pmatrix}.$$

Consider uniform weights. Find $j^\star \in \{1,2\}$ and $s^\star \in \mathbb{R}$ that minimize

$$\bar{w}_1(j,s)\mathcal{G}_1(j,s) + \bar{w}_2(j,s)\mathcal{G}_2(j,s)$$

(context of page 51 of the lecture notes).

Correction:
We see that $t_j = \tau_1$ when $x_j^2 = 0$ and $t_j = \tau_2$ when $x_j^2 = 1$. So splitting across the second variable should be optimal. Let us show this. Let $j^\star = 2$ and $s^\star = 0.5$ (or any number strictly between 0 and 1). Then we have $\bar{w}_1(j^\star, s^\star) = 1/2$ and $\bar{w}_2(j^\star, s^\star) = 1/2$. The first region

$$R_1(j^\star, s^\star) = \{x = {}^t(x^1, x^2) \in \mathbb{R}^2; x^2 \leq 0.5\}$$

contains x_1 and x_3 with corresponding values t_1 and t_3 both equal to τ_1. Hence we have $\hat{p}_{11} = 1$ and $\hat{p}_{12} = 0$ and thus $\mathcal{G}_1(j^\star, s^\star) = 0$. Similarly the second region

$$R_2(j^\star, s^\star) = \{x = {}^t(x^1, x^2) \in \mathbb{R}^2; x^2 > 0.5\}$$

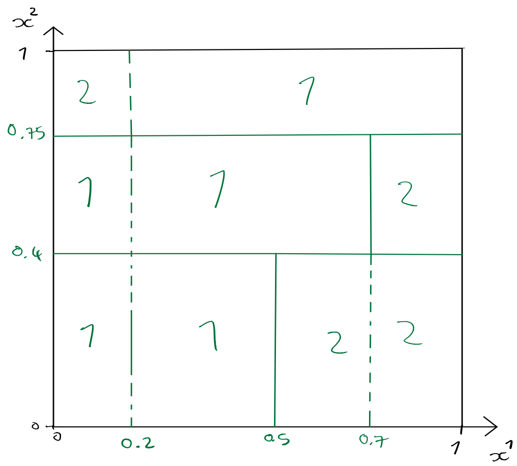
contains x_2 and x_4 with corresponding values t_2 and t_4 both equal to τ_2. Hence we have $\hat{p}_{21} = 0$ and $\hat{p}_{22} = 1$ and thus $\mathcal{G}_2(j^\star, s^\star) = 0$. Hence we have, since the Gini index is always non-negative,

$$\bar{w}_1(j^\star, s^\star)\mathcal{G}_1(j^\star, s^\star) + \bar{w}_2(j^\star, s^\star)\mathcal{G}_2(j^\star, s^\star) = 0 \leq \bar{w}_1(j,s)\mathcal{G}_1(j,s) + \bar{w}_2(j,s)\mathcal{G}_2(j,s),$$

for any $j \in \{1,2\}$ and $s \in \mathbb{R}$.

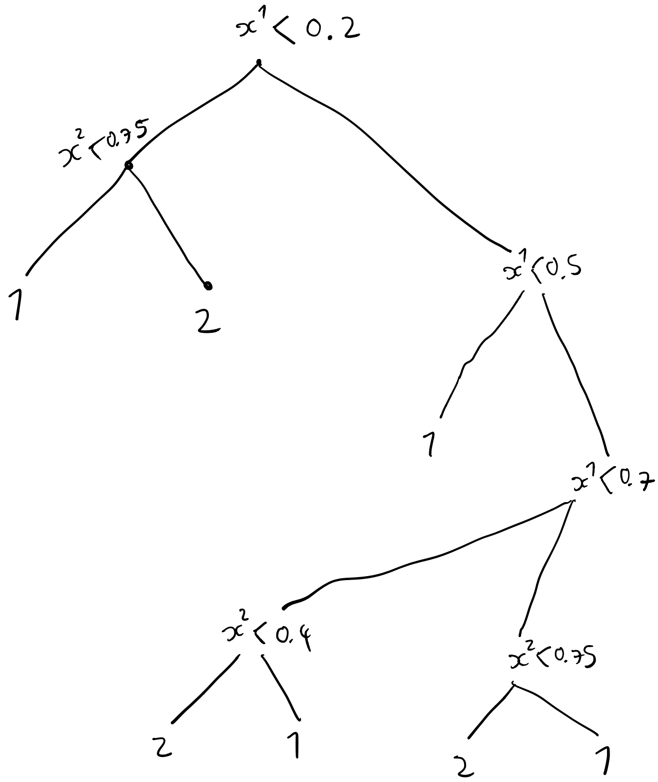
Exercise 22

The following plot is a partition of $[0,1]^2$ obtained from a classification tree.



Plot a classification tree leading to this partition. Ignore the problem of determining how input points located on the green straight and dashed lines (equality cases in the classification tree) are classified.

Correction:
With the convention of going left if the inequality is true.



Exercise 23

Consider two qualitative variables x^1 with values in $\{1,2\}$ and x^2 with values in $\{1,2\}$. Consider the corresponding data matrix

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 2 & 1 \\ 2 & 1 \\ 2 & 2 \end{pmatrix}.$$

Compute $d_{\chi^2}(1,2)$ (context of page 58 of the lecture notes).

Correction:

We have $p = m_1 = m_2 = 2$ and $n = 6$.

$$
\begin{aligned}
d_{\chi^2}(1,2) &= \frac{n}{p} \sum_{j=1}^{p} \sum_{\ell=1}^{m_j} \frac{\delta_{j\ell}(1,2)}{\#\{i \ s.t. \ x_i^j = \ell\}} \\
&= \frac{6}{2} \sum_{j=1}^{2} \sum_{\ell=1}^{2} \frac{\delta_{j\ell}(1,2)}{\#\{i \ s.t. \ x_i^j = \ell\}} \\
&= 3 \left(\frac{\delta_{11}(1,2)}{\#\{i \ s.t. \ x_i^1 = 1\}} + \frac{\delta_{12}(1,2)}{\#\{i \ s.t. \ x_i^1 = 2\}} + \frac{\delta_{21}(1,2)}{\#\{i \ s.t. \ x_i^2 = 1\}} + \frac{\delta_{22}(1,2)}{\#\{i \ s.t. \ x_i^2 = 2\}} \right) \\
&= 3 \left(\frac{0}{3} + \frac{0}{3} + \frac{1}{3} + \frac{1}{3} \right) \\
&= 2.
\end{aligned}
$$