ML: Algo & Theory
SS 18

1	2	3	4	5	$\sum$

Florence Lopez (3878792),
florence.lopez@student.uni-tuebingen.de
Jennifer Them (),
jennifer.them@student.uni-tuebingen.de

Assignment 1

(Abgabe am 24. April 2018)

Exercise 1

a.) In the following we will abbreviate small with s, medium with me, large with l, female with f and male with m:

- $P(Y = s) = P(X = m, Y = s) + P(X = f, Y = s) = 0.1 + 0.3 = 0.4$
- $P(Y = me) = P(X = m, Y = me) + P(X = f, Y = me) = 0.15 + 0.1 = 0.25$
- $P(Y = l) = P(X = m, Y = l) + P(X = f, Y = l) = 0.25 + 0.1 = 0.35$
- $P(X = m) = P(X = m, Y = s) + P(X = m, Y = me) + P(X = m, Y = l) = 0.1 + 0.15 + 0.25 = 0.5$
- $P(X = f) = P(X = f, Y = s) + P(X = f, Y = me) + P(X = f, Y = l) = 0.3 + 0.1 + 0.1 = 0.5$

b.) Proof: Choose a random $\epsilon > 0$. According to the Tchebycheff Inequality, we set $k = \epsilon$ and $X = \overline{X}$. $\overline{X}$ needs to be a random variable within the same distribution as the empirical mean $\overline{x}$. Then the following applies:

$P(|\overline{X} - E(\overline{X})| \geq \epsilon) \leq \frac{Var(\overline{X})}{\epsilon^2}$.

Because $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, E(X_i) = E(X), Var(X_i) = Var(X)$ and since all the X_i are iid, we get:

$E(\overline{X}) = E(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{\sum_{i=1}^{n} E(X)}{n} = \frac{n \cdot E(X)}{n} = E(X)$.
$Var(\overline{X}) = Var(\frac{\sum_{i=1}^{n} X_i}{n^2}) = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \frac{\sum_{i=1}^{n} Var(X)}{n^2} = \frac{n \cdot Var(X)}{n^2} = \frac{Var(X)}{n}$.

Therefore we get the following: $P(|\overline{X} - E(X)| \geq \epsilon) \leq \frac{Var(X)}{n \cdot \epsilon^2}$. The $lim_{n\to\infty} \frac{Var(X)}{n \cdot \epsilon^2} = 0$. This leads to: $P(|\overline{X} - E(X)| > \epsilon) \leq P(|\overline{X} - E(X)| \geq \epsilon)$, which proves the weak law of large numbers.

Exercise 2

a.) $P(X = me|Y = f) = \frac{P(X=me \cap Y=f)}{P(Y=f)} = \frac{0.1}{0.5} = 0.02$

b.) Two random variables X, Y are indepent if and only if $P(X \cap Y) = P(X) \cdot P(Y)$ or if $P(X|Y) = P(X)$. This means that the random variable Y has nor influence on the probability of X and vice-versa.

c.) In the following we will abbreviate a positive test with $+$, a negative test with $-$, cancer with c and no cancer with $\not{c}$:

- $P(A = +|B = c) = 0.95$
  $P(A = -|B = \not{c}) = 0.95$
  $P(B = c) = 0.01$
- $P(B = c|A = +) = \frac{P(A=+|B=c) \cdot P(B=c)}{P(A=+)}$
- $P(A = +)$ can be calculated with marginalisation:
  $P(A = +) = P(A = +|B = c) \cdot P(B = c) + P(A = +|B = \not{c}) \cdot P(B = \not{c})$, with $P(A = +|B = \not{c})$ being 0.05. This results in $P(A = +) = 0.95 \cdot 0.01 + 0.05 \cdot 0.99 = 0.059$
- This results in: $P(B = c|A = +) = \frac{0.95 \cdot 0.01}{0.059} = 0.16$

Exercise 3

a.) $A \cdot x = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 5 & 7 & 8 \end{pmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_1 \end{bmatrix} \cdot x_1 + \begin{bmatrix} a_2 \end{bmatrix} \cdot x_2 + \begin{bmatrix} a_3 \end{bmatrix} \cdot x_3 = \begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix} \cdot x_1 + \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} \cdot x_2 + \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix} \cdot x_3 = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_1 + 2x_2 + x_3 \\ 5x_1 + 7x_2 + 8x_3 \end{bmatrix}$

b.) Columns: If the columns of A should form a Basis of $\mathbb{R}^3$, one needs to test if the column-vectors are linearly independent and if they are able to produce every single possible vector of $\mathbb{R}^3$. This is done in the following:
$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda_1 \cdot \begin{bmatrix} 1 \\ 1 \\ 5 \end{bmatrix} + \lambda_2 \cdot \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} + \lambda_3 \cdot \begin{bmatrix} 2 \\ 1 \\ 8 \end{bmatrix}.$

This results in the following LGS:

$$
\begin{array}{rr}
\text{I)} & \lambda_1 + \lambda_2 + 2\lambda_3 = x \\
\text{II)} & \lambda_1 + 2\lambda_2 + \lambda_3 = y \\
\text{III)} & 5\lambda_1 + 7\lambda_2 + 8\lambda_3 = z \\
\hline
\text{I)} & x - \lambda_2 - 2\lambda_3 = \lambda_1 \\
\lambda_1 \text{ in II)} & (x - \lambda_2 - 2\lambda_3) + 2\lambda_2 + \lambda_3 = y \\
\Leftrightarrow & x + \lambda_2 - \lambda_3 = y \\
\Leftrightarrow & y + \lambda_3 - x = \lambda_2 \\
\hline
\lambda_1, \lambda_2 \text{ in III)} & 5(x - \lambda_2 - 2\lambda_3) + 7(y + \lambda_3 - x) + 8\lambda_3 = z \\
\Leftrightarrow & 3x + 2y + 10\lambda_3 = z \\
\Leftrightarrow & (z - 3x - 2y)/10 = \lambda_3
\end{array}
$$

$\lambda_1, \lambda_2, \lambda_3$ are therefore all three $\in \mathbb{R}^3$ and explicit, which means that all three column vectors are linearly indepent and also able to build all possible vectors in $\mathbb{R}^3$, which means that there are a basis of $\mathbb{R}^3$.

Rows: If the rows of A should form a Basis of $\mathbb{R}^3$, one needs to test if the row-vectors are linearly independent and if they are able to produce every single possible vector of $\mathbb{R}^3$. This is done in the following:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda_1 \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + \lambda_2 \cdot \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \lambda_3 \cdot \begin{bmatrix} 5 \\ 7 \\ 8 \end{bmatrix}.$$

This results in the following LGS:

$$\begin{aligned}
\text{I)} && \lambda_1 + \lambda_2 + 5\lambda_3 = x \\
\text{II)} && \lambda_1 + 2\lambda_2 + 7\lambda_3 = y \\
\text{III)} && 2\lambda_1 + 1\lambda_2 + 8\lambda_3 = z \\
\hline
\text{I)} && x - \lambda_2 - 5\lambda_3 = \lambda_1 \\
\lambda_1 \text{ in II)} && (x - \lambda_2 - 5\lambda_3) + 2\lambda_2 + 7\lambda_3 = y \\
\Leftrightarrow && x + \lambda_2 + 2\lambda_3 = y \\
\Leftrightarrow && y - x - 2\lambda_3 = \lambda_2 \\
\hline
\lambda_1, \lambda_2 \text{ in III)} && 2(x - \lambda_2 - 5\lambda_3) + (y - x - 2\lambda_3) + 8\lambda_3 = z \\
\Leftrightarrow && x - 2\lambda_2 - 4\lambda_3 + y = z \\
\Leftrightarrow && 3x - y = z
\end{aligned}$$

Since λ_3 can be chosen freely the row vectors of A are not linearly indepent from each other, which means that they are not building a Basis of the $\mathbb{R}^3$.

c.) Considering $b = (2, 3, 12)$:
$A \cdot x = b \Leftrightarrow \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 5 & 7 & 8 \end{pmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 12 \end{bmatrix}$. This results in the following LGS:

$$\begin{aligned}
\text{I)} && x_1 + x_2 + 2x_3 = 2 \\
\text{II)} && x_1 + 2x_2 + x_3 = 3 \\
\text{III)} && 5x_1 + 7x_2 + 8x_3 = 12 \\
\hline
\text{I)} && 2 - x_2 - 2x_3 = x_1 \\
x_1 \text{ in II)} && 2 - x_2 - 2x_3 + 2x_2 + x_3 = 3 \\
\Leftrightarrow && 2 + x_2 - x_3 = 3 \\
\Leftrightarrow && 1 + x_3 = x_2 \\
\hline
x_1, x_2 \text{ in III)} && 5(2 - x_2 - 2x_3) + 7(1 + x_3) + 8x_3 = 12 \\
\Leftrightarrow && 10 - 5x_2 - 10x_3 + 7 + 7x_3 + 8x_3 = 12 \\
\Leftrightarrow && 5 + 7 = 12 \checkmark
\end{aligned}$$

This means that x_3 can be chosen freely. We chose $x_3 = 1$, which leads to $x_1 = -2$ and $x_2 = 2$. This leads to $x = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$

d.) – Column rank: The column rank is 3, since A has 3 linearly indepent column vectors.

– Row rank: The row rank is 2, since there are only 2 linearly indepent row vectors, namely $r_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ and $r_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

– Rank of A: The rank of A is 3, since the maximum rank of column and row rank is 3.

Exercise 4

a.) – $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ has the independent eigenvectors $u_1 = (1, 0)$ and $u_2 = (0, 1)$ and the eigenvalue $\lambda = 1$.

– $B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ has the independent eigenvectors $u_1 = (1, 1)$ with eigenvalue $\lambda_1 = 3$ and $u_2 = (-1, 1)$ with eigenvalue $\lambda_2 = -1$.

– $C = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has the independent eigenvector $u_1 = (1, 0)$ with eigenvalue $\lambda = 1$. In the case of this matrix, one can observe the following:

  * $x - axis > 0, y - axis > 0$: $C \cdot x$ is larger than x.
  * $x - axis > 0, y - axis < 0$: $C \cdot x$ is smaller than x.
  * $x - axis < 0, y - axis < 0$: $C \cdot x$ is larger than x.
  * $x - axis < 0, y - axis > 0$: $C \cdot x$ is smaller than x.

b.) In the following we will always use this equation: $A \cdot u_i = \lambda_i \cdot I \cdot u_i$.

(1) $(A + \alpha \cdot I) \cdot u_i = A \cdot u_i + \alpha \cdot I \cdot u_i = \lambda_i \cdot I \cdot u_i + \alpha \cdot I \cdot u_i = \lambda_i \cdot u_i + \alpha \cdot u_i = (\lambda_i + \alpha) \cdot u_i$. This results in the eigenvalue $\lambda_i + \alpha$.

(2) Since A is a symmetric matrix, the following applies: $A^T = A$. This results in $A^T A = AA$. Therefore we have: $AA \cdot u_i = A \cdot (\lambda_i \cdot u_i) = \lambda_i \cdot (A \cdot u_i) = \lambda_i \cdot (\lambda_i \cdot u_i) = \lambda_i^2 \cdot u_i$. This results in the eigenvalue λ_i^2.

(3) Analogous to (2) $A^T = A$, so this results in the eigenvalue λ_i^2.

(4) In general the following applies: $A \cdot u_i == \lambda_i \cdot u_i$. Now, we multiply A^{-1} to this equation and get: $A^{-1} \cdot A \cdot u_i = \lambda_i \cdot A^{-1} \cdot u_i \Leftrightarrow I \cdot u_i = \lambda_i \cdot A^{-1} \cdot u_i \Leftrightarrow \lambda_i^{-1} \cdot u_i = A^{-1} \cdot u_i$. Therefore the eigenvalues would be λ_i^{-1}.

c.) A matrix S can be decomposed in the following way: $S = U \cdot D \cdot V^T$. Given SS^T and $S^T S$, the following to equations apply:

$$S^T S = (UDV^T)^T \cdot (UDV^T) = VD^T U^T \cdot UDV^T = VD^T \cdot DV^T = V(D^T D)V$$
$$SS^T = (UDV^T) \cdot (UDV^T)^T = UDV^T \cdot VD^T U^T = UD \cdot D^T U^T = U(D^T D)U$$