Bio144, 6./7. April 2017

Practical part 7: Exercise on linear algebra

1. Let us use the following matrices and vectors:

$$\mathbf{A} = \begin{bmatrix} 4 & 2 & 3 \\ 1 & 4 & 6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 3 & 6 \\ -1 & -1 & 0 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$$

Calculate all of the following expressions, if they are defined. Solve at least a)-d) by hand.

a) $2 \cdot \mathbf{A}$	**b)** $\mathbf{A} + \mathbf{B}$	**c)** $\mathbf{A} \cdot \mathbf{B}^{\mathrm{T}}$	**d)** $\mathbf{A} \cdot \boldsymbol{x}$
e) $\mathbf{A} \cdot \mathbf{B}$	**f)** $\mathbf{B}^{\mathrm{T}} \cdot \boldsymbol{y}$	**g)** $\mathbf{A} \cdot \mathbf{A}^{\mathrm{T}}$	**h)** $\mathbf{A}^{\mathrm{T}} \cdot \mathbf{A}$
i) $\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{x}$	**j)** $\boldsymbol{x} \cdot \boldsymbol{x}^{\mathrm{T}}$		

R-hints:

- A matrix can be created for example as
  `A <- matrix(c(4,2,3,0,3,6),byrow=TRUE,nrow=2)`
- `t(A)` corresponds to $\mathbf{A}^{\mathrm{T}}$.
- Make sure you understand the difference between **A*B** and **A%*%B** .

2. a) Repeat the example given in slides 35–38 of lecture 6 using R.

b) Repeat the procedure from a) 100 times. In each iteration, store the estimated coefficient vector $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$. Note that the estimates will be a bit different in each iteration, because another residual vector `t.e` is used.

c) Plot the histograms of the βs, the coefficients as a function of the simulation number, and scatterplots of $\hat{\beta}_1$ vs. $\hat{\beta}_0$, $\hat{\beta}_2$ vs. $\hat{\beta}_0$ and $\hat{\beta}_2$ vs. $\hat{\beta}_1$.

R-hints: The 100 simulations can be generated elegantly as follows:

- Generate an error matrix of dimension 5×100. Each column contains the errors of one experiment:
  `t.E <- matrix(rnorm(500),ncol=...)`
- Use `t.E` to generate the observations:
  ```
  t.y <- matrix(rep(t.y,100),nrow=5,byrow=F)
  t.Y <- t.y + t.E
  ```
- Define a results matrix of dimension 3×100 that contains in each column the estimates $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ of an experiment, and fill it with the results of the

100 iterations:

```
r.coef <- matrix(NA,ncol=3, nrow=100)
for (i in 1:100) {
 r.coef[i,] <- lm(t.Y[,i] ~ x1 + x2)$coefficients
}
```