CK0146: Reality check (2017.1)

This is a reality check by my friend Pekka. The check is not obligatory, but you are encouraged to use it to validate your knowledge before the course. Clearly, you will not get any bonus points for it.

Submit your answers by WED March 22 at 23:59:59 Fortaleza time either electronically as PDF file in SIGAA or on paper to FC's office. If returned on paper, please notify FC by email about it.

Exercise R.1 (Algebra and probabilities). Let Ω be a finite set of all possible outcomes and $P : \Omega \to \mathbb{R}$ a probability measure that (by definition) satisfies $P(\omega) \geq 0$ for all $\omega \in \Omega$ and $\sum_{w \in \Omega} P(\omega) = 1$. Let $f : \Omega \to \mathbb{R}$ be an arbitrary function on Ω. Define the *expectation* of f by $\mathrm{E}[f(\omega)] = \sum_{\omega \in \Omega} P(\omega) f(w)$. The *variance* of f is defined by $\mathrm{Var}[f(\omega)] = \mathrm{E}[(f(\omega) - \mathrm{E}[f(\omega)])^2]$.

Use such definitions to show that:

- $\mathrm{E}[\cdot]$ is a linear operator (see `http://mathworld.wolfram.com/LinearOperator.html`)
- $\mathrm{Var}[f(\omega)] = \mathrm{E}[f(\omega)^2] - \mathrm{E}[f(\omega)]^2$. [*Hint*]: Proof is short if you use linearity.

Exercise R.2 (Matrices). Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a symmetric matrix with distinct eigenvalues $\lambda_i \in \mathbb{R}$ $(i = 1, \ldots, N)$, defined by $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ $(\mathbf{v}_i \in \mathbb{R}^N)$ and eigenvectors that satisfy $\mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$ where the Kronecker delta is $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. Let the spectral decomposition

$$\mathbf{B} = \sum_{i=1}^{N} \lambda_i \mathbf{v}_i^T \mathbf{v}_i.$$

Show that $\mathbf{B}$ has the same eigenvectors and eigenvalues of $\mathbf{A}$ (showing that $\mathbf{A} = \mathbf{B}$ is not needed).

Exercise R.3 (Algorithms). Fibonacci numbers $F(i)$ are defined for $i \in N$ recursively as $F(i+2) = F(i+1) + F(i)$, with $F(1) = F(2) = 1$. Using pseudo-code, write down an algorithm that takes $n \in N$ as an input and outputs the Fibonacci numbers from $F(1)$ to $F(n)$. Analyse the time complexity of your algorithm using $\mathcal{O}$-notation. Briefly discuss the efficiency of your algorithm.

Exercise R.4 (Data). The data stored in file `reality_check_2017_1.txt` has 4000 data items (rows), each having 16 real-valued variables (columns). Write a small program that loads the data, finds the two variables showing the largest variances and makes a scatterplot of the data items using these two variables. Attach a printout of your code and the scatterplot produced by your program as an answer. [*Note*] The program should choose the variables and plot without user interaction.