Machine Learning, Spring 2018: Exercise Sheet 1

Note: you'll need to know the material from the Appendix A in the lecture notes to work out these exercises.

Problem 1 (easy warm-up) In the lecture notes I (Section 2) I was describing the TICS system with the use of two random variables which I called Image and Caption. Possible values of the Image RV were all vectors in $[0, 1]^{144000}$, and of Caption, any word sequence of length at most 20 made from a given vocabulary. Generally speaking, a RV always comes with a set of technically possible values, called the *measure space* of that RV. In order to get more familiar with this concept of a RV and its associated measure space, consider a scientific study carried out by an experimental psychologist where subjects first listened to a recording of a series of 10 randomly picked numbers (between 0 and 9) spoken by a synthetic voice, then had to wait 5 seconds, then had to repeat as much as they could remember from that sequence, starting from the sequence's beginning in the right order. Per subject, this trial was repeated 20 times with freshly randomly picked number sequences. In some of the trials, the voice was speaking in a neutral tone, in other trials the voice was speaking in an excited tone. The purpose of the experiment was to find out to what extent (if any) information given by an excited speaker is better remembered than information from a neutral speaker. Your task: declare a choice of random variables, each one describing some aspect of a trial, such that together they convey all the information that is relevant for this study. For each RV that you think of, (i) give a brief description in plain English of what it measures, (ii) give it a telling name, (iii) describe its sample space, (iv) state whether it is discrete or continuous. – The relevant information in these trials can be captured by observations (that is, RVs) in various ways, there is not a uniquely correct choice of them. Give at least five RVs.

Problem 2 (very easy, a 1-liner). Here we use the standard shorthand notation $P(x, y)$ for $P(X = x, Y = y)$, $P(x \mid y)$ for $P(X = x \mid Y = y)$, etc. Prove the following *factorization formula*

(1) $$P(x, y, z) = P(x)\, P(y \mid x)\, P(z \mid x, y),$$

starting from the definition of conditional (discrete) probabilities

(2) $$P(u \mid v_1, \ldots, v_n) = P(u, v_1, \ldots, v_n) \,/\, P(v_1, \ldots, v_n).$$

Note: in statistics and machine learning in general, factorizing joint distributions into products of simpler distributions is a super common strategy.