DIT852 Introduction to Data Science and AI, SP1 2021

Assignment 5: Probability for data science

1. From a group of 5 women and 7 men, how many different committees consisting of 2 women and 3 men can be formed? What if two of the men are feuding and refuse to serve on the committee together?

2. Compute the probabilities for the following events assuming equally likely outcomes. In each subproblem define the sample space (the set of all outcomes) and compute its cardinality. Likewise, define the event as a set of outcomes and compute its cardinality. Use Cartesian products where appropriate.

    a) A fair four-sided die with face values 1, 2, 3, and 4 is rolled three times. What is the probability that the sum of rolls is larger than 8?
    b) A standard deck of 52 playing cards contains 4 kings. What is the probability that two cards drawn randomly from the deck are both kings?
    c) What is the probability for getting at least one king when drawing three cards from a deck of 52 cards? Solve using the multiplication rule.

3. Two percent of women aged 45 who participate in routine screening have breast cancer. Ninety percent of those with breast cancer have positive mammographies. Eight percent of the women who do not have breast cancer will also have positive mammographies. Given that a woman has a positive mammography, what is the probability that she has breast cancer?

4. Our discussion of the k-nearest neighbor classifier (KNN) focused on the dependency of the decision boundary on k. This hyperparameter k however also has an effect on how errors are handled. Consider a test data point x of class 0 (assuming binary classification) and the k nearest neighbors of x, $x1,x2,\ldots,xk$ with class labels $y1,y2,\ldots,yk$. Assume that x is far enough from the decision boundary, so that $y1=y2=\cdots=yk=0$ if there would be no errors. Also assume that labels are flipped independently with probability p = 0.15 due to error. What is the probability that the KNN misclassifies x when k = 1, and k = 3, respectively?

Submitting work

Please upload your homework as one PDF file (a scan of handwritten notes is acceptable). Include the name of the group and the students who worked on the assignment at the beginning of the document.

Deadline: Monday 4 October 2021 at 23:59.