STAT561 Homework 1

Spring 2019

January 18, 2019

1 Probability

1.1 Among employees of a certain firm, 70% know C/C++, 60% know Fortran, and 50% know both languages. What portion of programmers

(a) does not know Fortran?

(b) does not know Fortran and does not know C/C++?

(c) knows C/C++ but not Fortran?

(d) knows Fortran but not C/C++?

(e) If someone knows Fortran, what is the probability that he/she knows C/C++ too?

(f) If someone knows C/C++, what is the probability that he/she knows Fortran too?

1.2 A computer program is tested by 5 independent tests. If there is an error, these tests will discover it with probabilities 0.1, 0.2, 0.3, 0.4, and 0.5, respectively. Suppose that the program contains an error. What is the probability that it will be found

(a) by at least one test?

(b) by at least two tests?

(c) by all five tests?

2 Linear Algebra

Show that $\boldsymbol{A}^T(r\boldsymbol{B}+\boldsymbol{C}) = r(\boldsymbol{A}^T\boldsymbol{B}) + \boldsymbol{A}^T\boldsymbol{C}$ for random variables $\boldsymbol{A}; \boldsymbol{B}; \boldsymbol{C}$ and scalar r.

3 Expectation and Variance

3.1

(a) Show that $E(c\boldsymbol{X}) = cE(\boldsymbol{X})$ for constant c and random variable $\boldsymbol{X}$.

(b) Show that $Var(a) = 0$ for constant a.

(c) Show that $Cov(\boldsymbol{X}; \boldsymbol{Y}) = E(\boldsymbol{XY})$ if $E(\boldsymbol{X}) = E(\boldsymbol{Y}) = 0$ for random variables $\boldsymbol{X}; \boldsymbol{Y}$.

3.2 Suppose we have a discrete random variable $\boldsymbol{X}$ that takes on values 1, 3 with probability (1/3; 2/3), and, independently, $\boldsymbol{Z}$ that takes on values 5; 10 with probability (1/2; 1/2). What is expected value $E[\boldsymbol{X} - \boldsymbol{Z}]$ and $Var(\boldsymbol{X} - \boldsymbol{Z})$?

3.3

(a) Find the expectation of for any set of random variables, not necessarily independent and identically distributed (IID). What if they are IID?

(b) Find the variance of for any set of random variables, not necessarily independent and identically distributed (IID). What if they are IID?

4 Linear Regression on Real-data

Important Note:

This problem is for you to test your understanding of linear regression and ability of applying your knowledge in a real-life case. Please don't use scikit-learn built-in linear regression package but solve the problem from scratch. More specifically, by solving the problem from scratch, we mean you should go through the MLE approach as discussed in class step by step.

Firstly, download the datafile Advertisement.csv. The features of this dataset are:

- TV-ad: money spent on TV for each product in some market
- Radio-ad: money spent on Radio for each product in some market
- Newspaper-ad: money spent on Newspaper for each product in some market

Note: numbers all in thousands of dollars.
The responses of this dataset are:

- Sales (in thousands of pieces): sales of each product in some market

Let's play with the data and give the company some feedbacks. Try to answer the following questions and present your answers in a clear format (such as tables, graphs, etc.). (Although this case is rather simple, it is always important to think about the best way of presenting your data analysis results.):

(a) Is there a relationship between total ads and sales? What kind of relationship is that relationship (Hint: positive, negative)?

(b) How strong is the relationship between total ads expense and sales?

(c) Among the three types of ads, which one has the strongest relationship with sales? Which one has the weakest?

(d) If the company want you to predict the sales for next year on a given cost of advertisement in some market, can you predict that? For example, if the company plan to spend \$50,000 in TV, how many widgets can they expect to sale?

Evaluation of a model can be the hardest part in the data analysis. However, it is crucial. Show the company how confident do you feel about your model by whatever means you have in mind for this rather simple case. Try to convince the company that your analysis make sense. (Hint: R^2 values, confidence intervals, etc.).