CSCI 347
Homework 01

Problem 1 (2 points)

What are the two main types of attributes typically found in data?

Problem 2 (14 points)

Consider the following data matrix

$$D = \begin{array}{cccc} & X_1 & X_2 & X_3 \\ x_1 & 0.3 & 23 & 5.6 \\ x_2 & 0.4 & 1 & 5.2 \\ x_3 & 1.8 & 4 & 5.2 \\ x_4 & 6.0 & 50 & 5.1 \\ x_5 & -0.5 & 34 & 5.7 \\ x_6 & 0.4 & 19 & 5.4 \\ x_7 & 1.1 & 11 & 5.5 \end{array}$$

1. (2 points) What is the estimated mean of X_3?
2. (2 points) What is the estimated covariance between X_1 and X_3?
3. (2 points) What is the estimated multi-dimensional mean of D?
4. (2 points) What is the estimated variance of X_2?
5. (2 points) What is the covariance matrix of D?
6. (2 points) What is the estimated correlation between X_1 and X_3?
7. (2 points) What is the total variance D?

Problem 3 (6 points)

Given $a, b \in \mathbb{R}^4$ (that is a fancy way of saying that a and b are 4-dimensional vectors with real values) where

$$\begin{aligned} a &= \begin{bmatrix} 2.0 & 5.0 & -2.6 & 6.0 \end{bmatrix} \\ b &= \begin{bmatrix} 15.0 & 2.5 & 4.0 & 4.0 \end{bmatrix} \end{aligned}$$

1. (2 points) What is $\|a - b\|_2$?
2. (2 points) What is $\|a - b\|_1$?
3. (2 points) What is the cosine of the angle between a and b?

Problem 4 (3 points)

The following questions reference the Heart Disease data set from the UCI Machine Learning Repository:
`https://archive.ics.uci.edu/ml/datasets/Heart+Disease`

1. (1 point) One attribute is named “cigs”. What information is stored in the “cigs” attribute?
2. (1 point) How man rows (i.e., observations, entities, instances) are there in the data set?
3. (1 point) How man attributes are there in the data set?

Tips and Acknowledgements

Acknowledgements: Homework problems adapted from assignments of Veronika Strnadova-Neeley.