

# 18s1: COMP9417 Machine Learning and Data Mining

---

**Lectures:** Supervised Learning – Classification

**Topic:** Questions from lectures

**Last revision:** Sun Apr 8 17:25:20 AEST 2018

## Introduction

Some questions and exercises from the course lectures covering aspects of learning linear models (models “linear in the parameters”) for classification tasks.

## Nearest-neighbour classifier

**Question 1** Which  $p$ -norm corresponds to *Euclidean* distance ? Express Euclidean distance in terms of this norm.

**Question 2** Construct a small two dimensional data set and show, for two different query points, how classification by  $k$  nearest neighbour is affected by (a) the number of neighbours (i.e., the value of  $k$ ), and (b) the distance of the exemplars (data points) from the queries.

**Question 3** Outline how the idea of computing the arithmetic mean  $\mu$  of a set of labelled data points (positive or negative) in some Euclidean space can be used to construct a linear classifier. [HINT: consider the basic linear classifier referred to in the lecture notes on slides 15–19.]

**Question 4 (Challenge)** Complete the proof that the arithmetic mean minimises squared Euclidean distance (see slide 15 of the lecture notes). Find the minimum by taking the gradient (vector of partial derivatives) and setting to zero.

## Naive Bayes classifier

**Question 5** Consider the example application of Bayes Theorem on slides 48–52 in the lecture notes. Suppose the a second laboratory test is ordered for the same patient, and this test also returns a positive result. What are the posterior probabilities of *cancer* and  $\neg$ *cancer* following these two tests? Note: you can assume that the two tests are independent.

**Question 6** Work through the example of applying Naive Bayes to text on slides 109–118. Be sure you are clear on the difference between the multinomial and multivariate Bernoulli models.