

17s1: COMP9417 Machine Learning and Data Mining

Lectures: Linear Models for Classification

Topic: Questions from lectures

Last revision: Mon Apr 3 10:21:06 AEST 2017

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 $\vec{\mu}$ of a set of labelled data points in some Euclidean space can be used to construct a linear classifier. [HINT: consider the basic linear classifier referred to in the lecture notes.]

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

Naive Bayes classifier

Question 5 Consider example application of Bayes rule in the lecture notes. Suppose the doctor decides to order a second laboratory test for the same patient, and suppose the second test returns a positive result as well. What are the posterior probabilities of *cancer* and \neg *cancer* following these two tests? Assume that the two tests are independent.

Question 6 Work through the example of applying Naive Bayes to text on slides 110–116. Be sure you are clear on the difference between the multinomial and multivariate Bernoulli models.