Machine Learning and Pattern Recognition, Tutorial Sheet Number 2

School of Informatics, University of Edinburgh, Instructor: Chris Williams

1. If there were questions on tutorial sheet 1 which were not covered in the last tutorial, please check the answer sheet which will be available after the Friday in the week in which the tutorials are held. You can use those to check where you may have got stuck, and to help you in formulating questions to ask in the tutorials.

2. [Maximum likelihood training of a Gaussian classifier]
A training set consists of one dimensional examples from two classes. The training examples from class 1 are $\{0.5, 0.1, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.35, 0.25\}$ and from class 2 are $\{0.9, 0.8, 0.75, 1.0\}$. Fit a (one dimensional) Gaussian using Maximum Likelihood to each of these two classes. Also estimate the class probabilities π_1 and π_2 using Maximum Likelihood. What is the probability that the test point $x = 0.6$ belongs to class 1? Sketch the form of the decision boundary/ies, i.e. the location(s) where $p(\text{class } 1|x) = p(\text{class } 2|x) = 0.5$ (you are not required to calculate the location(s) exactly).

3. [A simple Bayesian classifier]
Suppose you wish to predict a binary class label y, given one variable x. You know the class label is created from x by thresholding at some value a: $y(x) = H(x - a)$, $H(x) = 1$ if $x \geq 0$, $H(x) = 0$ if $x < 0$[1], but that value a is unknown. You know that the true threshold is somewhere between 1 and 9 with equal probability. Consider the input value $x = 7$. Calculate the probability that the label corresponding to this input value is 1 under the model.

You are now told that the label corresponding to $x = 7$ is in fact 1. How does this update the distribution for a?

4. [Maximum likelihood for linear regression]
Given training data $\mathcal{D} = \{(x^n, y^n), n = 1, \ldots, N\}$, you decide to fit a regression model $y = mx + c$ to this data. Derive an expression for m and c in terms of $\mathcal{D}$ using maximum likelihood.

[1] $H(x)$ is known as the Heaviside function.