EEL709: Minor Test I

February 15, 2015

Paper code: X47*
Maximum Marks: 25

1. The following are some admissions statistics for the two BNon† degree programmes at the University of Nonsensical Studies. The numbers in the cells are to be interpreted as <No. of successful applicants>/<Total no. of applicants>.

	BNon Witchcraft	BNon Horoscopy
Girls	3/20	30/60
Boys	12/84	12/36

We would like to set up a probabilistic model for this, involving the following parameters (assume that every applicant to this University must choose just one of the two programmes listed above):

- π: The prior probability of an applicant to this University being a girl.
- p_g: The probability of a girl applicant choosing Witchcraft.
- p_b: The probability of a boy applicant choosing Witchcraft.
- q_{gw}: The probability of a girl applicant to Witchcraft being successful.
- q_{bw}: The probability of a boy applicant to Witchcraft being successful.
- q_{gh}: The probability of a girl applicant to Horoscopy being successful.
- q_{bh}: The probability of a boy applicant to Horoscopy being successful.

(a) Write down the likelihood (denote it $\mathcal{L}$) of the above data, given these parameters. Be careful and clear with your notation, and keep in mind that you need to account for *all* of the applicants included in the above statistics. [4]

(b) Use this likelihood function to obtain the maximum likelihood estimate for q_{bw}. Clearly show your working, and try to keep it as concise as possible. (Hint: Making appropriate use of the symbol $\mathcal{L}$ introduced above can greatly simplify your working.) [3]

(c) Give the maximum likelihood estimates for the other 6 parameters. (Just write down the answers, no working needs to be shown.) [3]

2. The diastolic blood pressure readings (in $mmHg$) of 5 individuals from a given population are found to be as follows: $\{89, 92, 87, 90, 93\}$.

(a) Let us assume that the underlying distribution is uniform over a limited range, i.e., we have

$$p(x|a,b) = \begin{cases} \frac{1}{b-a}, \text{if } a \leq x \leq b. \\ 0, \text{otherwise.} \end{cases}$$

Here x is an individual's diastolic blood pressure reading, and a and b are respectively the lower and upper limits of the range. Given the above data, what are the maximum likelihood estimates of a and b? (Full derivation not needed, but some justification should be provided.) [3]

---

*Please write this code on the cover page of your answer script.
†Bachelor of Nonsense.

(b) Assuming that the underlying distribution really is uniform, do you think these are good estimates of a and b? Why or why not? [1]

(c) Now let us assume a normal underlying distribution:

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Furthermore, suppose someone has told us beforehand that μ is expected to be $72mmHg$, with a standard deviation of $6mmHg$. We wish to carry out Bayesian inference, using this information as our prior. Given the above data, compute MAP estimates of μ for two different assumed values of σ: $\sigma = 1$ and $\sigma = 10$. [4]

(d) Which of these two do you think gives a better estimate of the true population mean? What is the problem with the other estimate: is it underfitting, or overfitting? [2]

3. Consider a supervised two-class classification problem in two dimensions, with the following training set.

x_1	x_2	t
4	4	-1
4	-4	-1
-4	4	-1
2	2	1
[illegible]	-2	1
-2	2	1

(a) Draw a graph depicting this training set. What will happen if we attempt to train a hard-margin linear (i.e., no kernel) SVM on this data? Explain. [1]

(b) Now suppose you can map the input feature space $\mathbf{x} = (x_1, x_2)$ to some new feature space $\phi(\mathbf{x})$. Give the simplest (i.e., lowest dimensional) mapping ϕ you can think of in order to allow a hard-margin linear SVM to be trained in the new space. [1]

(c) Depict, in your graph drawn in part (a) above, the decision boundary that will be learnt in part (b). What is the equation of this boundary (in terms of the *original features*, x_1 and x_2)? [2]

(d) Rather than explicitly applying the mapping ϕ to the data and learning a linear SVM in the transformed space, we could have achieved the same effect by using the *kernel trick* to learn a non-linear SVM in the input space itself. Write down the kernel function $K(x_1, x_2)$ corresponding to your choice of ϕ. [1]

$K(\underline{x}, \underline{x}')$