EEL709: Minor I

February 8, 2014

Maximum Marks: 20

1. **Attention Pretension and Retention Tension: How attentive were you in class? Short answers, please!**

   (a) $P(B) = \sum_{\forall A_i} P(B|A_i)P(A_i)$. What is a basic assumption here?

   (b) Consider two measures of error between the actual observed data t_i and the predicted data according to a model, $y(x_i, \mathbf{w})$, for data points x_i and model parameters $\mathbf{w}$: the sum-of-squared errors, and the RMS error. Give two basic reasons why the RMS error is intuitively a 'better' measure.

   (c) What is S-fold cross-validation, and where is it useful?

   (d) Prof. P. C. Mahalanobis is regarded as the founding father of statistics: the Indian Statistical Institute was the first teaching and research institute devoted to statistics, anywhere in the world. One of the most improtant contributions of the great man was the Mahalanobis distance. What is the physical significance of the Mahalanobis distance as the distance between two points in a distribution, and how does this score over the Euclidean distance? Given two separate distributions in k-dimensional space, how will you find the Mahalanobis distance between a point in one distribution, with one in another? *(1+2+1+2+1 marks)*

2. **Sequential Bayesian Filtering** Consider a system's state described by a feature vector at time t, $\mathbf{X}_t$. Suppose we have observations of the system's state, $\mathbf{Z}_1 \ldots \mathbf{Z}_t$ for time instants $1 \ldots t$. The shortcut notation $\mathbf{Z}_{1:t}$ denotes the set $\mathbf{Z}_1, \ldots \mathbf{Z}_t$.

   (a) Making suitable assumptions, prove the Chapman-Kolmogorov result. This is also called the 'prediction rule': given all observations from time 1 to time $t-1$, we are predicting the state of the system at time t.

   $$P(\mathbf{X}_t|\mathbf{Z}_{1:t-1}) = \sum_{\forall \mathbf{X}_{t-1}} P(\mathbf{X}_t|\mathbf{X}_{t-1})P(\mathbf{X}_{t-1}|\mathbf{Z}_{1:t-1}) \quad (1)$$

   (b) Again, make suitable assumptions to prove the 'updating rule'. (Please note that there is no word in the Queen's language called 'updation'!) Given the prediction of the last stage and the observation at time t, we wish to update our probability estimate of the state of the system at time t, given all observations from time 1 to time t.

   $$P(\mathbf{X}_t|\mathbf{Z}_{1:t}) \propto P(\mathbf{Z}_t|\mathbf{X}_t)P(\mathbf{X}_t|\mathbf{Z}_{1:t-1}) \quad (2)$$

   (2+3 marks)

3. The marks of 6 randomly chosen students from last year's offering of EEL709 are observed to be as follows: $\{59.80, 48.54, 30.06, 40.10, 52.99, 0.00\}$. Suppose we model these as coming from a *Gaussian distribution*, i.e.,

   $$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}. \quad (3)$$

Here x is a random variable denoting the marks for a given student.

(a) Given a data set $\{x_1, ..., x_N\}$, what is its likelihood under the assumption of the above distribution? Derive a general expression for the maximum likelihood estimate (MLE) of the parameter μ. Now plug in the data given above (you may round off to integers, if you wish) to obtain the MLE for these specific observations. Do you think this is likely to be a good estimate of the actual value of μ? Why or why not? *(3 marks)*

(b) Supposing I now tell you that on average, courses taught by last year's EEL709 instructor have a class average of α marks, with a standard deviation of β marks. Suggest a Bayesian approach to incorporate this piece of information into your estimation of μ, and use it to obtain a general expression for your modified estimator of μ (in terms of a generic data set $\{x_1, ..., x_N\}$). Now, given that $\alpha = 50$ and $\beta = 5$, apply your modified estimator to the above EEL709 data. Compare the estimate obtained with the MLE computed above, and comment on the difference. *(5 marks)*

(You may make any assumptions you feel are required; but please be clear about your notation, and show all working used to obtain your estimates.)