ELL784/EEL709: Major Test

May 5, 2016

Maximum Marks: 24

Note: Please follow a consistent notation to distinguish between vectors and scalars. The suggested notation is to use an underbar for vectors; if you wish to use something else, please specify it at the start of your answer script.

1. Consider data $\{x_1, x_2, ..., x_N\}$ drawn from a univariate Gaussian $\mathcal{N}(\mu, \sigma^2)$. Derive the maximum likelihood estimator for σ. Prove that this is a biased estimator. Hence find the correction factor to be applied to the estimator to make it unbiased. (Full working must be shown for all steps.) **[3]**

2. Suppose you run PCA on a high-dimensional labeled data set, and find that only a small number of principal components capture most of the variance (say $> 90\%$). Is it always safe to replace the original features with the smaller number of principal components, for the purposes of supervised learning? If yes, give a general justification. If no, explain using a simple counter-example. **[2]**

3. Bioinformatics is a major application area of machine learning. One common type of biological data is DNA sequences, which are strings over an alphabet of 4 symbols, usually denoted $\{A, C, T, G\}$. Suppose I want to define a kernel function over such data for use in an SVM. I propose a kernel K that counts the number of position-wise matches between two DNA sequences. For instance, $K(ACTGG, ATCG) = 2$, and $K(AACTCG, ACCTGGA) = 4$. Prove whether or not K is a valid kernel. **[3]**

4. In class, we looked at the logistic sigmoid as the standard activation function used in neural networks. However, these days many other activation functions are becoming popular, especially in the context of deep learning approaches. The most common kind of hidden unit used in these approaches is known as a *ReLU* (rectified linear unit), which is a neuron with the activation function $f(a) = max(0, a)$. Since this is not differentiable, a smooth approximation called *softplus* may also be used: $f(a) = ln(1 + e^a)$.

   Suppose you have a neural network with a single hidden layer, which uses the softplus activation function. At the output layer you have a single logistic regression unit. Give the modified version of the backpropagation algorithm for this case, by deriving a general expression for the error gradient with respect to any of the weights in the first layer. All steps in your working should be shown, and all notation should be clearly defined. **[5]**

5. Consider a Naïve Bayes model for spam detection, which has been trained on e-mails that contain only three words, $\{x_1, x_2, x_3\}$. The parameters learnt by the model are as follows:

Word	Spam class	Non-spam class
x_1	$\mu_{11} = 0.5$	$\mu_{10} = 0.67$
x_2	$\mu_{21} = 0.67$	$\mu_{20} = 0.33$
x_3	$\mu_{31} = 0.33$	$\mu_{30} = 0.33$

   Now we want to classify an e-mail which contains words x_1 and x_2 but not x_3.

   (a) Compute the *likelihood* of this e-mail under both classes, spam and non-spam. Hence report the maximum likelihood (ML) classification of the e-mail. **[1]**

   (b) For a full Naïve Bayes classification, we also need to consider the class prior. Let π be the *a priori* probability of an e-mail being spam. For the given e-mail, find the range of values of π for which the MAP classification from Naïve Bayes will match the ML classification from part (a). **[2]**

6. Over three successive financial quarters, the direction of movement (either *increase* or *decrease*) in a company's average stock price for that quarter (relative to the preceding quarter) is observed. Suppose I want to model the dependence of this direction of movement on whether the company made a profit in that quarter or not (which I cannot observe). Let's say I assume the following: If the company makes a profit, the probability of the average stock price increasing is 0.8; and if the company makes a loss, the probability of the average stock price increasing is 0.2. Further, suppose that a profitable quarter will be followed by another profitable quarter with probability 0.6; and a loss-making quarter will be followed by another loss-making quarter with probability 0.3. Finally, I assume that the *a priori* probability of any quarter being profitable is 0.7.

   (a) Draw an appropriate Hidden Markov Model to represent this situation. Specify clearly your notation for random variables, and the corresponding initialisation, emission, and transition probabilities. **[2]**

   (b) Suppose the actual movements over the first three quarters of 2015 were observed to be $\{increase, decrease, increase\}$. Based on this and the above model, I wish to estimate the probability that the company made a profit in third quarter. Use the forward-backward algorithm to compute this. Referring to the notation used in class, which α or β value(s) need to be computed for this purpose? Show the steps of the recursion involved in doing so. **[4]**

   (c) Use your model to predict the probability of the company's average stock price increasing in the fourth quarter of 2015, relative to the third quarter. **[2]**