ELL784/EEL709: Introduction to Machine Learning

Minor Test II

22 March 2016

Maximum marks: 20
Closed book/notes; one two-sided A4 cheat sheet permitted

Please show all steps in your working fully and clearly, except where indicated otherwise. In this question paper, bold symbols indicate vectors. In your answers, please follow a consistent notation to denote vectors (e.g., an underbar).

1. Consider the following two-class data set:

x_1	x_2	t
1	1	-1
0	0	-1
2	3	1
1	2	1

For your reference, the general form of the (vanilla/hard-margin) SVM optimisation problem is as follows:

$$min \ \frac{1}{2}||\mathbf{w}||^2$$
$$s.t.$$
$$(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + w_0)t_i \geq 1 \ \ \forall i \in \{1, 2, ..., N\}.$$

(a) Visualise the given data set on a 2-D plot (you may use two different symbols of your choice for the two classes). Just by visual inspection, can you tell what will be the decision boundary learnt by an SVM on this data set? Draw the decision boundary on your plot and write down its equation. **[2]**

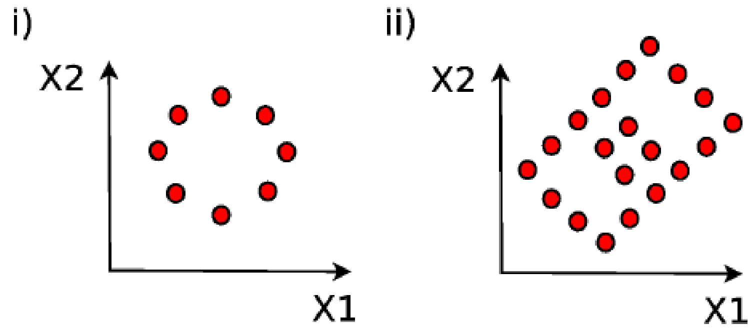
(b) Write down the primal form of the Lagrangian corresponding to the above optimisation problem for *this specific data set*. Do not write the general form, do not use summation signs. (Hint: the expression is long but simple. Please be careful and do not leave out any terms.) **[2]**

(c) Compute the partial derivatives of your Lagrangian with respect to the parameters $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ and w_0. By setting these to 0, eliminate $\mathbf{w}$ and w_0 to obtain the *dual* form of the Langrangian, which will be a function only of the Lagrange multipliers $\boldsymbol{\mu}$. **[3]**

(d) Now look at your visualisation from part (a). Does this tell you which Lagrange multipliers must be 0 for the optimal solution? Plug in these 0 values to simplify your dual Lagrangian. Now, compute partial derivatives of the simplified dual with respect to the other (non-zero) Lagrange multipliers. Set the partial derivatives to 0, and thus obtain the values of these Lagrange multipliers which maximise the dual. **[3]**

(e) Plug back your obtained values from part (d) into the expressions obtained for $\mathbf{w}$ in part (c). Does the computed value of $\mathbf{w}$ match with the decision boundary you drew by visual inspection in part (a)? Why or why not? **[2]**

2. Consider the following data sets.



(a) Write down *unit vectors* corresponding to the first and second principal components for both of the above data sets. (If there are multiple possible sets of principal components, you need to consider only one of them.) **[2]**

(b) In which of the above cases will the first principal component capture a larger proportion of the variance in the data? Explain your answer. **[1]**

(c) In which of the above cases will the first two principal components (combined) capture a larger proportion of the variance? Explain your answer. **[1]**

3. (a) Give an example of a data set where K-means will not work well, even when there is a natural clustering and K has been appropriately specified. Draw a plot showing the data and specifying both what result you think K-means will give on it and what the appropriate clustering ought to be. **[2]**

(b) Why do you think K-means will fail on your data set? Is it a convergence problem: that it might fail to converge or converge to a local minimum? Or is it a problem with the definition of the objective function? In either case, explain precisely what the problem is. Can you suggest some rough idea for a different clustering approach that might work for your example? **[2]**