ELL 409 Minor 2

Profs. Jayadeva and Prathosh

December 27, 2020

1. (a) Establish the Equivalence between the MAP estimate with a Laplacian prior and L1 regularization. PDF of laplacian distribution is given by $p(x) = \frac{1}{2b} exp(-\frac{|x-\mu|}{b})$. (2)
   (b) Show that perturbing each input feature with an uncorrelated additive noise is equivalent to L2-regularization (You may use 1D features and MSE as Loss for simplicity). (2)
   (c) Can Bias and Variance both be reduce simultaneously? If so, state 2 such ways. (1)

2. Figure 1 shows a set of samples with the decision boundary of a SVM. The sketch is not to scale and is intended for illustration only. The SVM was solved using a linear kernel with C = 10. "X" denotes class 1. In this problem, the sample at (0,0) is misclassified.
   (a) What is the value of the Lagrange multiplier associated with the sample at (0,0) ?
   (b) What is the value of the Lagrange multiplier associated with the sample at (3,3) ?
   (c). Assume that the samples at (0,1), (1,0) and (1,1) have equal Lagrange multipliers. Using this assumption, determine their values without solving any optimization problem. Explain your steps. Note that this need not be the solution of the SVM optimization problem.

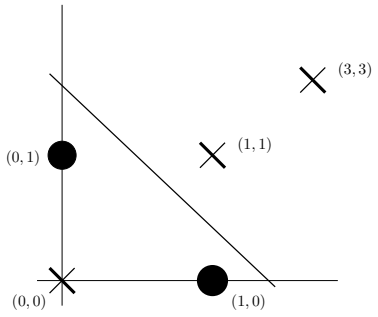

Figure 1: Training samples with the decision boundary

(1 + 1 + 2 marks)

3. A SVM is used to solve the binary classification problem with $x^i = (0,0)^T, (0,1)^T, (1,0)^T, (1,1)^T$, and the corresponding y_i = (-1, 1, 1, -1).
   A nonlinear map $\phi(x_1, x_2) = (x_1, x_2, (2x_1 - 1)(2x_2 - 1))$ is used to find a separating hyperplane in a higher dimensional space.
   (a) Determine the kernel function $K(p, q)$ in terms of the elements of two generic two dimensional vectors $p = (p_1, p_2)^T$ and $q = (q_1, q_2)^T$.
   (b) Determine the Kernel matrix for the training dataset.
   (c) A student attempting this question assumes that all Lagrange multipliers are equal to 1. Does this assumption satisfy all KKT conditions ?
   (d) For a general SVM using a kernel K, determine the margin in the image (kernel) space in terms of K and the

   (2 + 1 + 1 + 2 marks)

4. (a) Write out the Bias and variance terms of when MSE is used as as loss function and interpret them both (1).

   (b) What is regularization? State at least 3 different methods of accomplishing it (1.5).

   (c) Write down the Gradient update equation for L2-regularized logistic regression with cross entropy as the cost function in a binary classification problem (1.5).

   (d) Is an SVM inherently regularized? If so how and what is the loss functiona and the regularizer (1).