CSE/ISYE 6740 Homework 4

George Lan

October 26, 2019

- Submit your answers as an electronic copy on T-square.
- No unapproved extension of deadline is allowed. Zero credit will be assigned for late submissions. Email request for late submission may not be replied.
- For typed answers with LaTeX (recommended) or word processors, extra credits will be given. If you handwrite, try to be clear as much as possible. No credit may be given to unreadable handwriting.
- Explicitly mention your collaborators if any.

1 SVM. [20 points]

1. Explain why can we set the margin $c = 1$ in SVM formulation?
2. Using Lagrangian dual, show that the the weight vector can be represented as
$$w = \sum_{i=1}^{m} \alpha_i y^i x^i.$$
where α_i are the dual variables. What does this imply in terms of how to relate data to w?
3. Explain why only the data points on the "margin" will contribute to the sum above, i.e., playing a role in defining w. Hint: use the Lagrangian multiplier derivation and KKT condition we discussed in class.

2 Neural networks [20 points].

1. Consider a neural networks for a binary classification using sigmoid function for each unit. If the network has no hidden layer, explain why the model is equivalent to logistic regression.
2. Consider a simple two-layer network in the lecture slides. Given the cost function used to training the neural networks
$$\ell(w, \alpha, \beta) = \sum_{i=1}^{m} (y^i - \sigma(w^T z^i))^2$$
where $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid function. Show the that the gradient is given by
$$\frac{\partial \ell(w, \alpha, \beta)}{\partial w} = \sum_{i=1}^{m} 2(y^i - \sigma(u^i))\sigma(u^i)(1 - \sigma(u^i))z^i.$$
where $z_1^i = \sigma(\alpha^T x^i)$, $z_2^i = \sigma(\beta^T x^i)$. Also find the gradient of ℓ with respect to α and β.

3 Programming Problem: Subgradient descent and stochastic subgradient descent for SVM [60 points]

In the demo code, we used the gradient ascent method for solving the dual SVM problem. In this assignment, you are asked to:

a) Implement the subgradient descent method for solving the primal SVM problem. Make sure that your results match well with the ones obtained from the dual SVM solver.

b) Implement the stochastic subgradient descent method for solving the primal SVM problem. Compare your results (computation time and solution quality) with the subgradient descent method in part a).

(Note: for those who are not familiar with Matlab, you are encouraged to group with other students who know Matlab well. For those who prefer to using Python, extra credits will be given since more work needs to be done in Python for this assignment.)