University of Illinois Spring 2018

CS 446: Machine Learning
Homework 3: Binary Classification

Due on Tuesday, Feb 06, 2018, 11:59 a.m. Central Time

1. **[15 points]** Binary Classifiers

   (a) In order to use a linear regression model for binary classification, how do we map the regression output $\mathbf{w}^\top \mathbf{x}$ to the class labels $y \in \{-1, 1\}$?

Your answer:

   (b) In logistic regression, the activation function $\mathrm{g}(a) = \frac{1}{1+e^{-a}}$ is called sigmoid. Then how do we map the sigmoid output $\mathrm{g}(\mathbf{w}^\top \mathbf{x})$ to binary class labels $y \in \{-1, 1\}$?

Your answer:

   (c) Is it possible to write the derivative of the sigmoid function g w.r.t a, i.e. $\frac{\partial g}{\partial a}$, as a simple function of itself g? If so, how?

Your answer:

   (d) Assume quadratic loss is used in the logistic regression together with the sigmoid function. Then the program becomes:

   $$\min_{\mathbf{w}} f(\mathbf{w}) := \frac{1}{2} \sum_i \Big(y_i - \mathrm{g}(\mathbf{w}^\top \mathbf{x}_i) \Big)^2$$

   where $y \in \{0, 1\}$. To solve it by gradient descent, what would be the $\mathbf{w}$ update equation?

Your answer:

(e) Assume $y \in \{-1, 1\}$. Consider the following program for logistic regression:

$$\min_{\mathbf{w}} f(\mathbf{w}) := \sum_i \log\left(1 + \exp(-y^{(i)}\mathbf{w}^T\phi(x^{(i)}))\right).$$

The above program for binary classification makes an assumption on the samples/data points. What is the assumption?

Your answer: