

Machine Learning  
V. Adamchik      CSCI 567      Spring 2019  
Discussion 4      University of Southern California

Gradient Descent

Perceptron

Logistic Regression

## Problem 1

---

Why is the Hessian of logistic loss positive semidefinite?

## Problem 2

---

Can we apply Newton's method to the perceptron loss to minimize classification error?

$$F(\mathbf{w}) = \sum_{n=1}^N \max(0, -y_n \mathbf{w}^T x_n)$$

## Problem 3

---

Which of the following surrogate losses is not an upper bound of the 0-1 loss?

- (A) perceptron loss  $\max\{0, -z\}$
- (B) hinge loss  $\max\{0, 1-z\}$
- (C) logistic loss  $\log(1 + \exp(-z))$
- (D) exponential loss  $\exp(-z)$

## Problem 4

---

The following table shows a binary classification training set and the number of times each point is misclassified during a run of the perceptron algorithm. Which of the following is the final output of the algorithm? Assume  $w^{(0)} = 0$ .

$x$	$y$	Times misclassified
$(-3, 2)$	+1	5
$(-1, 1)$	-1	5
$(5, 2)$	+1	3
$(2, 2)$	-1	4
$(1, -2)$	+1	3

## Problem 5

---

Suppose we obtain a hyperplane  $w$  via logistic regression and are going to make a randomized prediction on the label  $y$  of a new point  $x$  based on the sigmoid model. What is the probability of predicting  $y = +1$ ?

(a)  $e^{-w^T x}$

(b)  $\frac{1}{1 + e^{-w^T x}}$

(c)  $\frac{1}{1 + e^{w^T x}}$

(d)  $\mathbb{I}[w^T x \geq 0]$

## Problem 6

---

Assume we have a training set  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)$ , the probability of seeing outcome  $y$  is given by

$$P(y|\mathbf{x}_n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\mathbf{y} - \mathbf{w}^T \mathbf{x}_n)^2}{2\sigma^2}\right)$$

Find the maximum likelihood estimations for  $\mathbf{w}$  and  $\sigma$