Machine Learning

V. Adamchik | CSCI 567 | Fall 2019

Discussion Set 4 | University of Southern California

Gradient Descent

Perceptron

Logistic Regression

Problem 1

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 2

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 $w^{(0)} = 0$.

$\boldsymbol{x}$	y	Times misclassified
(-3, 2)	+1	5
(-1, 1)	-1	5
(5, 2)	+1	3
(2, 2)	-1	4
(1, -2)	+1	3

Problem 3

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

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

Problem 4

Why is the Hessian of logistic loss positive semidefinite?

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?

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

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

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

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

Problem 6

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

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

Find the maximum likelihood estimations for $\boldsymbol{w}$ and σ

Problem 7

For a fixed multiclass problem, which of the following multiclass-to-binary reductions has the smallest testing time complexity?

(A) One-versus-all

(B) One-versus-one

(C) Tree reduction

(D) Both (A) and (C)

Problem 8

Show that one-versus-all can be seen as a special case of error-correcting-output-code (ECOC). Specifically, write down the code matrix M for ECOC for a problem with C labels so that executing ECOC is the same as doing one-versus-all. (Note: the entry of M should be either –1 or +1.)

Problem 9

In the lecture we derive the multiclass logistic regression by minimizing the multiclass logistic loss. In this problem you need to derive the multiclass perceptron algorithm in a similar way.

(1) Define the multiclass perceptron loss.

(2) Based on (1), write down the multiclass perceptron algorithm.