

MULTI-CLASS LOGISTIC REGRESSION

CONSIDER A MULTICLASS LOGISTIC REGRESSION PROBLEM OF THE FORM

$$(*) \left[\min_{X \in \mathbb{R}^{d \times K}} \sum_{i=1}^m \left[-x_{b_i}^T a_i + \log \left(\sum_{c=1}^K \exp(x_c^T a_i) \right) \right] \right]$$

LIKELIHOOD FOR SINGLE TRAINING EXAMPLE i WITH FEATURES $a_i \in \mathbb{R}^d$ AND LABEL $b_i \in \{1, 2, \dots, K\}$ IS GIVEN BY

$$P(b_i | a_i, X) = \frac{\exp(x_{b_i}^T a_i)}{\sum_{c=1}^K \exp(x_c^T a_i)}$$

PARAMETERS
↓
 $d \times K$

WHERE x_c IS COLUMN c OF MATRIX PARAMETER $X \in \mathbb{R}^{d \times K}$

TO MAXIMIZE LIKELIHOOD OVER m INDEPENDENT IDENTICALLY-DISTRIBUTED TRAINING SAMPLES, WE MINIMIZE NEGATIVE LOG-LIKELIHOOD:

$$\hat{f}(X) = \sum_{i=1}^m \left[-x_{b_i}^T a_i + \log \left(\sum_{c=1}^K \exp(x_c^T a_i) \right) \right]$$

FREE TIP

PARTIAL DERIVATIVE IS:

$$\frac{\partial f(X)}{\partial x_{jc}} = - \sum_{i=1}^m a_{ij} \left[\overset{\text{INDICATOR VARIABLE}}{\uparrow} I(b_i = c) - \frac{\exp(x_c^T a_i)}{\sum_{c'=1}^K \exp(x_{c'}^T a_i)} \right]$$

HOMEWORK (DEADLINE 21ST OF MAY)

1. RANDOMLY GENERATE A 1000×1000 MATRIX WITH ENTRIES FROM A $N(0, 1)$ DISTRIBUTION
2. GEN. $b_i \in \{1, 2, \dots, K\}$ ($K=50$) BY DRAWING

$$AX + E$$

WITH X, E SAMPLED FROM NORMAL DISTRIBUTION

$$X \in \mathbb{R}^{d \times K}$$

$$E \in \mathbb{R}^{m \times K}$$

CONSIDER MAX INDEX IN THE ROW AS CLASS LABEL!

3. SOLVE PROBLEM ~~*~~ WITH

(A) GRADIENT DESCENT

(B) BCGD WITH RANDOMIZED ROWS

(C) BCGD WITH GAUSS-SOUTHWELL RULE

USE BLOCKS $X_{j:c'}$ $c' \in \{1, \dots, K\}$

!!
[EACH ROW OF X IS ONE BLOCK!!!]

4. CHOOSE A PUBLICLY AVAILABLE DATASET
AND TEST METHODS ON THIS.

5. ANALYSE ACCURACY VS CPU TIME

6. DESCRIBE WHAT YOU DID ON A PDF
FILE

7. SUBMIT PROJECT