TRAINING SET = $\{(a_i, b_i)\}$ $a_i \in \mathbb{R}^d$, $b_i \in \{1, 2, \dots, k\}$, $i = 1 \dots m$

MULTI-CLASS LOGISTIC REGRESSION

CONSIDER A MULTICLASS LOGISTIC REGRESSION PROBLEM OF THE FORM

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

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

$$p(b_i \mid a_i, X) = \frac{\exp(x_{b_i}^T a_i)}{\sum_{c=1}^{k} \exp(x_c^T a_i)}$$

PARAMETERS

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

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:

INDICATOR VARIABLE

$$\frac{\partial f(X)}{\partial X_{jc}} = -\sum_{i=1}^{m} a_{ij} \left[\underbrace{I(b_i = c)}_{\begin{cases} = 1 & \text{if } b_i = c \\ = 0 & \text{otherwise} \end{cases}} - \frac{\exp(x_c^T a_i)}{\sum_{c'=1}^{k} \exp(x_{c'}^T a_i)} \right]$$

HOMEWORK (DEADLINE 21ST OF MAY)

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

$$AX + E$$

WITH X, E SAMPLED FROM NORMAL DISTRIBUTION

$$X \in \mathbb{R}^{d \times k} \qquad E \in \mathbb{R}^{m \times k}$$

CONSIDER MAX INDEX IN THE ROW AS CLASS LABEL!

3. SOLVE PROBLEM * WITH

(A) GRADIENT DESCENT
(B) BCGD WITH RANDOMIZED RULE
(C) BCGD WITH GAUSS-SOUTHWELL RULE

USE BLOCKS $X_{jc'}$ $c' \in \{1, .., k\}$

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

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

5. ANALYZE ACCURACY VS CPU TIME

6. DESCRIBE WHAT YOU DID ON A PDF FILE

7. SUBMIT PROJECT