Sentiment Analysis: Ordinal Logistic Regression

ChengXiang "Cheng" Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

Motivation: Rating Prediction

- Input: An opinionated text document **d**
- Output: Discrete rating **r** $\in${**1, 2, …, k**}
- Using regular text categorization techniques
  - Doesn't consider the order and dependency of the categories
  - The features distinguishing r=2 from r=1 may be the same as those distinguishing r=k from r=k-1 (e.g., positive words generally suggest a higher rating)
- Solution: Add order to a classifier (e.g., ordinal logistic regression)

Logistic Regression for Binary Sentiment Classification

Binary Response Variable: Y $\in$ {0,1}

Predictors: $X = (x_1, x_2, \ldots, x_M), \; x_i \in \Re$

$$Y = \begin{cases} 1 & X \text{ is POSITIVE} \\ 0 & X \text{ is NEGATIVE} \end{cases}$$

$$\log \frac{p(Y=1|X)}{p(Y=0|X)} = \log \frac{p(Y=1|X)}{1-p(Y=1|X)} = \beta_0 + \sum_{i=1}^{M} x_i \beta_i \quad \beta_i \in \Re$$

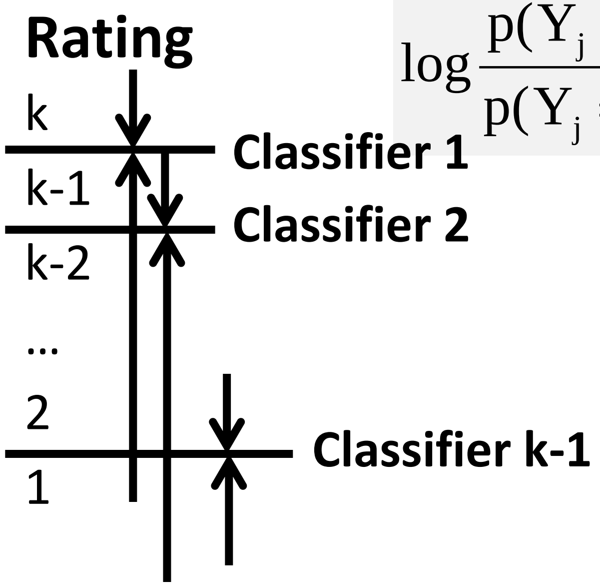
$$p(Y=1|X) = \frac{e^{\beta_0 + \sum_{i=1}^{M} x_i \beta_i}}{e^{\beta_0 + \sum_{i=1}^{M} x_i \beta_i} + 1}$$

Logistic Regression for Multi-Level Ratings

$$Y_j = \begin{cases} 1 & \text{rating is } j \text{ or above} \\ 0 & \text{rating is lower than } j \end{cases}$$

Predictors: $X = (x_1, x_2, \ldots, x_M),\ \ x_i \in \Re$

Rating: $r \in \{1, 2, \ldots, k\}$



$$\log \frac{p(Y_j = 1 \mid X)}{p(Y_j = 0 \mid X)} = \log \frac{p(r \geq j \mid X)}{1 - p(r \geq j \mid X)} = \alpha_j + \sum_{i=1}^{M} x_i \beta_{ji} \qquad \beta_{ji} \in \Re$$

$$p(r \geq j \mid X) = \frac{e^{\alpha_j + \sum_{i=1}^{M} x_i \beta_{ji}}}{e^{\alpha_j + \sum_{i=1}^{M} x_i \beta_{ji}} + 1}$$

Rating Prediction with Multiple Logistic Regression Classifiers

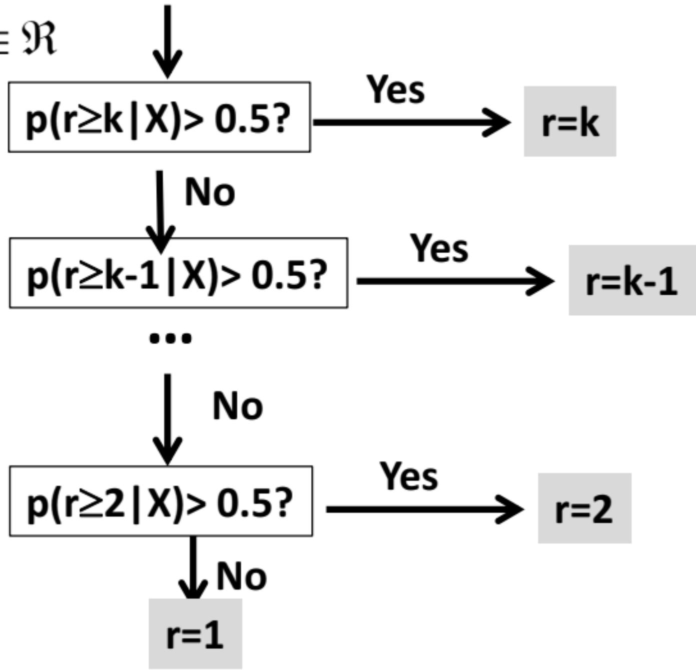
Text Object: $X = (x_1, x_2, \ldots, x_M),\ \ x_i \in \Re$

Rating: $r \in \{1, 2, \ldots, k\}$

After training k-1 Logistic Regression Classifiers

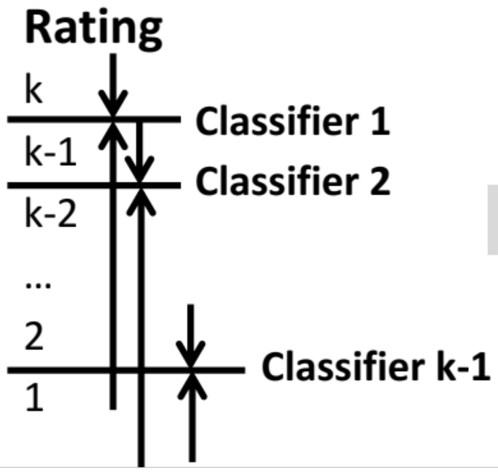
$$p(r \geq j \mid X) = \frac{e^{\alpha_j + \sum_{i=1}^{M} x_i \beta_{ji}}}{e^{\alpha_j + \sum_{i=1}^{M} x_i \beta_{ji}} + 1}$$

j=k, k-1, …, 2



Problems with k-1 Independent Classifiers?

$$\log \frac{p(Y_j = 1 \mid X)}{p(Y_j = 0 \mid X)} = \log \frac{p(r \geq j \mid X)}{1 - p(r \geq j \mid X)} = \alpha_j + \sum_{i=1}^{M} x_i \beta_{ji} \qquad \beta_{ji} \in \Re$$



$$p(r \geq j \mid X) = \frac{e^{\alpha_j + \sum_{i=1}^{M} x_i \beta_{ji}}}{e^{\alpha_j + \sum_{i=1}^{M} x_i \beta_{ji}} + 1}$$

k-1个分类器　M个β, 1个α.

How many parameters are there in total? **(k-1)*(M+1)**

The k-1 classification problems are dependent.
The positive/negative features tend to be similar!

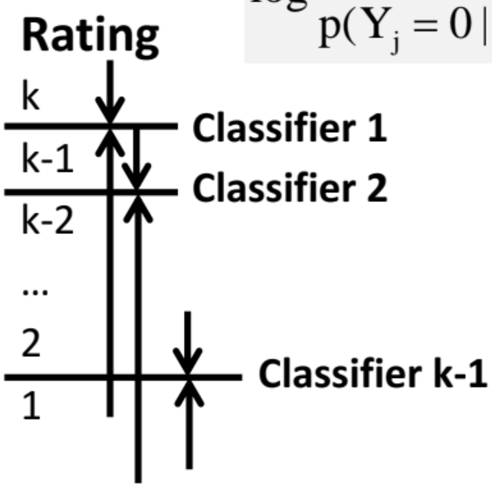
Ordinal Logistic Regression

共用β，α特殊.

Key Idea: $\forall i = 1, \ldots, M,\ \forall j = 3, \ldots, k,\ \beta_{ji} = \beta_{j-1i}$

➔ **Share training data** ➔ **Reduce # of parameters**

$$\log \frac{p(Y_j = 1 \mid X)}{p(Y_j = 0 \mid X)} = \log \frac{p(r \geq j \mid X)}{1 - p(r \geq j \mid X)} = \alpha_j + \sum_{i=1}^{M} x_i \beta_i \qquad \beta_i \in \Re$$



$$p(r \geq j \mid X) = \frac{e^{\alpha_j + \sum_{i=1}^{M} x_i \beta_i}}{e^{\alpha_j + \sum_{i=1}^{M} x_i \beta_i} + 1}$$

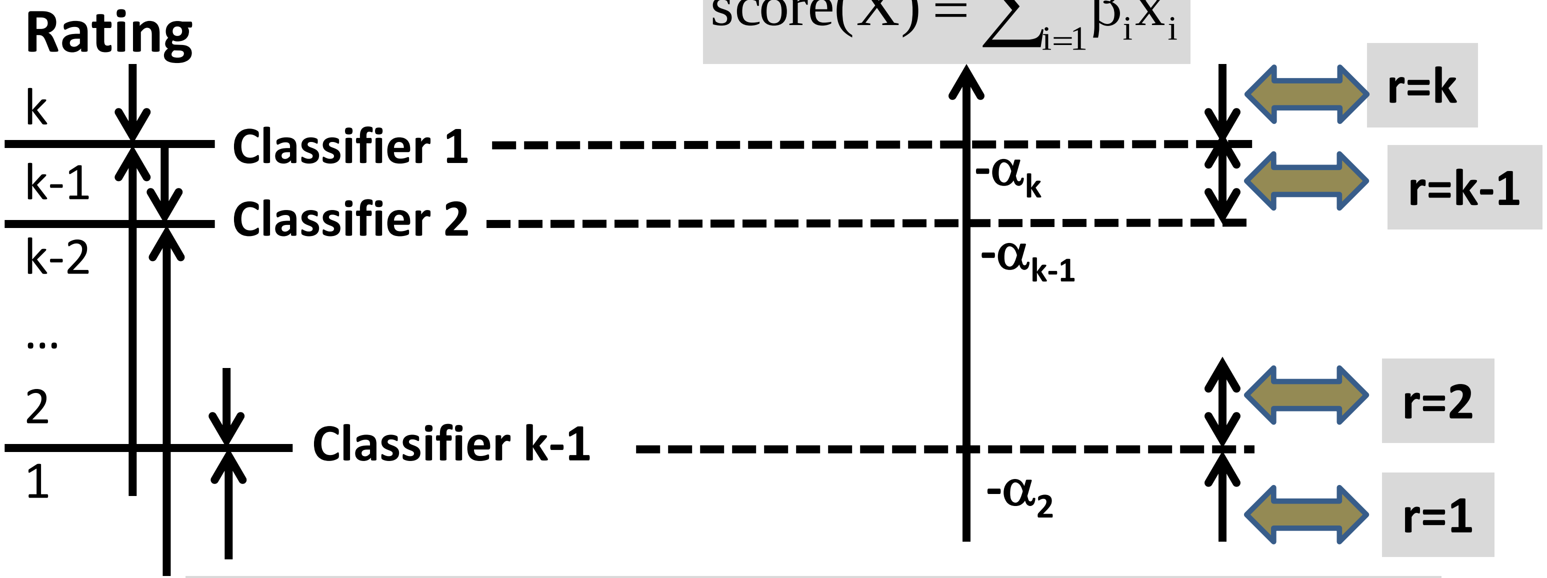
How many parameters are there in total? **M+k-1**

M个β，
k-1个α.

Ordinal Logistic Regression: Rating Prediction

$$p(r \geq j \mid X) \geq 0.5 \Leftrightarrow \frac{e^{\alpha_j + \mathrm{score}(X)}}{e^{\alpha_j + \mathrm{score}(X)} + 1} \geq 0.5 \Leftrightarrow \mathrm{score}(X) \geq -\alpha_j$$

$$\mathrm{score}(X) = \sum_{i=1}^{M} \beta_i x_i$$



$r=j \Leftrightarrow$ score $\in [-\alpha_j, -\alpha_{j+1})$, define $\alpha_1 = \infty$, $\alpha_{k+1} = -\infty$