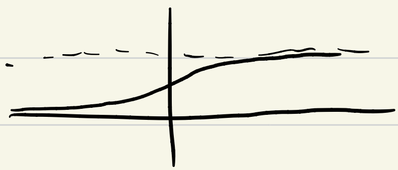
CS 371N Lecture 4: Multiclass

Announcements

- A1 due in one week
- Social impact response on Tues, open for 1 week

Recap Logistic Regression

$$P(y=+1|\bar{x}) = \frac{e^{\bar{w}^T f(\bar{x})}}{1+e^{\bar{w}^T f(\bar{x})}}$$



Minimize negative log likelihood

$$\underset{\bar{w}}{\operatorname{argmin}} \sum_{i=1}^{D} -\log P(y=y^{(i)}|\bar{x}^{(i)})$$

Update from SGD

$$\bar{w} \leftarrow \bar{w} + \alpha f(\bar{x}^{(i)}) (1-P(y=+1|\bar{x}^{(i)})) \quad \text{if } y^{(i)}=+1$$

$$\bar{w} \leftarrow \bar{w} - \alpha f(\bar{x}^{(i)}) (1-P(y=-1|\bar{x}^{(i)})) \quad \text{if } y^{(i)}=-1$$

Counter and Indexer

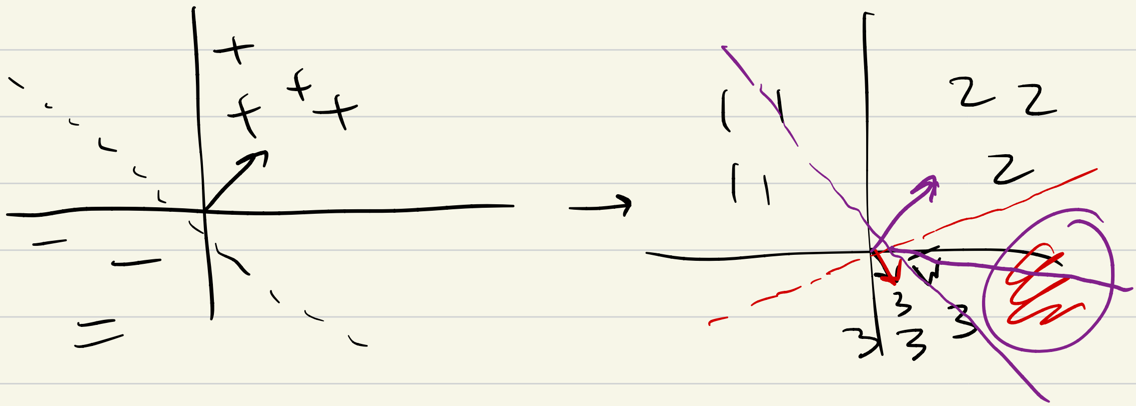
Indexer: $\left\{ \begin{array}{l} 0 \leftrightarrow \text{"the"} \\ \underline{1} \leftrightarrow \text{"of"} \\ \quad \vdots \\ 113 \leftrightarrow \text{"of the"} \end{array} \right\}$

Counter: mapping from obj to float

~10 feats on an example $\left\{ \begin{array}{l} 113 \rightarrow \underline{1} \\ \quad \vdots \end{array} \right\}$ sentence contains "of the" (1 time)

Today
- Sentiment analysis
- Multiclass: examples, perc, LR

Multiclass basics



Output space: $\mathcal{Y} = \{1, 2, 3\}$

one-vs-all: 1 vs (2, 3)

2 vs (1, 3)

3 vs (1, 2)

Reconcile by taking highest prob

MC: like one-vs-all, but trained "better"

Two ways of thinking about it:

1. Different weights per class
2. Different features per class

Different weights

$\bar{w}_1 \quad \bar{w}_2 \quad \bar{w}_3$ weight vecs per class in $\mathcal{Y}$

pred label $y_{pred} = \underset{y \in \mathcal{Y}}{\operatorname{argmax}} \underbrace{\bar{w}_y^T f(\bar{x})}_{\text{highest score}}$

Ex Headline classification

$\bar{x}$ = "too many drug trials, too few patients"

$$f(\bar{x}) = \begin{array}{ccc} [\,1 & 1 & 0\,] \\ \text{drug} & \text{patients} & \text{baseball} \end{array}$$

$$y = \{\underset{1}{\text{health}}, \underset{2}{\text{sports}}, \underset{3}{\text{science}}\}$$

"the word drug w/ the label health"

$\bar{w}_1 = [2, 5.6, -3]$

$\bar{w}_2 = [1.2, -3.1, 5.7]$

$\bar{w}_3 = [1, 1.2, -0.5]$

$\cdot\,[1\ 1\ 0]$

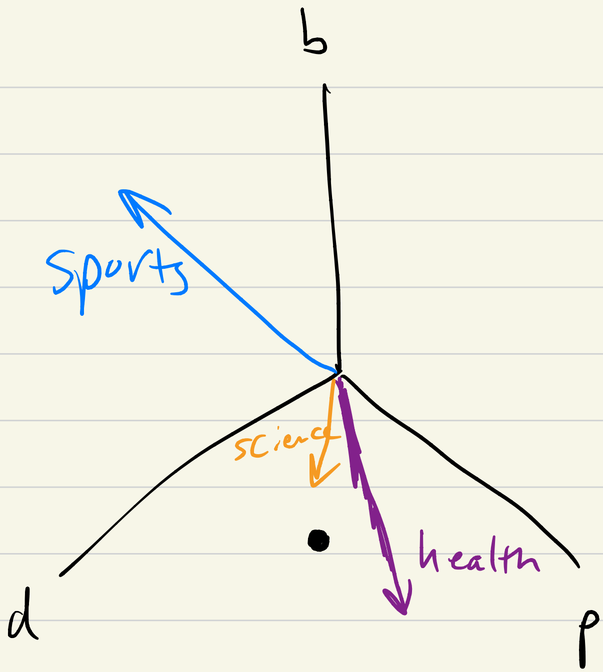
Score for class 1 (health): $\boxed{7.6}$ $(2+5.6)$

2 (sports): -1.9

argmax

3 : 2.2

$\boxed{\text{return health}}$



Multiclass Perceptron

for t in epochs

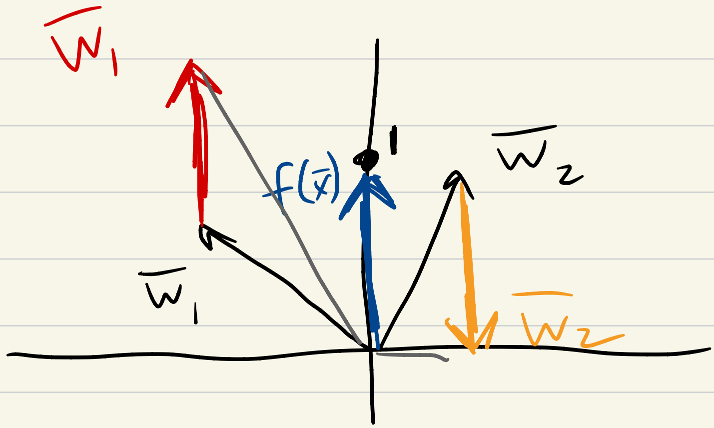
for i in data $\quad (\bar{x}^{(i)}, y^{(i)})$

$$y_{pred} = \underset{y \in \mathcal{Y}}{\operatorname{argmax}} \ \bar{w}_y^T f(\bar{x}^{(i)})$$

if $y_{pred} \neq y^{(i)}$:

$$\bar{w}_{y^{(i)}} \leftarrow \bar{w}_{y^{(i)}} + \alpha f(\bar{x}^{(i)})$$

$$\bar{w}_{y_{pred}} \leftarrow \bar{w}_{y_{pred}} - \alpha f(\bar{x}^{(i)})$$



$y^{(i)} = 1 \quad \bar{w}_2 \cdot f(\bar{x}) > \bar{w}_1 \cdot f(\bar{x}) \Rightarrow$

$y_{pred} = 2$

Update $\bar{w}_1$ $(y^{(i)})$

Update $\bar{w}_2$ (y_{pred})

Multiclass LR

$$P(y = \hat{y} \mid \bar{x}) = \frac{e^{\bar{w}_{\hat{y}}^T f(\bar{x})}}{\sum_{y' \in \mathcal{Y}} e^{\bar{w}_{y'}^T \cdot f(\bar{x})}}$$

distribution over $y \in \mathcal{Y}$

$$P(y = \text{class } 1 \mid \bar{x}) = \frac{e^{\bar{w}_1^T f(\bar{x})}}{e^{\bar{w}_1^T f(\bar{x})} + e^{\bar{w}_2^T f(\bar{x})} + e^{\bar{w}_3^T f}}$$

<u>Softmax</u>: probs are ≥ 0

probs sum to 1 over classes

Same as binary LR if we assume negative class has score 0

Update: SGD of negative log likelihood

For $y^{(i)}$:

$$\bar{w}_{y^{(i)}} \leftarrow \bar{w}_{y^{(i)}} + \alpha f(\bar{x}^{(i)}) \left(1 - P(y = y^{(i)} \mid \bar{x})\right)$$

For all other y':

$$\bar{w}_{y'} \leftarrow \bar{w}_{y'} - \alpha f(\bar{x}^{(i)}) \left(P(y = y' \mid \bar{x})\right)$$

Classes 1, 2, 3 w/probs

$[\underline{0.1}, 0.8, 0.1]$ $y^{(i)} = 1$

no ypred here (doesn't matter)

$\bar{w}_1$: This is $y^{(i)}$

add $\alpha f(\bar{x}^{(i)}) \cdot 0.9$

$\bar{w}_2$: This is not $y^{(i)}$

subtract $\alpha \cdot f(\bar{x}^{(i)}) \cdot 0.8$

$\bar{w}_3$: subtract $\alpha \cdot f(\bar{x}^{(i)}) \cdot 0.1$

Suppose we have $[0.99, 0.005, 0.005]$

Ex 2 (Instapoll)

[1 1 0] feats

$y^{(i)} = 1$ classes 1 2 3

Initialize

$\bar{w}_1, \bar{w}_2, \bar{w}_3$ all zeroes

① Predicted Class probs $[\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}]$

② LR update $(\alpha = 1)$

Class 1: $[\frac{2}{3}\ \frac{2}{3}\ 0]$ add $(1 - \frac{1}{3}) \cdot f(\bar{x})$

Class 2, 3: $[-\frac{1}{3}\ -\frac{1}{3}\ 0]$ sub $\frac{1}{3} \cdot f(\bar{x})$

③ What happens if we train on this repeatedly?