CS 378 Lecture 5

Announcements

- Pronouns
- A1 due Tues, A2 out Tues
- Pytorch in class Tuesday: install it
  if you can

Today

- Multiclass Logistic Regression
- Multiclass examples
- Feedforward neural nets

Recap "Different weights:" "Different feats"

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

classes ↗

d

$f(\bar{x}) = [$

$\bar{w}_{health} =$

$\bar{w}_{sports} =$

$\bar{w}_{science}$

Multiclass

$\bar{w}_{ypred}$

$\bar{w}_{y^{(i)}}$

Multicla

$P(y = \hat{y} |$

rug patients baseball

1 1 0]

$[2, 5.6, -3]$

$[1.2, -3.1, 5.7]$

$= [0, -2, -1.7]$

perceptron:

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

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

ss LR

$\boxed{DW}$

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

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

"score" of $y=-1 = 0$

"score" of $y=+1 = \bar{w}^T f(\bar{x}$

$$\text{binary} \left[\frac{e^{\text{score } y=+1}}{e^{\text{score of } y=-1} + e^{\text{score of } y=+}} \right.$$

$$P(y=\hat{y}|\bar{x}) = \frac{e^{\bar{w}^T f(\bar{x},\hat{y})}}{\sum_{y'} e^{\bar{w}^T f(\bar{x},y')}}$$

Update (negative gradient)

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

$\bar{w}_{y_{other}} \leftarrow \bar{w}_{y_{other}} - \alpha f(\bar{x}) P(y=y_{ot}$

$^{T}f(\bar{x})$

$^{T}f(\bar{x})$

$)$

1

$\boxed{\text{DF}}$

Y_{other}: all $y \neq y^{(i)}$

$\bar{x}))$

ner $|\bar{x})$

Poll 1:

① $P(y|\bar{x}) = [1/3, 1/3, 1/3]$

② $\bar{w}_1 = [2/3, 2/3, 0]$

$\bar{w}_2 = [-1/3, -1/3, 0]$

$\bar{w}_3 = [-1/3, -1/3, 0]$

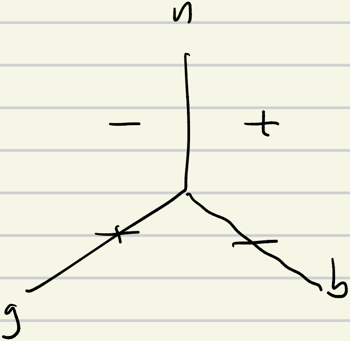
③ Weights keep slowly getting more "extreme" ($\bar{w}_1$ gets more +, others more negative)

Poll 2:

① All 4 ② MC Perc / MC LR
③ LR ④ Perc

Neural Nets

	$f(\bar{x})=[g$	b	$n]$
good	1	0	0
bad	0	1	0
not good	1	0	1
not bad	0	1	1



$\bar{w}_{\hat{y}}^T \underbrace{f(\bar{x})}$ for MC LR or perc

Replace $f(\bar{x})$ w/ nonlinear fcn

① $\underbrace{\bar{z}}_{\substack{\text{intermediate}\\ \text{feats}\\ \text{d-len}\\ \text{vector}}} = \underbrace{g}_{\text{nonlinearity}}(\underbrace{V}_{\substack{d\times n\\ \text{matrix}}}\underbrace{f(\bar{x})}_{\text{feats } n})$

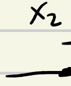
② $\underset{\hat{y}}{\text{argmax}}\ \bar{w}_{\hat{y}}^T \bar{z}$

nonlinear

$g = \tanh$

$g = \text{ReLU}$

Ex

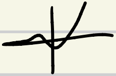


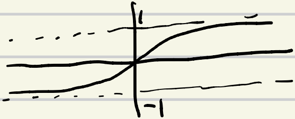
$V = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

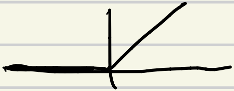
$d = \overset{3\times 2}{\overset{\shortparallel}{n}}$

$V f(\bar{x}) =$

fcn g:

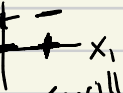






max(0, z)

x₁ XOR x₂ → x_1 XOR x_2



← will end up learning this

$\left.\begin{matrix}0\\1\\1\end{matrix}\right]$

$f(\bar{x}) = [x_1, x_2]$

$g = \tanh$

approx

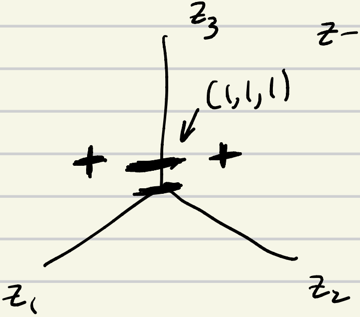
$\tanh(0) = 0$ $\tanh(1) = 1$

$\tanh(2) = 1$

$[x_1, x_2, x_1 + x_2]$

$g(Vf(\bar{x}))$

$\bar{Z} = [\tanh(x_1), \tanh(x_2),$



$\tanh(x_1 + x_2)]$

e