CS 378 Lecture 3

Classification 2: Perceptron (cont'd), Logistic Regression

Announcements
- Pronouns
- AI
- TX Votes

<u>Recap</u> Binary classification: $\bar{w}^T f(\bar{x}) \overset{?}{>} 0$

Step 1: Define feature extractor f

Step 2: Learning. Labeled data
$(\bar{x}^{(i)}, y^{(i)})_{i=1}^{D} \Rightarrow$ weights $\bar{w}$ that minimize some loss

SGD:
for t up to epochs
  for i up to D
    sample $j \sim \{1, \dots, D\}$
    $\bar{w} \leftarrow \bar{w} - \alpha \frac{\partial}{\partial \bar{w}} \text{loss}(\bar{x}^{(i)}, y^{(i)}, \bar{w})$

Perceptron

Initialize $\overline{w} = \overline{0}$

for t up to epochs

  for i up to D:

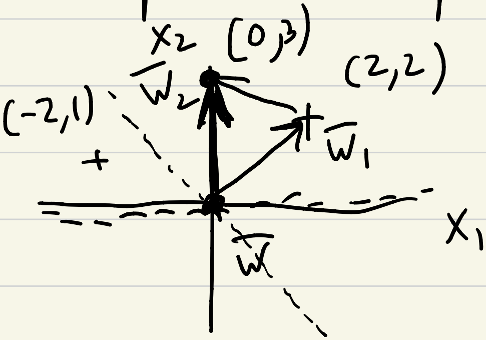
    [randomly sample example or just use i]

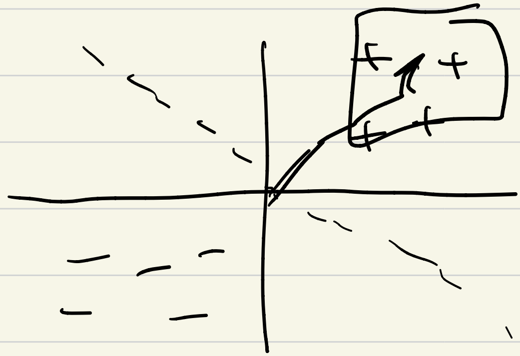
    $y_{pred} \leftarrow +1$ if $\overline{w}^T f(\overline{x}^{(i)}) > 0$

      -1 otherwise

$$\overline{w} \leftarrow \begin{cases} \overline{w} & \text{if } y_{pred} = y^{(i)} \\ \overline{w} + \alpha f(\overline{x}^{(i)}) & \text{if } y^{(i)} = +1, y_{pred} = -1 \\ \overline{w} - \alpha f(\overline{x}^{(i)}) & \text{if } y^{(i)} = -1, y_{pred} = +1 \end{cases}$$

Stop if no updates for a whole epoch





$\boxed{\text{rule: } \bar{w}^T f(\bar{x}) \overset{?}{>} 0}$

Ex

	y	g	b	n	
good	+1	1	0	0	1
not good	−1	1	0	1	2
bad	−1	0	1	0	

(feats: g, b, n)

$\bar{w}$ set $\alpha = 1$

0 0 0

$f(\bar{x})$

good: 1 0 0 $y_{pred} = -1 \neq y^{(i)}$

$\bar{w}$ 0 0 0 $\Rightarrow$ $\boxed{1\ 0\ 0}$

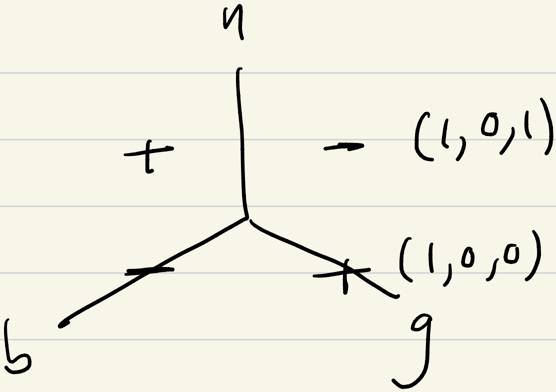
not good: $\boxed{1\ 0\ 1}$ $y_{pred} = +1 \neq y^{(i)}$

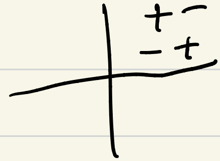
$\bar{w}$ 1 0 0 $\Rightarrow$ 0 0 −1

bad: $y_{pred} = -1$ ✓ $\bar{w} = 0\ 0\ -1$

epoch 2:

good: $y_{pred} = -1$ ✗ $\bar{w} \rightarrow 1\ 0\ -1$





	$f(\bar{x})$					$\bar{w}$ after each ex		
	g	b	n	ng	nb			
good	1	0	0			1	0	0
bad	0	1	0			1	0	0
not good	1	0	1			0	0	-1
not bad	0	1	1			0	1	0

Ways to fix it: bigrams ✓

model { kernel methods
neural nets

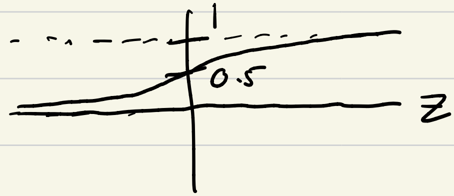
Logistic Regression

Discriminative probabilistic model

$P(y|\bar{x})$ generative: $P(\bar{x}, y)$

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

$\frac{e^z}{1+e^z}$ logistic function



maps scores $\Rightarrow$ probs

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

$$P(y=+1|\bar{x}) + P(y=-1|\bar{x}) = 1$$

Decision boundary: predict +1 if $P(y=+1|\bar{x}) > 0.5$

$$\Longleftrightarrow \bar{w}^T f(\bar{x}) > 0$$

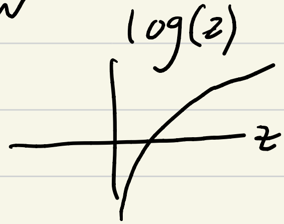
Learning

For dataset $(\bar{x}^{(i)}, y^{(i)})_{i=1}^{D}$, maximize likelihood of data

logistic regression w/ weights $\bar{w}$

$$\max_{\bar{w}} \prod_{i=1}^{D} P(y^{(i)} | \bar{x}^{(i)})$$

max this with respect to $\bar{w}$



$$\Rightarrow \max_{\bar{w}} \log \prod_{i=1}^{D} P(\text{—})$$

$$= \max_{\bar{w}} \sum_{i=1}^{D} \log P(y^{(i)} | \bar{x}^{(i)})$$

$$= \min_{\bar{w}} \sum_{i=1}^{D} \underbrace{-\log P(y^{(i)} | \bar{x}^{(i)})}_{\text{loss}(\bar{x}^{(i)}, y^{(i)}, \bar{w})}$$

$$\frac{\partial}{\partial \bar{w}} \text{loss}\left(\bar{x}^{(i)}, y^{(i)}, \bar{w}\right)$$

$\left\{ \text{Assume } y^{(i)} = +1 \right.$

$$= \frac{\partial}{\partial \bar{w}} - \log \frac{e^{\bar{w}^T f(\bar{x})}}{1 + e^{\bar{w}^T f(\bar{x})}}$$

$$= \frac{\partial}{\partial \bar{w}} \left[- \bar{w}^T f(\bar{x}) + \log\left(1 + e^{\bar{w}^T f(\bar{x})}\right) \right]$$

[Calculus]

Logistic regression update

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

if $y^{(i)} = +1$

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

if $y^{(i)} = -1$