#### Logistic Regression

### https://powcoder.com

Logisite regression defines the conditional probability directly and is a discriminative modernther traject generalize in the probability directly and

- Start with the scoring function  $\theta \cdot f(x, y)$  that measures the completification the complete point of the co
- To make sure it's not negative, we exponentiate it and get  $\exp \theta \cdot f(x, \frac{https://powcoder.com}{})$
- We then normalize it by dividing it over all possible labels  $y \in \mathcal{Y}$  and  $\operatorname{Atch} \operatorname{div} \operatorname{div} \operatorname{div} \operatorname{atch} \operatorname{powcoder}$

$$p(y|\mathbf{x}; \boldsymbol{\theta}) = \frac{\exp \boldsymbol{\theta} \cdot \boldsymbol{f}(\mathbf{x}, y)}{\sum_{y' \in \mathcal{Y}} \exp \boldsymbol{\theta} \cdot \boldsymbol{f}(\mathbf{x}, y')}$$

#### Logistic Regression

The weights are estimated by maximum conditional likelihood. Give a data set  $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$ , the maximum conditional likelihood is: Assignment Project Exam Help

$$\log p(y^{(1:N)} s \text{ signature of the property of the property$$

$$\begin{array}{c} \text{https:} \underset{=}{\overset{N}{\nearrow}} p \underset{x}{\text{powcoder}}, \text{com} \\ \sum_{i=1}^{n} \exp \theta \cdot \textbf{\textit{f}}(\textbf{\textit{x}}^{(i)}, \textbf{\textit{y}}') \\ \text{Or they can be estimated by minimizing the logistic regression loss} \end{array}$$

(or cross-entropy loss):

$$\mathcal{L}_{\mathsf{LOGREG}} = -\sum_{i=1}^{N} \left( \boldsymbol{\theta} \cdot \boldsymbol{f}(\boldsymbol{x}^{(i)}, y^{(i)}) - \log \sum_{y' \in \mathcal{Y}} \exp \left( \boldsymbol{\theta} \cdot \boldsymbol{f}(\boldsymbol{x}^{(i)}, y') \right) \right)$$

## Logistic Regression Objective

https://powcoder.com

Assignment Project Exam Help

Loss of a single is maked the Glat Exmontelp

 $\ell_{\mathsf{LOGREG}} = \frac{\theta \cdot f(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) + \log \sum_{\mathbf{y}' \in \mathcal{Y}} \exp \left(\theta \cdot f(\mathbf{x}^{(i)}, \mathbf{y}')\right)}{\mathsf{https://powcoder.com}}$ 

Add WeChat powcoder

#### Gradient of Logistic Regression

https://powcoder.com
The gradient with respect to the loss of a single example

Assignment Project Exam Help  $\frac{\partial}{\partial \theta} = -\mathbf{f}(\mathbf{x}^{(i)}, y^{(i)}) + \frac{1}{\mathbf{Assignment Project Exam Help}}$   $\times \sum_{y' \in \mathcal{Y}} \exp\left(\theta \cdot \mathbf{f}(\mathbf{x}^{(i)}, y')\right) \times \mathbf{f}(\mathbf{x}^{(i)}, y')$   $+ \frac{1}{\mathbf{f}(\mathbf{x}^{(i)}, y')} + \frac{1}{$  $= -\mathbf{f}(\mathbf{x}^{(i)}, \mathbf{y}_{\mathbf{A}}^{(i)}) + \sum_{\mathbf{y}' \in \mathbf{Y}} \underbrace{\exp\left(\mathbf{\theta} \cdot \mathbf{f}(\mathbf{x}^{(i)}, \mathbf{y}')\right)}_{\text{exp}\left(\mathbf{x}^{(i)}, \mathbf{y}'\right)} \times \mathbf{f}(\mathbf{x}^{(i)}, \mathbf{y}')$  $= -\boldsymbol{f}(\boldsymbol{x}^{(i)}, y^{(i)}) + \sum_{y' \in \mathcal{Y}} P(y'|\boldsymbol{x}^{(i)}; \boldsymbol{\theta}) \times \boldsymbol{f}(\boldsymbol{x}^{(i)}, y')$  $= -f(x^{(i)}, y^{(i)}) + E_{Y|X}[f(x^{(i)}, y)]$ 

Application of the chain rule in calculus, expectation

#### Gradient of Logistic Regression

## https://powcoder.com

$$\frac{\partial}{\partial \theta} = Assignment \sum_{y' \in \mathcal{Y}} Bye'' Example by Estable for the position of the positi$$

- This is a vehitips in powe oder.com
  - The gradient equals to the difference between the expected feature  $\mathbf{x}_{i}$  the cutrent  $\mathbf{x}_{i}$  for  $\mathbf{x}_{i}$  and the observed feature counts  $\mathbf{f}(\mathbf{x}^{(i)}, y^{(i)})$
  - ► The loss is minimized if the feature counts under the current model and the observed feature counts are the same
- ➤ The power of Logistic Regression model is that you can use arbitrary number of features without making any independence assumptions. The allows creative feature engineering to improve the performance of the model.

#### Digression: Expectation

Expectation is the mean or average of a random variable, for discrete case. Let X be a random variable.

discrete case. Let X be a random variable: Assignment Project Exam Help

$$\mathbb{E}[X] = \sum x P(x) = \mu$$

Assignment Project Example p  $\mathbb{E}[g(X)] = \sum_{g(X)} g(X) p(X)$ 

 $\frac{\text{https://powcoder.com}}{\mathbb{E}[g(X,Y)] = \sum g(x,y)p(x,y)}$ 

## Add WeChat powcoder

Let X be the random variable which is the value of rolling a single dice:

$$\mathbb{E}[x] = \sum_{x=1}^{6} xP(y) = \frac{1}{6} \sum_{x=1}^{6} x = \frac{21}{6} = 3.5$$

### Digression: Linearity of Expectations

https://powcoder.com
$$\mathbb{E}[X+Y] = \sum_{P(x,y)(x+y)} P(x,y)(x+y)$$
Assignment Project Exam Help
$$= \sum_{P(x,y)x+\sum_{P(x,y)y}} P(x,y)y$$
Assignment Project Exam Help
$$= \sum_{P(x,y)x+\sum_{P(x,y)y}} P(x,y)y$$

$$= \sum_{P(x,y)} P(x,y) \times \sum_{P(x,y)y} P(x,y)y$$
https://powcoder.com
$$= \sum_{P(x)x+\sum_{P(y)y}} P(y)y$$
Add WeChat powcoder

Let the Y be the random variable for the sum of two dice rolled. Expected value of Y:

$$\mathbb{E}[Y] = \mathbb{E}[X] + \mathbb{E}[X]$$

When two random variables are independent:

$$\mathbb{E}[X, Y] = \mathbb{E}[X] \times \mathbb{E}[X]$$

#### Digression: Variance

Variance of a random variable tend to be consistent over trials/experiments or whether they vary a lot Assignment Project Exam Help

## Assignaverty reflect paying their

$$\mathbb{E}[(X - \mathbb{E}[X]^2] = \sum_{x} \frac{\text{powcoder.com}}{(x - \mathbb{E}[X])^P(x)}$$

$$= \sum_{x} \frac{x^2 P(x) - 2\mathbb{E}[X]}{x} \sum_{x} x P(x) + (\mathbb{E}[X])^2 \sum_{x} P(x)$$

$$= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2$$

$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

 $ightharpoonup \sigma^2$  denotes the variance, and  $\sigma$  is the standard deviation.

#### Regularization

## https://powcoder.com

- ► L2 regulalisation meanth Perperce the Expansed Healp  $\|\theta\|_2^2$ to the minimization objective
- Regularization for the Cipactr Extended of formance on the training data against the norm of the weights, and thus helps relieve overfitting.

  The overall loss of a training set with a regularization term:

$$\mathcal{L}_{\text{LOGREG}} = \frac{\lambda}{2} \|\boldsymbol{\theta}\|_{2}^{2} - \sum_{i=1}^{k} \left( \frac{\mathbf{WeChat powcoder}}{\mathbf{\theta} \cdot \boldsymbol{f}(\boldsymbol{x}^{(i)}, y^{(i)})} - \log \sum_{y' \in \mathcal{Y}} \exp \left( \boldsymbol{\theta} \cdot \boldsymbol{f}(\boldsymbol{x}^{(i)}, y) \right) \right)$$

#### Regularized gradient

Derivative of the Laterps://powcoder.com

$$\frac{\lambda}{2} \|\boldsymbol{\theta}\|_{2}^{2} \text{ssignment}_{j}^{2} \text{Project}_{2}^{\lambda} \sum_{j=1}^{N} \text{am Help}$$

$$\frac{\lambda}{\partial \theta_{k}} \sum_{j=1}^{N} \|\boldsymbol{\theta}\|_{2}^{2} = \frac{\lambda}{\partial \theta_{k}} \sum_{j=1}$$

$$Ad \frac{\partial}{\partial \theta_k} \underbrace{W}_{j} = \underbrace{Chat \ prowkcoder}_{0 \ otherwise}$$

Gradient of regularized loss:

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_{\mathsf{LOGREG}} = \lambda \boldsymbol{\theta} - \sum_{i=1}^{N} \left( \boldsymbol{f}(\boldsymbol{x}^{(i)}, y^{(i)}) - E_{Y|X}[\boldsymbol{f}(\boldsymbol{x}^{(i)}, y')] \right)$$

# Batch Optimization vs online optimization <a href="https://powcoder.com">https://powcoder.com</a>

Assignment Project Exam Help Gradient Descent vs Stochastic Gradient Descent. In batch optimization, each update to the weights is based on the entire dataset. One such algorithm is gradient descent, which iteratively updates the weights,

https://powcoder.com

## Add WeChat poweoder

where  $\nabla_{\theta} \mathcal{L}$  is the gradient computed over the entire training set,  $\eta^{(t)}$  is the **learning rate** at iteration t.

#### Variations of Gradient Descent

## https://powcoder.com

## Assignment Project Exam Help

- The data. In **stochasttic gradient descent**, the approximate gradient is computed by sampling a single instance, and an update is made immediately.
- In mini-batch stochastic gradient descent, the gradient is computed over a small subset of instances.

#### Generalized gradient descent algorithm

https://powcoder.com
The function BATCHER partitions the training set into B batches such that each instance appears in exactly one batch. In stochastic gradient descent, B = A\S\si\g\adin\cap\ta\ser\to\pe€t1\bix\am\back\ta\ser\ta\se gradient descent, 1 < B < N.

```
1: procedures BPATTON POLICE POLICE BATCHER, T<sub>MAX</sub>)
             t \leftarrow 0
  2:
             \boldsymbol{\theta}^{(0)} \leftarrow \mathbf{0}
  3:
             repeat https://powcoder.com
  4:
                   (\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}, \cdots, \boldsymbol{b}^{(\widehat{B})}) \leftarrow \mathsf{BATCHER}(\mathsf{N})
  5:
                   for n \neq 1 We Chat powcoder
  6:
  7:
                         \boldsymbol{\theta}^{(t)} \leftarrow \boldsymbol{\theta}^{(t-1)} - \eta^{(t)} \nabla \mathcal{L}(\boldsymbol{\theta}^{(t-1)}; \boldsymbol{x}^{(b_1^{(n)}, b_2^{(n)}, \cdots)}, \boldsymbol{y}^{(b_1^{(n)}, b_2^{(n)}, \cdots)})
 8:
                         if Converged(\theta^{(1,2,\cdots,t)}) then return \theta^{(t)}
 9:
                         end if
10:
                   end for
11:
             until t = T_{MAX}
12:
13: end procedure
```

Binary logistic regression is a special case of multinominal logistic regression <a href="https://powcoder.com">https://powcoder.com</a>

$$P(y = 1 | \mathbf{x}) = \frac{\text{Assignment}(\sum_{k} \theta_{k} f_{k}(y = 1, \mathbf{x})) + \exp(\sum_{k} \theta_{k} f_{k}(y = 0, \mathbf{x}))}{\exp(\sum_{k} \theta_{k} f_{k}(y = 1, \mathbf{x})) + \exp(\sum_{k} \theta_{k} f_{k}(y = 0, \mathbf{x}))}$$

$$\times \frac{\exp(-\sum_{k} \theta_{k} f_{k}(y = 1, \mathbf{x}))}{\exp(\sum_{k} \theta_{k} f_{k}(y = 1, \mathbf{x}))}$$

$$= \frac{1}{1 + \exp(\sum_{k} \theta_{k} (f_{k}(y = 0, \mathbf{x}) - f_{k}(y = 1, \mathbf{x})))}$$

$$= \frac{1}{1 + \exp(\sum_{k} \theta_{k} (f_{k}(y = 0, \mathbf{x}) - f_{k}(y = 1, \mathbf{x})))}$$

$$= \frac{1}{1 + \exp(\sum_{k} -\theta_{k} f'_{k}(\mathbf{x}))} = \sigma(\theta \cdot \mathbf{f}(\mathbf{x}))$$

Note: For binary classification, you only need to pay attention to the positive class.

### Logistic Regression: features and weights

## https://powcoder.com f(x, y)

		Acc	<del>ionme</del>	nt F	roie	ct Ex	<del>am.</del>	Helr	
	not <sup>4</sup>	turir	iyenme	ok -	overa	ir story	good	Jokes	bias
POS	_	$f_4$	$f_7$	$f_{10}$	$f_{13}$	$f_{16}$	$f_{19}$	$f_{22}$	$f_{25}$
NEG	f <sub>2</sub> A <sub>5</sub>	Ssig	nnaent	<b>VRP</b>	idat	<b>PAW</b>	<b>POPE</b>	Mp .	$f_{26}$
NEU	$f_3$	$f_6$	f <sub>9</sub>	$f_{12}$	$f_{15}$	$f_{18}$	$f_{21}$	$f_{24}$	f <sub>27</sub>
		h	ttps://r	)OW	Code	er.coi	n		
POS	-1	2	-2.5	0.5	0.2	.08	1.5	8.0	1.2
NEG	2	-2 🔥	148W	e <b>c</b> h	at 1	<u>06</u>	đ <mark>ệr</mark>	-1.2	8.0
NEU	-0.4	-0.9	1.5	2	y po	-0.2	-1.2	-0.3	0.4

e.g., 
$$f_1(x,y)=1$$
 if x="not"  $\wedge$  y= "POS",  $heta_1=-1$ 

Note: The features f(x, y) and  $\theta$  are presented in a table rather than a vector due to limitation of space. Mathematically they should still be viewed as vectors

## https://powcoder.com

Assignment Project Exam Help funny painful ok overall story good jokes bias POS  $f_1$  $f_{25}$  $f_{26}$ NEU  $f_3$  $f_6$  $f_{12}$  $f_{15}$  $f_{18}$  $f_{21}$  $f_{24}$  $f_{27}$ https://pow**e**oder.com POS -1 -2.5 0.5 2 0.2 .08 1.5 8.0 1.2 NEG 2 d<sup>8</sup>We@hat1po -1.28.0 NEU -0.4 -0.9 -0.30.4

$$p(y|\mathbf{x}) = \frac{\exp(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y))}{\sum_{y}' \exp(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y'))} = \frac{\exp(\sum_{k} \theta_{k} f_{k}(\mathbf{x}, y))}{\sum_{y}' \exp(\sum_{k} \theta_{k} f_{k}(\mathbf{x}, y'))}$$

## https://powcoder.com

POS	not $f_1$	Fun f <sub>4</sub>	ny painfi Signm	ent P	overa LOJE f <sub>13</sub>	$ct_{f_{16}}^{\parallel \text{story}}$	g000 7119	Helf Helf	bias f <sub>25</sub>
NEG	$f_2$	$f_{5}$	f <sub>8</sub> 1 1	$f_{11}$	$f_{14}$	$f_{17}$	$f_{20}$	$f_{23}$	$f_{26}$
NEU	fz	5\$18	gningen	typeo	Hat	PA:M	ng pl		f <sub>27</sub>
				$\epsilon$	)				
POS	-1	2	1ttp\$5//	pow(	ode	er.eor	<b>11</b> .5	0.8	1.2
NEG	2	-2	1.8	-0.5	0.1	-0.6	-2	-1.2	0.8

$$p(y = POS|\mathbf{x})$$

$$= \frac{\exp(\theta_4 f_4 + \theta_{25} f_{25})}{\exp(\theta_4 f_4 + \theta_{25} f_{25}) + \exp(\theta_5 f_5 + \theta_{26} f_{26}) + \exp(\theta_6 f_6 + \theta_{27} f_{27})}$$

$$= \frac{\exp(2 + 1.2)}{\exp(2 + 1.2) + \exp(-2 + 0.8) + \exp(-0.9 + 0.4)} = 0.9643$$

## https://pow.com

POS	not $f_1$	fur f <sub>4</sub>	signm Signm		verall stor Ject E	X 2 000 X 2 119	Helf Helf T <sub>22</sub>	bias f <sub>25</sub>
NEG	$f_2$	$f_{5}$	f <sub>8</sub>	11 f <sub>1</sub>	$f_{17}$	$f_{20}$	$f_{23}$	$f_{26}$
NEU	fz	\$\$1	gnaten	tvr <sub>12</sub> oje	et Pas	MASE		f <sub>27</sub>
				heta				
POS	-1	2	http\$5//	powca	oder.00	<b>M</b> .5	8.0	1.2
NEG	2	-2	1.8	-0.5 0.	1 -0.6	-2	-1.2	0.8

$$p(y = NEG|\mathbf{x})$$

$$= \frac{\exp(\theta_5 f_5 + \theta_{26} f_{26})}{\exp(\theta_4 f_4 + \theta_{25} f_{25}) + \exp(\theta_5 f_5 + \theta_{26} f_{26}) + \exp(\theta_6 f_6 + \theta_{27} f_{27})}$$

$$= \frac{\exp(-2 + 0.8)}{\exp(2 + 1.2) + \exp(-2 + 0.8) + \exp(-0.9 + 0.4)} = 0.0118$$

## https://powcoder.com

POS	not $f_1$	fur f <sub>4</sub>	Signm Signm		ject Ex	g000 719	Help F <sub>22</sub>	bias f <sub>25</sub>
NEG	$f_2$	$f_{5_{\bullet}}$	f <sub>8</sub>	$f_{11}$ $f_{14}$	$f_{17}$	$f_{20}$	$f_{23}$	$f_{26}$
NEU	f <sub>3</sub> AS	581	gnagen	tyrzoyes	at Pasan	ng pr	elp	f <sub>27</sub>
				$\boldsymbol{\theta}$				
POS	-1	2	http\$5//	powco	der.cor	<b>11</b> .5	0.8	1.2
NEG	2	-2	1.8	-0.5 0.1	-0.6	-2	-1.2	8.0

$$p(y = NEU|\mathbf{x})$$

$$= \frac{\exp(\theta_6 f_6 + \theta_{27} f_{27})}{\exp(\theta_4 f_4 + \theta_{25} f_{25}) + \exp(\theta_5 f_5 + \theta_{26} f_{26}) + \exp(\theta_6 f_6 + \theta_{27} f_{27})}$$

$$= \frac{\exp(-9 + 0.4)}{\exp(2 + 1.2) + \exp(-2 + 0.8) + \exp(-0.9 + 0.4)} = 0.0238$$

#### Logistic Regression: Parameter estimation

## https://powcoder.com f(x, y)

		A gg	ianmo	nt [	Proje	of Ev	am	Halr	
	not '	funn	<mark>ignme</mark> yepamru	ok 1	overa	It story	good	jokes	bias
POS	-	$f_4$	$f_7$		$f_{13}$	$f_{16}$	$f_{19}$	$f_{22}$	$f_{25}$
NEG	$f_{\mathbf{A}}$	ssig	nnaetht	<b>MPE</b>	jaat	PAW	POPE	Mp	$f_{26}$
NEU	$f_3$	$f_6$	f <sub>9</sub>	$f_{12}$	$f_{15}$	f <sub>18</sub>	$f_{21}$	$f_{24}$	$f_{27}$
		h	ttps://r	ow	Code	er.cor	n		
POS	-1	2	-2.5	0.5	0.2	.08	1.5	8.0	1.2
NEG	2	-2 <u>A</u>	d <sup>1</sup> 18W	e (01)	At 1no	0.60	đ <mark>ệ</mark> r	-1.2	8.0
NEU	-0.4	-0.9	-1.5	2	1	-0.2	-1.2	-0.3	0.4

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}_{\mathsf{LOGREG}} = -\sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}^{(i)}, y^{(i)}) + \sum_{i=1}^{N} \sum_{y' \in Y} p(y'|\boldsymbol{x}^{(i)}) \boldsymbol{f}(\boldsymbol{x}^{(i)}, y')$$

### Logistic Regression: parameter estimation

## https://powcoder.com f(x, y)

	not	Auss	<u>igama</u>	ensk P	160g/c	C tstbry	<b>@ba</b> d	Heli	) bias
POS		$f_4$	$f_7$	$f_{10}$	$f_{13}$	$f_{16}$	$f_{19}$	$f_{22}$	$f_{25}$
NEG	$f_{2\Delta}$	<b>E</b>	nAradah	Wae.	Haat	<b>POW</b>	epple	<b>A</b>	f <sub>26</sub>
NEU	$f_3$	$f_6$	f <sub>9</sub>	$f_{12}$	$f_{15}$	$f_{18}$	$f_{21}$	$f_{24}$	f <sub>27</sub>
		h	ttng·//	now	$\theta$	er cor	n		
POS	-1	2	<sup>1</sup> 2.5	P 8.5°	0.2	.08	1.5	8.0	1.2
NEG	2	-2	1.8	-0.5	0.1	-0.6	<u>-</u> 2	-1.2	8.0
NEU	-0.4	-0.9	<u>uq.5</u>	eyn	at p	O MCD	uę,	-0.3	0.4

$$\begin{split} & p(y = \textit{NEG}|x) \\ & = \frac{\exp(\theta_2 f_2 + \theta_5 f_5 + \theta_{26} f_{26})}{\exp(\theta_1 f_1 + \theta_4 f_4 + \theta_{25} f_{25})) + \exp(\theta_2 f_2 + \theta_5 f_5 + \theta_{26} f_{26})) + \exp(\theta_3 f_3 + \theta_6 f_6 + \theta_{27} f_{27}))} \\ & = \frac{\exp(2 - 2 + 0.8)}{\exp(2 - 1 + 1.2) + \exp(2 - 2 + 0.8) + \exp(-0.9 - 0.4 + 0.4)} = 0.1909 \end{split}$$

### Logistic Regression: parameter estimation

## https://powcoder.com f(x, y)

	not	Auss	<u>igama</u>	ensk P	160gra	Ctstbry	<b>a</b> bad	Hal	) bias
POS		$f_4$	$f_7$	$f_{10}$	$f_{13}$	$f_{16}$	$f_{19}$	$f_{22}$	$f_{25}$
NEG	$f_{2}$	\$io	nAadah	W/e6	Haat	<b>EQM</b>	sopt e	<b>1</b> 73	f <sub>26</sub>
NEU	$f_3$	$f_6$	f <sub>9</sub>	$f_{12}$	$f_{15}$	$f_{18}$	$f_{21}$	f <sub>24</sub>	f <sub>27</sub>
		h	ttng·//	now	$\theta$	er cor	n		
POS	-1	2 11	2.5	Py.5 <sup>v</sup>	0.2	.08	1.5	8.0	1.2
NEG	2	-2	1,8	-0.5	0.1	-0.6	<u>-</u> 2	-1.2	8.0
NEU	-0.4	-0.9	a <u>q</u> 5	eyn	at p	O MCO	uer.	-0.3	0.4

$$p(y = POS|x)$$

$$= \frac{\exp(\theta_1 f_1 + \theta_4 f_4 + \theta_{25} f_{25}))}{\exp(\theta_1 f_1 + \theta_4 f_4 + \theta_{25} f_{25})) + \exp(\theta_2 f_2 + \theta_5 f_5 + \theta_{26} f_{26})) + \exp(\theta_3 f_3 + \theta_6 f_6 + \theta_{27} f_{27}))}$$

$$= \frac{\exp(2 - 1 + 1.2)}{\exp(2 - 1 + 1.2) + \exp(2 - 2 + 0.8) + \exp(-0.9 - 0.4 + 0.4)} = 0.7742$$

### Logistic Regression: Parameter estimation

## https://powcoder.com f(x, y)

	not	Auss	ygame	intk P	160era	Ctstbry	<b>aba</b> d	Flores	) bias
POS	$f_1$	$f_4$	$f_7$	$f_{10}$	$f_{13}$	$f_{16}$	$f_{19}$	$f_{22}$	$f_{25}$
NEG	$f_{2\Delta}$	(Sign	nAeddhi	Vøe(	Haat	<b>POW</b>	ROTE OF	173	f <sub>26</sub>
NEU	$f_3$	$f_6$	f <sub>9</sub>	$f_{12}$	$f_{15}$	$f_{18}$	$f_{21}$	$f_{24}$	f <sub>27</sub>
		h	ttng•//1		eode	er cor	n		
POS	-1	2	(P <sub>2.5</sub> //)	0.5	0.2	.08	1.5	8.0	1.2
NEG	2	-2	1.8	-0,5	0.1	-0.6	<u>-</u> 2	-1.2	8.0
NEU	-0.4	-0.9	<u>uq.5</u> vv	eyn	at po	J MGO	uer L	-0.3	0.4

$$p(y = NEU|x)$$

$$= \frac{\exp(\theta_3 f_3 + \theta_6 f_6 + \theta_{27} f_{27}))}{\exp(\theta_1 f_1 + \theta_4 f_4 + \theta_{25} f_{25})) + \exp(\theta_2 f_2 + \theta_5 f_5 + \theta_{26} f_{26})) + \exp(\theta_3 f_3 + \theta_6 f_6 + \theta_{27} f_{27}))}$$

$$= \frac{\exp(-0.9 - 4 + 0.4)}{\exp(2 - 1 + 1.2) + \exp(2 - 2 + 0.8) + \exp(-0.9 - 0.4 + 0.4)} = 0.0349$$

## Logistic Regression: Parameter estimation https://powcoder.com

	not	funn	y painfi	l ok p	overa	Il stery	good	Lipkes	bias
POS	$f_1$	$f_4$	$S_{7}$	$f_{10}$	$f_{13}$	$f_{16}$	$f_{19}$	f <sub>22</sub>	$f_{25}$
NEG	$f_2$	$f_{5}$	$f_8$	$f_{11}$	$f_{14}$	$f_{17}$	$f_{20}$	$f_{23}$	$f_{26}$
NEU	fzAS	SSIGI	nagen	typeo	<del>Madt</del>	PA:M	12 <b>4</b>	Mp	f <sub>27</sub>
				$\epsilon$	9				
POS	_1	2 h1	tt1205/	100 57	1000	er.080r	<u>1</u> .5	0.8	1.2
1 03	_T	<b>4</b> ]]	いりもり//		<b>Myu</b>		1.5	0.0	1.4
NEG	2	-2	1.8	-0.5	<b>-</b>	-0.6	-2	-1.2	0.8

$$gradient[\theta_1] = -0 + 0.7742, gradient[\theta_4] = -0 + 0.7742$$
  
 $gradient[\theta_2] = -1 + 0.1909, gradient[\theta_5] = -1 + 0.1909$   
 $gradient[\theta_3] = -0 + 0.0349, gradient[\theta_6] = -0 + 0.0349$ 

## Logistic Regression: Parameter estimation https://powycoder.com

	not	fun	ny painfi	ent Pro	erall story	/ goo	drjokeş	bias
POS	$f_1$	$f_4$		$f_{10} = f_{13}$	$f_{16}$	$f_{19}$	$f_{22}$	$f_{25}$
NEG	$f_2$	$f_{5}$	f <sub>8</sub>	$f_{11}$ $f_{12}$	$f_{17}$	$f_{20}$	$f_{23}$	$f_{26}$
NEU	f <sub>3</sub> A.S	5818	nngen	tyreoge	et Pra	M2H		f <sub>27</sub>
				heta				
POS	-1	2	nttp\$5//	powco	der.co	<b>m</b> 1.5	0.8	1.2
NEG	2	-2	1.8	-0.5 0.3	1 -0.6	-2	-1.2	0.8
NEU	-0.4	-0.9	<del>W 664</del>	/el hat	nowler	7445	-0.3	0.4

Training instance: y = NEG, x = "not funny at all"

$$\mathit{gradient}[\theta_{25}] = -0 + 0.7742$$

$$gradient[\theta_{26}] = -1 + 0.1909$$

$$gradient[\theta_{27}] = -0 + 0.0349$$

Note: The "bias" is the feature that always fires.

# Some observations about the gradient for Logistic Regression <a href="https://powcoder.com">https://powcoder.com</a>

## Assignment Project Exam Help

- The gladient of all the that the transfer to the partie of the same amount.
- A feature is penalized procewarde by an amount that is proportional to how badly off the prediction is.

## Discussion questiand WeChat powcoder

Assuming that the weights  $\theta$  are initialized as  $\mathbf{0}$ , how would the first iteration of the Logistic Regression Training proceed?

#### Adding a regularization term

	htt	ps://	po	WCC	odei	c.co	m	
POS -1								1.2
NEGA2SONEU -0.	sign	mle <sup>8</sup> n	t P	<del>'die</del>	<u>c</u> 4.6	$\Xi \hat{x}a$	m f	<del>Ielo</del>
NEU -0.	4 -0.9	-1.5	2	1	-0.2	-1.2	-0.3	0.4

Training intening int

$$\nabla_{\theta} \mathcal{L}_{\text{LOGREG}} = \lambda \theta - \sum_{i=1}^{httpS://powcodef.com} f(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) + \sum_{i=1}^{httpS://powcodef.com} p(\mathbf{y}'|\mathbf{x}^{(i)}) f(\mathbf{x}^{(i)}, \mathbf{y}')$$

$$Add WeChat powcoder$$

Let 
$$\lambda = 0.5$$
:

gradient
$$[\theta_1] = -0.5 - 0 + 0.7742$$
, gradient $[\theta_4] = 1 - 0 + 0.7742$   
gradient $[\theta_2] = 1 - 1 + 0.1909$ , gradient $[\theta_5] = -1 - 1 + 0.1909$   
gradient $[\theta_3] = -0.2 - 0 + 0.0349$ , gradient $[\theta_6] = -0.45 - 0 + 0.0349$