1.) $X = \{(x_i, y_i) \dots (x_n, y_n)\}$ $\quad x_i \in \mathbb{R}^d$ $\quad y_i \in \{0, 1\}$

$a \in \mathbb{R} \quad \sigma(a) = \frac{1}{1+e^{-a}}$

$L(y_i|x_i, w) = y_i \log(\sigma(w^T x_i)) + (1-y_i)\log(\sigma(-w^T x_i))$

① Make arg to σ the same to simplify things

$$\sigma(-a) = \frac{1}{1+e^{a}}$$

$$= \frac{e^{-a}}{e^{-a}+1}$$ (multiply top and bottom by e^{-a})

$$= \frac{e^{-a}+(1-1)}{e^{-a}+1} = \frac{e^{-a}+1}{e^{-a}+1} - \frac{1}{e^{-a}+1}$$

$$= 1 - \sigma(a)$$

$$\therefore \sigma(-a) = \sigma(a)$$

let $a_i = w^T x_i$

$L(y_i|x_i, w) = y_i \log(\sigma(a_i)) + (1-y_i)\log(1 - \sigma(a_i))$

② Use chain rule

let $z_i = \sigma(a_i)$

$L(y_i|x_i, w) = y_i \log(z_i) + (1-y_i)\log(1-z_i)$

$$\frac{\partial L(y_i|x_i,w)}{\partial w} = \frac{\partial L}{\partial z_i}\frac{\partial z_i}{\partial a_i}\frac{\partial a_i}{\partial w}$$ y_i doesn't depend on w

$$L(y_i|x_i,w) = y_i \log(z_i) + (1-y_i)\log(1-z_i)$$

(3) Find $\frac{\partial L}{\partial z_i}$

$$\frac{\partial L}{\partial z_i} = \frac{y_i}{z_i} + -\left(\frac{(1-y_i)}{(1-z_i)}\right)$$

chain rule

log derivative rule

$$= \frac{y_i(1-z_i) - z_i(1-y_i)}{z_i(1-z_i)}$$

$$= \frac{y_i - z_i}{z_i(1-z_i)}$$

(4) Find $\frac{\partial z_i}{\partial a_i}$ $\quad z_i = \sigma(a_i)$

$$\sigma(x) = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1}$$

$$\frac{\partial \sigma(x)}{\partial x} = -(1+e^{-x})^{-2} \cdot -e^{-x} = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \cancel{\frac{e^{-x}}{1+2e^{-x}+e^{-2x}}} \; \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{(1+e^{-x})}$$ ... From previous result

$$= \cancel{\sigma(x)(1-x)} \; \sigma(x)\,(1-\sigma(x))$$

$$\frac{\partial z_i}{\partial a_i} = z_i(1-z_i)$$

$$a_i = w^T x_i$$

⑤ Find $\dfrac{\partial a_i}{\partial w} = x_i$ (power rule/constant rule)

⑥ Put it together (Yay!)

$$\frac{\partial L(y_i | x_i, w)}{\partial w} = \frac{\partial L}{\partial z_i} \frac{\partial z_i}{\partial a_i} \frac{\partial a_i}{\partial w}$$

$$= \frac{y_i - z_i}{z_i(1-z_i)} \cdot z_i(1-z_i) \cdot x_i$$

$$\frac{\partial L}{\partial w} = (y_i - z_i) \cdot x_i$$

$$= (y_i - \sigma(w^T x_i)) \cdot x_i$$

$$\cancel{\frac{\partial L(y_i|x_i,w)}{\partial w_j} = (y - \sigma(w^T x_i))[x_{ij}]}$$

$$\boxed{\frac{\partial L(y_i | x_i, w)}{\partial w_j} = (y_i - \sigma(w^T x_i))\, x_{ij}}$$

$$\frac{\partial}{\partial w}(1-y_i(w^T x_i+b)) = -y_i x_i$$

$$\frac{\partial}{\partial w}(0) = 0$$

3.) $X = \{(x_i, y_i) \dots (x_n, y_n)\}$ $\quad x_i \in \mathbb{R}^d$

$y_i \in \{-1, 1\}$

$$f(w) = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n} \max(0, 1-y_i(w^T x_i + b))$$

$$\frac{\partial f}{\partial w} = w + C \cdot \sum_{i=1}^{n}\left([1-y_i(w^T x_i + b) > 0] \cdot -y_i x_i\right)$$

where $[x] = \begin{cases} 0 & \text{if } x \text{ false} \\ 1 & \text{if } x \text{ true} \end{cases}$

iverson bracket

$$\boxed{\frac{\partial f(w)}{\partial w_j} = w_j + C\sum_{i=1}^{n}\left([1-y_i(w^T x_i + b) > 0] \cdot -y_i x_{ij}\right)}$$

CSCI 5525 HW2

Daniel Chang

February 2023

2: Logistic Regression

Error rate for Logistic Regression												
η	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.00000	0.50000	0.54375	0.53125	0.51875	0.48750	0.50000	0.63125	0.58750	0.55000	0.47500	0.53250	0.04565
0.00001	0.15000	0.21250	0.14375	0.21250	0.15625	0.15000	0.20000	0.19375	0.13125	0.11250	0.16625	0.03380
0.00010	0.01250	0.00625	0.00625	0.03750	0.01250	0.00625	0.03750	0.01875	0.01875	0.01250	0.01688	0.01120
0.00100	0.01250	0.00625	0.00000	0.03125	0.01875	0.00625	0.03750	0.01875	0.01250	0.00625	0.01500	0.01125
0.01000	0.01250	0.00625	0.00000	0.02500	0.01250	0.00625	0.03125	0.01250	0.00625	0.00625	0.01188	0.00904

Best $\eta = 0.01$.
Logistic Regression test error rate: 0.005

4: SVM

Error rate for SVM												
η, c	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
(1e-05, 0.01)	0.01250	0.00625	0.00000	0.03125	0.01875	0.00625	0.03125	0.01875	0.01250	0.00625	0.01438	0.01010
(1e-05, 0.1)	0.01250	0.00625	0.00000	0.02500	0.01875	0.00625	0.03125	0.01250	0.00625	0.00625	0.01250	0.00927
(1e-05, 1)	0.00625	0.00625	0.00000	0.02500	0.01875	0.00625	0.03125	0.01250	0.01250	0.01250	0.01313	0.00904
(1e-05, 10)	0.00625	0.00625	0.00000	0.02500	0.01875	0.00625	0.03125	0.00625	0.01250	0.01250	0.01250	0.00927
(1e-05, 100)	0.01250	0.00625	0.00000	0.02500	0.01875	0.00625	0.03125	0.00625	0.01250	0.00625	0.01250	0.00927
(0.0001, 0.01)	0.01250	0.00625	0.00000	0.02500	0.01250	0.00625	0.03125	0.01250	0.00625	0.00625	0.01188	0.00904
(0.0001, 0.1)	0.00625	0.00625	0.00000	0.02500	0.01250	0.00000	0.03125	0.01250	0.00625	0.00625	0.01063	0.00970
(0.0001, 1)	0.00625	0.00625	0.00000	0.02500	0.01875	0.00625	0.03125	0.01250	0.01250	0.01250	0.01313	0.00904
(0.0001, 10)	0.01250	0.00625	0.00000	0.02500	0.01875	0.00625	0.03125	0.00625	0.01250	0.01250	0.01313	0.00904
(0.0001, 100)	0.00625	0.00625	0.00000	0.02500	0.01250	0.00625	0.03125	0.00625	0.31250	0.01250	0.04188	0.09066
(0.001, 0.01)	0.01250	0.00625	0.00000	0.02500	0.01250	0.00625	0.03125	0.01250	0.00625	0.00625	0.01188	0.00904
(0.001, 0.1)	0.00625	0.00625	0.00000	0.02500	0.01250	0.00000	0.03125	0.01250	0.00625	0.00625	0.01063	0.00970
(0.001, 1)	0.00625	0.00625	0.00000	0.02500	0.01875	0.00625	0.03125	0.01875	0.01250	0.00625	0.01313	0.00946
(0.001, 10)	0.00625	0.00625	0.00000	0.03125	0.01875	0.00625	0.03125	0.01250	0.08750	0.36250	0.05625	0.10488
(0.001, 100)	0.01250	0.00625	0.00000	0.02500	0.01875	0.00625	0.03125	0.01250	0.02500	0.04375	0.01813	0.01264

Best η and c: 0.0001, 0.1
Test error rate: 0.0075