ELE 888
Midterm
Page#1

Q5 dataset has 235 samples, 25 = class A, 20 = class B
190 = class C, if k-NN, & k = 100
→ probability of class C = higher, but A & B possible

Q6

a) Linear Regression → since amount of mortgage is numerical
→ not classification

b) $\hat{y} = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n$
→ need general model

$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h(x^i) - y^i)^2$ where ~~$h(x) = \theta_1 x + \theta_0$~~

→ error (for training) total i.e. if training set given

$h(x_i) = \theta_0 + x_1\theta_1 + x_2\theta_2 + \dots + x_m\theta_m$
→ Model

c) θ_0 bias weight = 100,000, θ_1 walk-score weight = 2000, θ_2 bedroom weights = 10000
θ_3 annual income weight = $x_4 + 0.5 = 8.5$ ($x_4 \to 8$), estimated mortgage amount (#) for following
x_1 ← walk = 62, bed = 4 (x_2), income = 50k (x_3)
mortgage = ?

$h(x_1^i) = 100,000 + (2000)(62) + (10,000)(4) + (8.5)(50,000)$
$= 689 \times 10^3$ → \$689k → estimated mortgage

Q7 Logistic Regression, given $\log_e(\text{odds}) = \underbrace{-25.8 + (0.5)\frac{\text{\# of cells white blood}}{1000}}_{z}$

white blood cell count	coronavirus
20, 800	"−" = 0
17, 600	"−" = 0
35, 200	"+" = 1
30, 100	"+" = 1
53, 800	"+" = 1

a) $p = \frac{\text{odds}}{1-\text{odds}}$, 80% < p (or 0.8)
& we know
$p > \frac{1}{1+e^{-z}} > \frac{e^z}{e^z+1} > 0.8$
$0.8 = \frac{e^z}{e^z+1} \to e^z = (0.8)(e^z+1)$

→ but $\log_e(\text{odds}) = z = \log\left(\frac{p}{1-p}\right)$
→ let c = # of white blood cells
→ $-25.8 + 0.5\left(\frac{c}{1000}\right) = \log\left(\frac{0.8}{1-0.8}\right) \to \log(4)$

→ $0.5\left(\frac{c}{1000}\right) = \log(4) + 25.8 \to c = \frac{1000(\log(4) + 25.8)}{0.5}$
→ $c \geq 52,804$ → ∴ ($c > 52,804$) for ($p > 80\%$)

Page#2

cont'd threshold

$\boxed{Q7}$ b) $p \geq 60\% \rightarrow$ positive & $p < 60\% \rightarrow$ negative

$\hookrightarrow$ find accuracy (compared to given table)

$$Logloss = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log_e(p_i) + (1-y_i)\log_e(1-p_i)\right]$$

$\hookrightarrow N=5$

$$Accuracy = \frac{\text{\# of correct classifications}}{\text{Total classifications}}$$

~~given~~ given $P = \frac{1}{1+e^{-z}} \geq 60\%$ then "+"

if $P = \frac{1}{1+e^{-z}} < 60$ then "−", $Z = -25.8 + 0.5\left(\frac{C}{1000}\right)$ ← # of cells

	# of cells	test result
1	20 800	"−" = 0
2	17,600	"−" = 0
3	35,200	"−" = 0
4	30,100	"−" = 0
5	53,800	"+" = 1

$Z_1 = -25.8 + 0.5\left(\frac{20800}{1000}\right) = -15.4$, $P = \frac{1}{1+e^{15.4}} << 60\% \Rightarrow$ result 0

$Z_2 = -17$, $P = \frac{1}{1+e^{17}}$ $<< 60\% \Rightarrow$ result 0

$Z_3 = -8.2$, $P = \frac{1}{1+e^{8.2}}$ $<< 60\% \rightarrow$ result = 0

$Z_4 = -10.75$, $P << 60\% \rightarrow$ result = 0

$Z_5 = 1.1$, $P = \frac{1}{1+e^{-1.1}} =$ 0.75 or 75% $\hookrightarrow$ result = 1

$\rightarrow$ correct = 3

so Accuracy = $3/5 = 0.6$ or 60% accuracy

c) $\hookrightarrow z = -25.8 + 0.5\left(\frac{17600}{1000}\right) = -17$ $\hookrightarrow$ negative = 0

so $P = \frac{1}{1+e^{17}} \rightarrow$ sto "A" in calculator

$$Logloss = -\frac{1}{5}\sum_{i=1}^{5}\left(\overset{0}{y_i \log_e(p_i)} + (1-y_i)\log_e(1-p_i)\right)$$

$$= -\frac{1}{5}(1)\log_e(1-P) \rightarrow \text{log loss} = 8\times10^{-9}$$

$\boxed{Q8}$ Total score = Round$\left(\sum_{j=0}^{n} w_j \times x_j\right)$

$\hookrightarrow$ predict total score given 1st quarter

1st Quarter score	total score
27	104
36	123

$w_0 = 5$, $w_1 = 3.5$, $\lambda = 0.01$, $\alpha = 0.01$

a) error/cost = ?, error $= \frac{1}{2m}\sum_{i=1}^{m}\left(y^{(i)} - \sum_{j=0}^{n} w_j \times x_j^{(i)}\right)^2 + \frac{\lambda}{2}\sum_{j=1}^{n}(w_j)^2$

~~$\hookrightarrow$ total score$_1$ = Round$(5 + 3.5(27))$ = (100 = total score$_1$)~~

~~Total score$_2$ = $(5 + 3.5(36))$ = 131 = total score$_2$~~

$$Error = \frac{1}{2(2)}\sum_{i=1}^{2}\left[(104 - \cancel{100})^2 + (123 - \cancel{131})^2\right] + \frac{0.01}{2}\left[5^2 + 3.5^2\right]$$

$\hookrightarrow 5\times27 = 135$ $\quad\hookrightarrow 3.5\times36 = 126$

Page# 3

cont'd

$\boxed{Q8}$ a) ~~Error = 20.18~~ Error = -5.313

b) find new weights & errors after 1 Epoch (e.g cycle)

~~Error = 20.18~~

$$w_{new} \leftarrow w_{old} - \alpha\left(\frac{1}{m}\sum_{i=1}^{m} x_k^{(i)} \times \left(y^i - \sum_{j=0}^{n} w_{old}\, x_j^{(i)}\right)\right) - (\alpha)(\lambda)(w_{old})$$

$$= w_{old} - (0.01)\left(\frac{1}{2}\, \cancel{(27+36)}\left[\cancel{(104-100)} + (123-131)\right]\right) - (0.01)(0.01)\, w_{old}$$

0.9 $\quad [27(104-100) + 36(123-131)]$

$w_{new} = w_{old} - 1.26 - 0.0001\, w_{old}$, now we can find $w_{0,new}$ & $w_{1,new}$

weights	Initial	New
w_0	5	~~3.75~~
w_1	3.5	~~2.25~~

$w_{0,new} =$ ~~3.7495~~ → sto "B" → $\boxed{5.8995}$ — $w_{0,new}$

$w_{1,new} =$ ~~2.24965~~ → sto "C" → $\boxed{4.39965}$ — $w_{1,new}$

c) New cost based on updated weights

Total score$_{1,new} = 5.8$ ~~3.7495~~ + ~~2.24965~~ $(27) =$ ~~64.49~~ 124.6

Total score$_{2,new} =$ " " $(36) =$ ~~84.7369~~ 164.2869

$$Error_{new} = \frac{1}{2(2)}\left[\left(\cancel{64.49} - 10\cancel{0}4\right)^2 + \left(\cancel{164.2869} - 131\right)^2 + \frac{0.01}{2}\left[5.8995^2 + 4.39965^2\right]\right]$$

$(104 - \cancel{124.6})$ $(123 - \cancel{164.289})$

$\to (5.8995)(27) = 159.3$ $\to (4.39965)(36) = 158.38$

~~= 428.565~~ ~~Error = 430.69~~

Error = 532.37

~~a lot bigger~~ → bigger then before → arithmetic error? :(

d) Total score $= w_0 + w_1(x_1)$

$= 5.8995 + 4.39965(40)$

Total score $= 181.886$

K-NN model parameter: not set by user

2. ~~1/~~ model hypeparameter: set by user

Epoch: complete cycle, batch size: samples at a time

1, 2, 3, 4 Regression tasks

Ridge Regression: reduces

Lasso Regression: removes features & reduces

Notes from multiple Choice Questions