HW2 Report

Qiong Wang 5906740674

1. Part A

   In this question I use 3 different method to preprocess the data and use cv to select the best model for logistic regression.

   For different preprocess method and different lambda, I use train set to do logistic regression and calculate average cv error (5 times) for different model. Find out the minimum cv error, and this model is considered the best model.

   When standardized data set, and when lambda = 0.8912509381337491
   cv error rate = 0.05897650465861448

   When log-transform data set, and when lambda = 0.8912509381337491
   cv error rate = 0.05897650465861448

   When binarized data set, and when lambda = 0.44668359215096465
   cv error rate = 0.06745973557306417

   Then I calculate test error and train error.

preprocessing method	λ	average cv error	train error	test error
standerlized	0.158489319	0.079208979	0.07406199	0.080729167
log transform	0.891250938	0.058976505	0.053181077	0.055989583
binarize	0.446683592	0.067459736	0.063947798	0.072916667

   The data below is all data that will print after running the code. I print all lambda value and all cv error to check whether I select the minimum cv error or not.

/Users/qiongwang/PycharmProjects/660/HW2/venv/bin/python
"/Users/qiongwang/PycharmProjects/660/HW2/check spam.py"
standardized result:
when lambda = 0.01 cv error rate = 0.0929591929777247
when lambda = 0.011220184543019636 cv error rate = 0.09030613293264744
when lambda = 0.012589254117941675 cv error rate = 0.09073731322198886
when lambda = 0.014125375446227547 cv error rate = 0.09008950575664743
when lambda = 0.01584893192461114 cv error rate = 0.09141539044390079
when lambda = 0.017782794100389236 cv error rate = 0.10076619598928926
when lambda = 0.01995262314968881 cv error rate = 0.08640234889201392
when lambda = 0.022387211385683413 cv error rate = 0.08400390411794556
when lambda = 0.025118864315095822 cv error rate = 0.08381650645656347
when lambda = 0.028183829312644564 cv error rate = 0.08465634772334707
when lambda = 0.031622776601683826 cv error rate = 0.08637428804428082
when lambda = 0.035481338923357586 cv error rate = 0.09207160332106867
when lambda = 0.039810717055349776 cv error rate = 0.08291977294471953
when lambda = 0.04466835921509637 cv error rate = 0.08945194660091305
when lambda = 0.050118723362727303 cv error rate = 0.08814475416257306
when lambda = 0.056234132519034995 cv error rate = 0.08903265846031361
when lambda = 0.06309573444801943 cv error rate = 0.08398228378057326
when lambda = 0.07079457843841391 cv error rate = 0.09204255360135627
when lambda = 0.07943282347242829 cv error rate = 0.09228369067570785
when lambda = 0.08912509381337473 cv error rate = 0.09228297791733286
when lambda = 0.1000000000000002 cv error rate = 0.08137973968266188
when lambda = 0.11220184543019658 cv error rate = 0.08663314133805944
when lambda = 0.125892541179417 cv error rate = 0.08989173778197301
when lambda = 0.14125375446227575 cv error rate = 0.08945845774498662
when lambda = 0.15848931924611173 cv error rate = 0.07920897947127448
when lambda = 0.17782794100389274 cv error rate = 0.08183239262070341
when lambda = 0.1995262314968885 cv error rate = 0.08464925866707751
when lambda = 0.2238721138568346 cv error rate = 0.09095560992211027
when lambda = 0.25118864315095873 cv error rate = 0.09010836495797303
when lambda = 0.2818382931264462 cv error rate = 0.0896661337031972
when lambda = 0.3162277660168389 cv error rate = 0.09030073266423932
when lambda = 0.3548133892335766 cv error rate = 0.08401578342419391
when lambda = 0.39810717055349853 cv error rate = 0.09208280839642202
when lambda = 0.44668359215096465 cv error rate = 0.09315120109418051
when lambda = 0.501187233627274 cv error rate = 0.09139888784008521
when lambda = 0.5623413251903511 cv error rate = 0.08682776290188987
when lambda = 0.6309573444801956 cv error rate = 0.08595943056384958
when lambda = 0.7079457843841406 cv error rate = 0.09271480033133628

when lambda = 0.7943282347242846 cv error rate = 0.09183471711198021
when lambda = 0.8912509381337491 cv error rate = 0.08598560356507612
when lambda = 1.000000000000004 cv error rate = 0.08551635812576652
when lambda = 1.1220184543019676 cv error rate = 0.0918406952925842
when lambda = 1.2589254117941726 cv error rate = 0.08880989257254401
when lambda = 1.4125375446227613 cv error rate = 0.09509809096337951
when lambda = 1.5848931924611207 cv error rate = 0.08816559752910424
when lambda = 1.77827941003893 cv error rate = 0.09052784573597117
when lambda = 1.995262314968889 cv error rate = 0.08596144683528861
when lambda = 2.2387211385683514 cv error rate = 0.09311100409033413
when lambda = 2.5118864315095926 cv error rate = 0.08770364662595598
when lambda = 2.8183829312644666 cv error rate = 0.0926912279349914
when lambda = 3.1622776601683955 cv error rate = 0.08532251995402385
when lambda = 3.548133892335775 cv error rate = 0.09380324658229144
when lambda = 3.9810717055349936 cv error rate = 0.08814536418100205
when lambda = 4.466835921509653 cv error rate = 0.09684529932641117
when lambda = 5.01187233627275 cv error rate = 0.09161774960990932
when lambda = 5.623413251903525 cv error rate = 0.09399589682340925
when lambda = 6.309573444801969 cv error rate = 0.09270934227170857
when lambda = 7.079457843841417 cv error rate = 0.09813699729665504
when lambda = 7.943282347242862 cv error rate = 0.09358128335034954
when lambda = 8.912509381337514 cv error rate = 0.08620795207181531
when lambda = 10.000000000000062 cv error rate = 0.09467873861031395
when lambda = 11.220184543019698 cv error rate = 0.08637991947756729
when lambda = 12.589254117941753 cv error rate = 0.09662366357804708
when lambda = 14.125375446227642 cv error rate = 0.08705675739888152
when lambda = 15.84893192461124 cv error rate = 0.09792650241117806
when lambda = 17.78279410038934 cv error rate = 0.09599940282406438
when lambda = 19.952623149688932 cv error rate = 0.10491505974969983
when lambda = 22.38721138568356 cv error rate = 0.10012277423539007
when lambda = 25.118864315095976 cv error rate = 0.10358400595891692
when lambda = 28.183829312644722 cv error rate = 0.09966858661940636
when lambda = 31.62277660168402 cv error rate = 0.091194069336621
when lambda = 35.481338923357825 cv error rate = 0.09860972947287983
when lambda = 39.81071705535002 cv error rate = 0.10163607584776513
when lambda = 44.66835921509662 cv error rate = 0.10425824969659614
when lambda = 50.11872336272761 cv error rate = 0.10860228082679968
when lambda = 56.23413251903537 cv error rate = 0.10098694560561994
when lambda = 63.09573444801982 cv error rate = 0.10859420931979746
when lambda = 70.7945784384143 cv error rate = 0.1079596553074813
when lambda = 79.43282347242878 cv error rate = 0.10706965768334276

when lambda = 89.12509381337532 cv error rate = 0.11883840290754044
the best model is when lambda = 0.15848931924611173
cv error rate = 0.07920897947127448
test error = 0.08072916666666663
train error = 0.07406199021207183

log transform result:
when lambda = 0.01 cv error rate = 0.07202260278810524
when lambda = 0.011220184543019636 cv error rate = 0.07527934991299201
when lambda = 0.012589254117941675 cv error rate = 0.07438949997752553
when lambda = 0.014125375446227547 cv error rate = 0.07399314852985572
when lambda = 0.01584893192461114 cv error rate = 0.07380360617210235
when lambda = 0.017782794100389236 cv error rate = 0.07768531396685352
when lambda = 0.01995262314968881 cv error rate = 0.07071216119897517
when lambda = 0.022387211385683413 cv error rate = 0.07354984492689398
when lambda = 0.025118864315095822 cv error rate = 0.07332338682231776
when lambda = 0.028183829312644564 cv error rate = 0.06352916851277501
when lambda = 0.031622776601683826 cv error rate = 0.07178748242183741
when lambda = 0.035481338923357586 cv error rate = 0.07791261967598384
when lambda = 0.039810717055349776 cv error rate = 0.07377007442224826
when lambda = 0.04466835921509637 cv error rate = 0.07531416591217022
when lambda = 0.050118723362727303 cv error rate = 0.07225906519491687
when lambda = 0.056234132519034995 cv error rate = 0.06572766850956435
when lambda = 0.06309573444801943 cv error rate = 0.07136722467299794
when lambda = 0.07079457843841391 cv error rate = 0.07924258827608788
when lambda = 0.07943282347242829 cv error rate = 0.07702829843385783
when lambda = 0.08912509381337473 cv error rate = 0.06920701456980838
when lambda = 0.1000000000000002 cv error rate = 0.07026529380413904
when lambda = 0.11220184543019658 cv error rate = 0.06986064610583487
when lambda = 0.125892541179417 cv error rate = 0.07007781908779775
when lambda = 0.14125375446227575 cv error rate = 0.07505924884257031
when lambda = 0.15848931924611173 cv error rate = 0.07680504453134529
when lambda = 0.17782794100389274 cv error rate = 0.07179222130184348
when lambda = 0.1995262314968885 cv error rate = 0.06657781587717437
when lambda = 0.2238721138568346 cv error rate = 0.0685419146873173
when lambda = 0.25118864315095873 cv error rate = 0.06987005965338111
when lambda = 0.2818382931264462 cv error rate = 0.06308813803111735
when lambda = 0.3162277660168389 cv error rate = 0.06637165533316625
when lambda = 0.3548133892335766 cv error rate = 0.07008259649528359
when lambda = 0.39810717055349853 cv error rate = 0.06725414652000539
when lambda = 0.44668359215096465 cv error rate = 0.0633303731386411

when lambda = 0.501187233627274 cv error rate = 0.0698525232288596
when lambda = 0.5623413251903511 cv error rate = 0.06485432117791345
when lambda = 0.6309573444801956 cv error rate = 0.06878549183538496
when lambda = 0.7079457843841406 cv error rate = 0.062250248823306764
when lambda = 0.7943282347242846 cv error rate = 0.06835081196663517
when lambda = 0.8912509381337491 cv error rate = 0.05897650465861448
when lambda = 1.000000000000004 cv error rate = 0.06569231312567025
when lambda = 1.1220184543019676 cv error rate = 0.06941946151425837
when lambda = 1.2589254117941726 cv error rate = 0.06245976125805053
when lambda = 1.4125375446227613 cv error rate = 0.06351819460230002
when lambda = 1.5848931924611207 cv error rate = 0.06093887615341631
when lambda = 1.77827941003893 cv error rate = 0.07006310159054274
when lambda = 1.995262314968889 cv error rate = 0.06398156460095161
when lambda = 2.2387211385683514 cv error rate = 0.07073098829406743
when lambda = 2.5118864315095926 cv error rate = 0.06332225026166571
when lambda = 2.8183829312644666 cv error rate = 0.06638912754522153
when lambda = 3.1622776601683955 cv error rate = 0.06483690033583112
when lambda = 3.548133892335775 cv error rate = 0.06245464994574035
when lambda = 3.9810717055349936 cv error rate = 0.06463456685480917
when lambda = 4.466835921509653 cv error rate = 0.0641810085209944
when lambda = 5.01187233627275 cv error rate = 0.06092699684716796
when lambda = 5.623413251903525 cv error rate = 0.06439375084278853
when lambda = 6.309573444801969 cv error rate = 0.063994580467852
when lambda = 7.079457843841417 cv error rate = 0.06634769124077733
when lambda = 7.943282347242862 cv error rate = 0.06831304219401169
when lambda = 8.912509381337514 cv error rate = 0.060281847777927666
when lambda = 10.000000000000062 cv error rate = 0.06137161680568659
when lambda = 11.220184543019698 cv error rate = 0.06246709432169162
when lambda = 12.589254117941753 cv error rate = 0.06573483462079321
when lambda = 14.125375446227642 cv error rate = 0.06245631304861532
when lambda = 15.84893192461124 cv error rate = 0.06658968234092966
when lambda = 17.78279410038934 cv error rate = 0.07051735341899268
when lambda = 19.952623149688932 cv error rate = 0.06723839520204444
when lambda = 22.38721138568356 cv error rate = 0.06772187012386577
when lambda = 25.118864315095976 cv error rate = 0.06702242299319994
when lambda = 28.183829312644722 cv error rate = 0.06331721600431517
when lambda = 31.62277660168402 cv error rate = 0.06635844682886738
when lambda = 35.481338923357825 cv error rate = 0.06700869436792445
when lambda = 39.81071705535002 cv error rate = 0.06920545420687974
when lambda = 44.66835921509662 cv error rate = 0.07531728021678119
when lambda = 50.11872336272761 cv error rate = 0.07420743837208565

when lambda = 56.23413251903537 cv error rate = 0.07399958904021631
when lambda = 63.09573444801982 cv error rate = 0.07267037172596691
when lambda = 70.7945784384143 cv error rate = 0.0783435688004468
when lambda = 79.43282347242878 cv error rate = 0.07942902917172356
when lambda = 89.12509381337532 cv error rate = 0.08250485125182205
the best model is when lambda = 0.8912509381337491
cv error rate = 0.05897650465861448
test error = 0.05598958333333337
train error = 0.053181076672104366

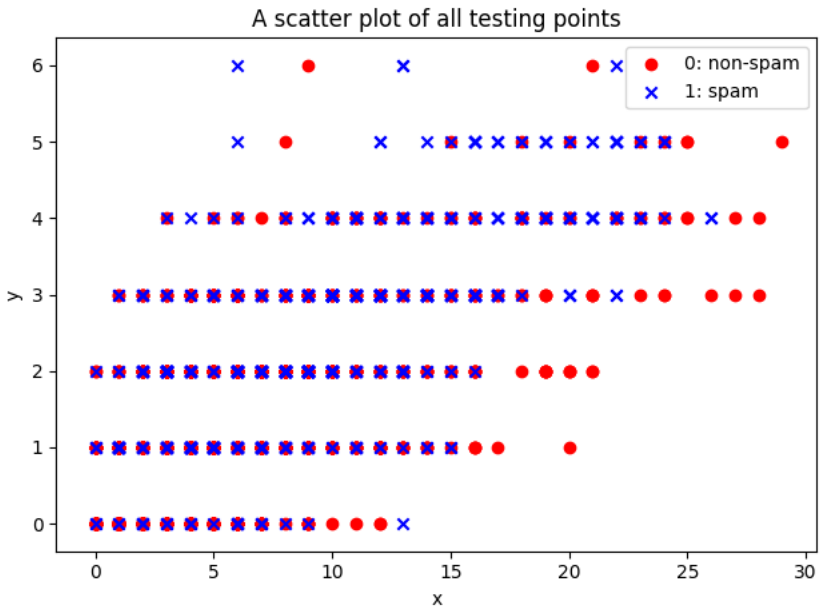
binarized result:
when lambda = 0.01 cv error rate = 0.08898483301548166
when lambda = 0.014125375446227547 cv error rate = 0.08444960926714307
when lambda = 0.01584893192461114 cv error rate = 0.08246582933610735
when lambda = 0.017782794100389236 cv error rate = 0.08379138011853615
when lambda = 0.01995262314968881 cv error rate = 0.08136856671354187
when lambda = 0.022387211385683413 cv error rate = 0.08445436098964243
when lambda = 0.025118864315095822 cv error rate = 0.0779171916035779
when lambda = 0.028183829312644564 cv error rate = 0.07856788220865207
when lambda = 0.031622776601683826 cv error rate = 0.08641028555283725
when lambda = 0.035481338923357586 cv error rate = 0.08249621467511703
when lambda = 0.039810717055349776 cv error rate = 0.07488188758965664
when lambda = 0.04466835921509637 cv error rate = 0.08073826998773548
when lambda = 0.050118723362727303 cv error rate = 0.0781368431867362
when lambda = 0.056234132519034995 cv error rate = 0.07919434545022563
when lambda = 0.06309573444801943 cv error rate = 0.08488053270661966
when lambda = 0.07079457843841391 cv error rate = 0.08094714029781758
when lambda = 0.07943282347242829 cv error rate = 0.08575513218136166
when lambda = 0.08912509381337473 cv error rate = 0.08227637045456004
when lambda = 0.1000000000000002 cv error rate = 0.08751075237746642
when lambda = 0.11220184543019658 cv error rate = 0.06966335972465687
when lambda = 0.125892541179417 cv error rate = 0.08466183788920778
when lambda = 0.14125375446227575 cv error rate = 0.0807254339157405
when lambda = 0.15848931924611173 cv error rate = 0.0761637289463376
when lambda = 0.17782794100389274 cv error rate = 0.07314225629763749
when lambda = 0.1995262314968885 cv error rate = 0.07617112622244482
when lambda = 0.2238721138568346 cv error rate = 0.07314036845113114
when lambda = 0.25118864315095873 cv error rate = 0.07510005586484547
when lambda = 0.2818382931264462 cv error rate = 0.07790273095618772
when lambda = 0.3162277660168389 cv error rate = 0.07551990907514783
when lambda = 0.3548133892335766 cv error rate = 0.08379717850423474

when lambda = 0.39810717055349853 cv error rate = 0.08225670217616055
when lambda = 0.44668359215096465 cv error rate = 0.06745973557306417
when lambda = 0.501187233627274 cv error rate = 0.07421181766228102
when lambda = 0.5623413251903511 cv error rate = 0.07355721651801483
when lambda = 0.6309573444801956 cv error rate = 0.07354215869468894
when lambda = 0.7079457843841406 cv error rate = 0.08097142545253733
when lambda = 0.7943282347242846 cv error rate = 0.07575514502385483
when lambda = 0.8912509381337491 cv error rate = 0.07638833773188725
when lambda = 1.000000000000004 cv error rate = 0.07745663411094628
when lambda = 1.1220184543019676 cv error rate = 0.07377269429086974
when lambda = 1.2589254117941726 cv error rate = 0.08029162091528441
when lambda = 1.4125375446227613 cv error rate = 0.07156469084908146
when lambda = 1.5848931924611207 cv error rate = 0.06963621711519075
when lambda = 1.77827941003893 cv error rate = 0.07290050920485702
when lambda = 1.995262314968889 cv error rate = 0.07485083444099838
when lambda = 2.2387211385683514 cv error rate = 0.07312242106682587
when lambda = 2.5118864315095926 cv error rate = 0.0785627580538486
when lambda = 2.8183829312644666 cv error rate = 0.0787751343645855
when lambda = 3.1622776601683955 cv error rate = 0.07486724714736115
when lambda = 3.548133892335775 cv error rate = 0.07571701566142042
when lambda = 3.9810717055349936 cv error rate = 0.07878517077305391
when lambda = 4.466835921509653 cv error rate = 0.07770778190877992
when lambda = 5.01187233627275 cv error rate = 0.0768263502276334
when lambda = 5.623413251903525 cv error rate = 0.08117563393757266
when lambda = 6.309573444801969 cv error rate = 0.08374858893105497
when lambda = 7.079457843841417 cv error rate = 0.08444631516762668
when lambda = 7.943282347242862 cv error rate = 0.08291574682308822
when lambda = 8.912509381337514 cv error rate = 0.0877069021979926
when lambda = 10.000000000000062 cv error rate = 0.08706529123564055
when lambda = 11.220184543019698 cv error rate = 0.08137665106303726
when lambda = 12.589254117941753 cv error rate = 0.09247800401970052
when lambda = 14.125375446227642 cv error rate = 0.08791251693603797
when lambda = 15.84893192461124 cv error rate = 0.08726892822972654
when lambda = 17.78279410038934 cv error rate = 0.09970119370974684
when lambda = 19.952623149688932 cv error rate = 0.09729223735496006
when lambda = 22.38721138568356 cv error rate = 0.089028991928493
when lambda = 25.118864315095976 cv error rate = 0.097932904394059
when lambda = 28.183829312644722 cv error rate = 0.10465305362383037
when lambda = 31.62277660168402 cv error rate = 0.10013601484592216
when lambda = 35.481338923357825 cv error rate = 0.10816638734243855
when lambda = 39.81071705535002 cv error rate = 0.10947856908940301

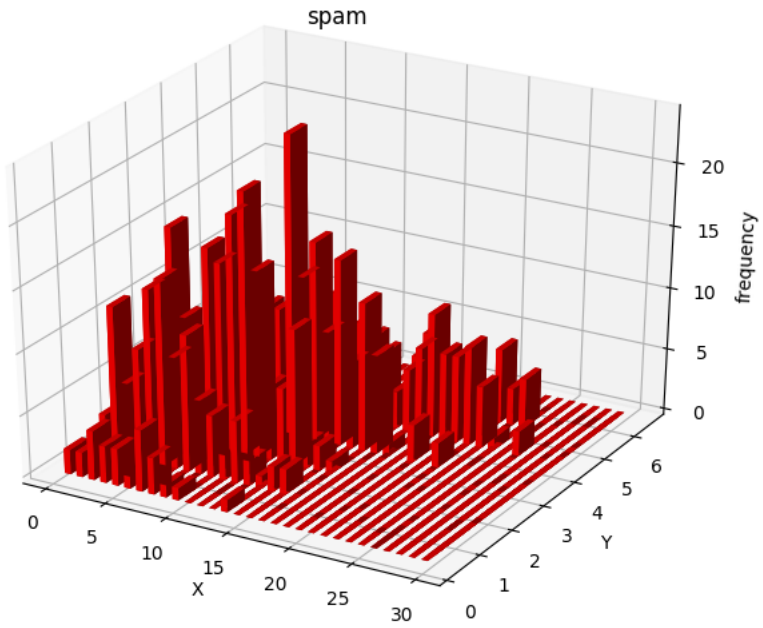
when lambda = 44.66835921509662 cv error rate = 0.11079215066813064
when lambda = 50.11872336272761 cv error rate = 0.11425414009875878
when lambda = 56.23413251903537 cv error rate = 0.11316448665343892
when lambda = 63.09573444801982 cv error rate = 0.11516484624324974
when lambda = 70.7945784384143 cv error rate = 0.11731159099227528
when lambda = 79.43282347242878 cv error rate = 0.11993332819633618
when lambda = 89.12509381337532 cv error rate = 0.12188033364797446
the best model is when lambda = 0.44668359215096465
cv error rate = 0.06745973557306417
test error = 0.07291666666666663
train error = 0.06394779771615011

2. Part B

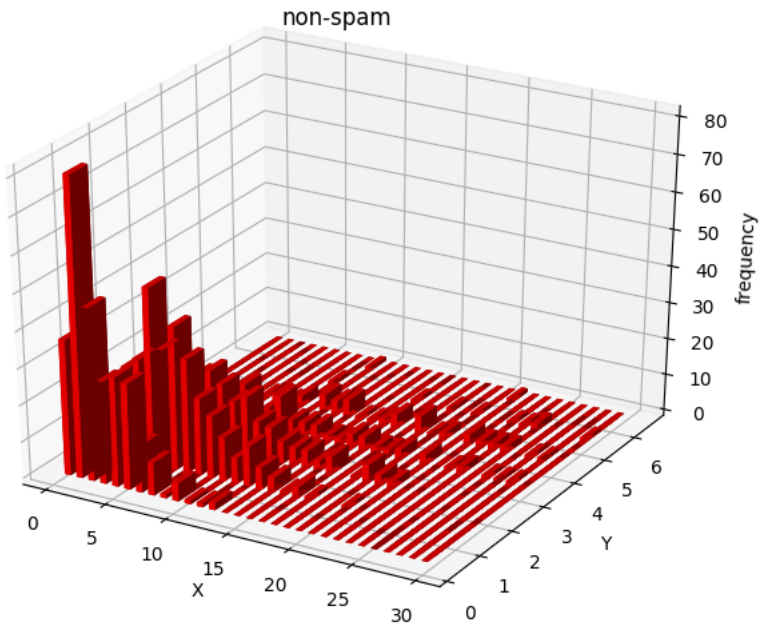
Question 1:



Question 2:



Question 3:



Question 4:

Compare these two-3D histograms, it is obviously that x value and y value of non-spam data set is smaller than spam' s, because there are more keywords and special characters for spam. So, non-spam histogram focuses more on (0, 0).

2. a

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

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

$$= \frac{1+e^{-a}}{(1+e^{-a})^2} - \frac{1}{(1+e^{-a})^2}$$

$$= \sigma(a) - \sigma(a)^2$$

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

b. this is Negative log likelihood.

$$NLL = -\sum_{i=1}^{N} [\, y_i \log u_i + (1-y_i) \log (1-u_i)]$$

$$\therefore \frac{\partial NLL}{\partial \vec{w}} = -\sum_{i=1}^{N} \left[y_i \cdot \frac{1}{u_i} \cdot \frac{\partial u_i}{\partial \vec{w}} + (1-y_i) \cdot \frac{1}{1-u_i} \cdot (-1) \cdot \frac{\partial u_i}{\partial \vec{w}} \right]$$

$$= -\sum_{1}^{N} \left[\frac{y_i}{u_i} \cdot \frac{\partial u_i}{\partial \vec{w}} - \frac{1-y_i}{1-u_i} \cdot \frac{\partial u_i}{\partial \vec{w}} \right]$$

$$= \sum_{1}^{N} \left[\frac{u_i - y_i}{u_i (1-u_i)} \cdot \frac{\partial u_i}{\partial \vec{w}} \right]$$

$\because \quad u_i = \frac{1}{1+e^{-w^T x}} \quad \Rightarrow \quad \frac{\partial u}{\partial w} = u_i(1-u_i) \cdot \frac{\partial}{w}(w^T x)$

$\therefore \quad \frac{\partial NLL}{\partial w} = \sum_1^N \frac{u_i - y_i}{u_i(1-u_i)} u_i(1-u_i) \cdot \frac{\partial}{\partial w}(w^T x_i)$

$$= \sum_1^N (u_i - y_i) \cdot x_i = x^T(\vec{u} - \vec{y})$$

c. $\because$ S is a diagonal matrix.

$\therefore \quad S = \sqrt{S}^T \cdot \sqrt{S}. \qquad \sqrt{S} = diag(\sqrt{u_i(1-u_i)})$

$\therefore$ for $\forall$ non zero vector v

$$v^T H v = v^T x^T \sqrt{S}^T \sqrt{S} x v = (\sqrt{S} x v)^T \cdot (\sqrt{S} x v) > 0$$

$\therefore$ H is positive definite

Which means that:

NLL is convex and it must have a minimum number.

3. (a). hypothesis set $\mathcal{H}$ denote all possible

$$\vec{w} = [w_0 \; w_1 \cdots w_j]$$

$$= [1, w_1 \cdots w_D]$$

$\because w_j = 1$ or 2 for $\forall j \in [1, \cdots D]$

$\therefore |\mathcal{H}| = 2^D$

(b)

$$\because P[|E_D(h) - E_{out}(h)| > \varepsilon] \leq 2e^{-2\varepsilon^2 N}$$

$\therefore$ for the best hypothesis $h_g \in \mathcal{H}$

$$P(|E_D(h_g) - E_{out}(h_g)| > \varepsilon)$$

$$\leq P(|E_D(h_i) - E_{out}(h_i)| > \varepsilon) \quad \text{for } \forall h_i \in \mathcal{H}$$

sum all:

$$2^D P\left(\left| E_D(h_g) - E_{out}(h_g) \right| > \epsilon \right)$$

$$\leq \sum_{i=1}^{2^D} P\left(\left| E_D(h_i) - E_{out}(h_i) \right| > \epsilon \right)$$

$$\leq 2^D \cdot 2e^{-2\epsilon^2 N}$$

$$\therefore P\left(\left| E_D(h_g) - E_{out}(h_g) \right| > \epsilon \right) \leq |\Phi| \cdot 2e^{-2\epsilon^2 N}$$

$$= 2^{D+1} e^{-2\epsilon^2 N}$$