CSCI 5525 Homework #2

Noah Hendrickson

Problem #1:

Noah Hendrickson

Problem #1:

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

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

$$= y_i \log\left(\frac{1}{1+\exp(w^T x_i)}\right) + (1-y_i)\log\left(\frac{1}{1+\exp(-w^T x_i)}\right)$$

(I'm assuming I dont have to mention log rules)

$$= -y_i \log(1+\exp(w^T x_i)) - (1-y_i)\log(1+\exp(-w^T x_i))$$

$$= -y_i \log(1+\exp(w^T x_i)) - \log(1+\exp(-w^T x_i)) + y_i \log(1+\exp(w^T x_i))$$

$$= -y_i \log\left(1+\frac{1}{\exp(w^T x_i)}\right) - \log(1+\exp(w^T x_i)) + y_i \log(1+\exp(w^T x_i))$$

$$= -y_i \log(\exp(w^T x_i)) - y_i \log(1+\exp(w^T x_i)) - \log(1+\exp(w^T x_i)) + y_i \log(1+\exp(w^T x_i))$$

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

now take derivative wrt w_j

$\log(x)' = \frac{x'}{x}$

$$\frac{\partial L}{\partial w_j} = \frac{\partial(-y_i w^T x_i)}{\partial w_j} + \frac{\partial(-\log(1+\exp(w^T x_i)))}{\partial w_j}$$

$$= -y_i x_j^i + \frac{x_j^i \exp(w^T x_i)}{1+\exp(w^T x_i)}$$

element$_j$ of w is only modifying element$_j$ of x so deriv results in rest going to 0

Problem #2:

Eta Vals	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Mean	Std
0.000001	0.0373	0.0438	0.0750	0.0125	0.0750	0.0688	0.0750	0.0625	0.0625	0.1132	0.0625	0.0256
0.00001	0.0186	0.0063	0.0188	0.0063	0.0188	0.0188	0.0188	0.0000	0.0188	0.0377	0.0625	0.0256
0.0001	0.0124	0.0063	0.0188	0.0063	0.0188	0.0125	0.0063	0.0000	0.0063	0.0314	0.0163	0.0098
0.001	0.0124	0.0188	0.0188	0.0063	0.0188	0.0125	0.0125	0.0000	0.0125	0.0314	0.0119	0.0086
0.01	0.0248	0.0125	0.0125	0.0063	0.0188	0.0125	0.0125	0.0000	0.0125	0.0377	0.0150	0.0098

The best eta value was 0.0001 and the resulting error from that run through the training and testing data was 0.005 using zero-one error.

Problem #3:

Problem #3: find grad wrt w_j (again)

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

$$= \frac{1}{2}\sum_k^n w_k^2 + C\sum_i^n \max(0, 1-y_i(w^Tx_i+b))$$

two cases:

hinge loss of 0

$= \frac{1}{2}\sum_k^n w_k^2 \rightarrow \frac{\partial}{\partial w_j} = w_j$ because all other w values will go to 0 and $(\frac{1}{2}w_j^2)' = w_j$

hinge loss of $1-y_i(w^Tx_i+b)$

$$= \frac{1}{2}\sum_k^n w_k^2 + C\sum_i^n(1-y_i(w^Tx_i+b))$$

$$\frac{\partial}{\partial w_j} = w_j - Cy_ix_j^i$$

thus, $\frac{\partial L}{\partial w_j} = w_j + \sum_i^n($

0 if hinge is 0,

$-Cy_ix_j^i$ if hinge is not 0

$)$

Problem #4:

Eta Vals	C Vals	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Mean	Stddev
0.00001	0.01	0.0186	0.1625	0.3813	0.0313	0.0938	0.0313	0.4688	0.0625	0.0813	0.4340	0.1765	0.1702
0.00001	0.1	0.0745	0.1313	0.1375	0.0500	0.0938	0.0750	0.1313	0.1000	0.0875	0.1761	0.1057	0.0358
0.00001	1	0.0124	0.0063	0.0188	0.0063	0.0188	0.0188	0.0125	0.0000	0.0125	0.0314	0.0138	0.0083
0.00001	10	0.0124	0.0063	0.0188	0.0063	0.0188	0.0125	0.0125	0.0000	0.0063	0.0314	0.0125	0.0084
0.00001	100	0.0124	0.0125	0.0188	0.0063	0.0125	0.0125	0.0125	0.0000	0.0125	0.0314	0.0131	0.0077
0.0001	0.01	0.4596	0.1250	0.5625	0.5563	0.4938	0.5688	0.5938	0.5188	0.1813	0.5723	0.4632	0.1602
0.0001	0.1	0.0186	0.0063	0.0313	0.0063	0.0188	0.0125	0.0438	0.0188	0.0125	0.0629	0.0232	0.0171
0.0001	1	0.0373	0.0313	0.0250	0.0188	0.0625	0.0688	0.0250	0.0188	0.0313	0.0377	0.0356	0.0163
0.0001	10	0.3292	0.0188	0.0188	0.1188	0.0188	0.0125	0.0188	0.4313	0.0125	0.0314	0.1011	0.1446
0.0001	100	0.0124	0.4938	0.3750	0.0063	0.0125	0.0125	0.0188	0.4375	0.0125	0.0377	0.1419	0.1941
0.001	0.01	0.5466	0.1438	0.5625	0.5563	0.5438	0.5688	0.5938	0.5188	0.5313	0.5723	0.5138	0.1250
0.001	0.1	0.5466	0.5125	0.4438	0.5563	0.5438	0.5688	0.4063	0.4813	0.5313	0.4277	0.5018	0.0553
0.001	1	0.5528	0.5250	0.4375	0.5563	0.5500	0.5688	0.4125	0.4938	0.5375	0.4403	0.5074	0.0546
0.001	10	0.5528	0.5250	0.4438	0.5563	0.5500	0.5688	0.4063	0.4938	0.5375	0.4403	0.5074	0.0550
0.001	100	0.5528	0.5250	0.4438	0.5563	0.5500	0.5688	0.4125	0.4938	0.5375	0.4403	0.5081	0.0538

The best eta and C value pair from this set was eta = 1e-5 and C = 10. Running a SVM with those values on the test and train data, we get an error rate of 0.005 using zero-one error.