

HOMEWORK 3 CMPS242 - FALL 2015

Ehsan Amid Andrei Ignat Sabina Tomkins

1 Problem 1

a) After saving the file in an aff file, we run the linear regression algorithm on the data and get the following values for the parameters:

$w = [-0.1343, 1.8477, -0.8966]^T$ and $b = 4.3608$ where $\hat{y} = w^T \mathbf{x} + b$

The root mean squared error on the training set is: 0.1897

b) The prediction for the new instance $\mathbf{x} = [3, 3, 5]$: $\hat{t} = 5.0180$

c) Now, we set the regularizer $\lambda = 0.2$ and run the first two parts again to get:
 $w = [-0.1527, 2.0598, -0.6439]^T$ and $b = 1.9483$

As can be seen, the weights become smaller, compared to (a), as we expected.

The root mean squared error on the training set becomes: 0.4614

d) By doing the calculations by hand, we find the following weights:

$w = [-0.1343, 1.8477, -0.8966]^T$ and $b = 4.3608$ and error = 0.1897, which is almost the same as part (a), because Weka uses a very small regularizer parameter by default ($1e-8$).

e) The least squared error solution does not depend on the order of the examples because what we are trying to minimize is the sum of the squared errors.

2 Problem Two

2.1 Part a

Select the Use training set test option and run the three classifiers. Report their results (accuracies). Which algorithm is best and why?

	Nearest Neighbor	Naive Bayes	Logistic Regression
Accuracy Training Set	100%	76.3021%	78.2552%
Accuracy 66% Split	72.7969 %	77.0115 %	80.0766 %

At first glance it may appear that 1 Nearest Neighbor is the best. However it's performance is so nice on the training set that one is agitated by a fear of the model having overfitted the data. Indeed that seems to be the case, when we perform a 66% split training, testing, we see that the 1 Nearest Neighbor is the least impressive of the lot.

2.2 Part b

3 Problem 4

Note that

$$\frac{d\sigma(a)}{da} = \frac{\exp(-a)}{(1 + \exp(-a))^2} = \sigma(-a)(1 - \sigma(-a))$$

and

$$\frac{dy_n}{d\mathbf{w}} = \sigma(-a)(1 - \sigma(-a))\mathbf{x}_n = y_n(1 - y_n)\mathbf{x}_n$$

Taking the derivative of $E(\mathbf{w})$ w.r.t. \mathbf{w} and substituting for these $\frac{dy_n}{d\mathbf{w}}$ from above, we have:

$$\nabla E(\mathbf{w}) = \sum_{n=1}^N \left(\frac{t_n}{y_n} - \frac{(1-t_n)}{(1-y_n)} \right) \frac{dy_n}{d\mathbf{w}} = \sum_{n=1}^N (t_n(1-y_n) - (1-t_n)y_n)\mathbf{x}_n = \sum_{n=1}^N (t_n - y_n)\mathbf{x}_n$$