CS221 Fall 2015 Homework 2

SUNet ID: lguan
Name: Leying Guan

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Problem 1

(a) The derivative will be $-\sum_{i=1}^{4} y_i\phi(x_i)\mathbb{I}(w_i^T x_i y_i < 1)$. Starting from $w^T = (0,0,0,0)$, the first derivative will be (-2,0,1,1,1,-1), the algorithm will step after one step: weights of pretty, bad, not, plot and scenery will appear in the model.

(b) I give the following example where cases can not be seperable using linear classifier in feaeture space {not, good, bad }: 1.(+1) Good 2.(-1)Not Good 3.(+1) Not Bad 4.(-1) Bad If we have perfect linear classifier, it means:

$$w_2 > 0 \quad (1)$$

$$w_1 + w_1 < 0 \quad (2)$$

$$w_3 < 0 \quad (3)$$

$$w_1 + w_3 > 0 \quad (4)$$

Equations (1),(2) indicate $w_1 > 0$ while Equations (3),(4) say $w_1 < 0$, it is a contradiction.

Thus we show no linear classifier can perfectly classify our given example.

Problem 2

(a) the loss Loss(x, y, w) = $(y - (1 + e^{-w^T\phi(x)})^{-1})^2$

(b) The gradient of the loss is as following:

$$\frac{dL}{dw} = -2(y - (1 + e^{-w^T\phi(x)})^{-1})\frac{e^{-w^T\phi(x)}\phi(x)}{(1 + e^{-w^T\phi(x)})^2}$$
$$= -2(y - \sigma(w^T\phi(x)))\sigma(w^T\phi(x))(1 - \sigma(w^T\phi(x))\phi(x)$$

(c) When $y = 0$, $\|Loss\| = 2\sigma(w^T\phi(x))\sigma(w^T\phi(x))(1-\sigma(w^T\phi(x))\|\phi(x)\|$. Obviously, when $\sigma(w^T\phi(x)) \to 0$ or $\sigma(w^T\phi(x)) \to 1$, we can have $\|Loss\| \to 0$, which correspond to $w^T\phi(x) \to -\infty$ and $w^T\phi(x) \to \infty$ respectively. Thus, any w whose magnitute goes ∞ and whose direction is not orthogonal to $\phi(x)$ will work.

(d) $\|Loss\| = 2\sigma(w^T\phi(x))\sigma(w^T\phi(x))(1 - \sigma(w^T\phi(x))\|\phi(x)\|$, the largest magnitute happens when $\sigma(w^T\phi(x)) = \frac{2}{3} \to e^{-w^T\phi(x)} = \frac{1}{2} \to w^T\phi(x) = \ln 2$, the largest magnitute $\|Loss\| = -\frac{8}{27}\|\phi(x)\|$

Problem 3

(a) Code

(b) Code

(c) Code

(d) Those wrongly predicted reviews are not straightforward – they are trying to be witty or sarcastic, for example, they will use a lot of positive words even though they do not like it, and thus a word-wise classifier is not supposed to capture their true meanings.

(e) Code

(f) If we let $n = 4$, test error decreases from 27.7% to 27.0%. The reason might be that $n = 4$ will enable use to capture most meaningful words while do not produce too many user-specific combinations such that our training sample size is still large enought to capture features which could be generalized to other data. Below is an example where n-grams is probably better:

The movie is not bad: the main character has terrifying and dark character while kind of pathetic at the same time, and the movie does not fail to convey his distorted mental stage.

The review consists a lot negative words, however, it is positive it self. Use character n-grams can capture it buy considering information from conseqtutive words.

Problem 4

(a) Starting with $\mu_1 = [-1, 0], \mu_2 = [3, 3]$:

Iter1: z = (1,1,1,2), $\mu_1 = [1, \frac{1}{3}], \mu_2 = [2, 2]$

Iter2: z = (1,1,1,2), $\mu_1 = [1, \frac{1}{3}], \mu_2 = [2, 2]$ - converge

Starting with $\mu_1 = [1, -1], \mu_2 = [0, 2]$:

Iter1: z = (1,2,1,2), $\mu_1 = [1, 0], \mu_2 = [1, 1.5]$

Iter2: z = (1,2,1,2), $\mu_1 = [1, 0], \mu_2 = [1, 1.5]$ - converge

(b) Code

(c) Code

(d) Those wrongly predicted reviews are not straightforward – they are trying to be witty or sarcastic, for example, they will use a lot of positive words even though they do not like it, and thus a word-wise classifier is not supposed to capture their true meanings.

(e) Code

(f) If we let $n = 4$, test error decreases from 27.7% to 27.0%. The reason might be that $n = 4$ will enable use to capture most meaningful words while do not produce too many user-specific combinations such that our training sample size is still large enought to capture features which could be generalized to other data. Below is an example where n-grams is probably better:

The movie is not bad: the main character has terrifying and dark character while kind of pathetic at the same time, and the movie does not fail to convey his distorted mental stage.

The review consists a lot negative words, however, it is positive it self. Use character n-grams can capture it buy considering information from consequtive words.