

Problem 1a.

	$y=+1$	$y=-1$	$y=-1$	$y=+1$
$w$	$\phi(x_1)$	$\phi(x_2)$	$\phi(x_3)$	$\phi(x_4)$
pretty : 0	pretty : 1	pretty : 0	pretty : 0	pretty : 1
good : 0	good : 1	good : 0	good : 1	good : 0
bad : 0	bad : 0	bad : 1	bad : 0	bad : 0
plot : 0	plot : 0	plot : 1	plot : 0	plot : 0
not : 0	not : 0	not : 0	not : 1	not : 0
scenery : 0	scenery : 0	scenery : 0	scenery : 0	scenery : 1

$$\nabla_w \text{Loss}_{\text{hinge}}(x, y, z) = \begin{cases} 0 & (w \cdot \phi(x)y \geq 1) \\ -\phi(x)y & (w \cdot \phi(x)y < 1) \end{cases}$$

$$\Rightarrow w = \begin{cases} w & (w \cdot \phi(x)y \geq 1) \\ w + \phi(x)y & (w \cdot \phi(x)y < 1) \end{cases}$$

① training on  $x_1$ 

$$w \cdot \phi(x)y = 0 \text{ or } y = +1 \text{ or } 2 \quad w = w + \phi(x) \text{이다.}$$

$$w = \{ \text{pretty: 1, good: 1, bad: 0, plot: 0, not: 0, scenery: 0} \}$$

② training on  $x_2$ 

$$w \cdot \phi(x)y = 0 \text{ or } y = -1 \text{ or } 3 \quad w = w - \phi(x) \text{이다.}$$

$$w = \{ \text{pretty: 1, good: 1, bad: -1, plot: -1, not: 0, scenery: 0} \}$$

③ training on  $x_3$ 

$$w \cdot \phi(x)y = -1 \text{ or } y = -1 \text{ or } 3 \quad w = w - \phi(x) \text{이다.}$$

$$w = \{ \text{pretty: 1, good: 0, bad: -1, plot: -1, not: -1, scenery: 0} \}$$

④ training on  $x_4$ 

$$w \cdot \phi(x)y = 1 \text{ or } y = 1 \text{ or } 3 \quad w = w \text{이다. (no update)}$$

⑤ training on  $x_1$ 

$$w \cdot \phi(x)y = 1 \text{ or } y = 1 \text{ or } 3 \quad w = w \text{이다. (no update)}$$

⑥ training on  $x_2$ 

$$w \cdot \phi(x)y = 2 \text{ or } y = 2 \text{ or } 3 \quad w = w \text{이다. (no update)}$$

① training on  $x_3$

$$W \cdot \phi(x) y = 1 \text{ 이므로 } W = W_0 \text{ (no update)}$$

③ training on  $x_4$

$$W \cdot \phi(x) y = 1 \text{ 이므로 } W = W_0 \text{ (no update)}$$

$$\therefore W = \{ \text{pretty: } 1, \text{ good: } 0, \text{ bad: } -1, \text{ plot: } -1, \text{ not: } -1, \text{ scenery: } 0 \}$$

Problem 1b.

$$(+) \text{ good} \rightarrow \phi(x) = \{ \text{not: } 0, \text{ good: } 1, \text{ bad: } 0 \}$$

$$(-1) \text{ bad} \rightarrow \phi(x) = \{ \text{not: } 0, \text{ good: } 0, \text{ bad: } 1 \}$$

$$(+1) \text{ not bad} \rightarrow \phi(x) = \{ \text{not: } 1, \text{ good: } 0, \text{ bad: } 1 \}$$

$$(-1) \text{ not good} \rightarrow \phi(x) = \{ \text{not: } 1, \text{ good: } 1, \text{ bad: } 0 \}$$

$$W = \{ \text{not: } a, \text{ good: } b, \text{ bad: } c \} \text{ 일때}$$

$$W \cdot \phi(x_1) = b > 0$$

$$W \cdot \phi(x_2) = c < 0$$

$$W \cdot \phi(x_3) = a + c > 0$$

$$W \cdot \phi(x_4) = a + b < 0$$

$$b > 0 \text{ 이므로 } a + b > a \text{ 이고, } c < 0 \text{ 이므로 } a + c < a \text{ 이다.}$$

따라서  $a + c < a < a + b$  을 만족한다.

하지만 이것은  $a + b < 0 < a + c$  이다.

이 features 를 사용하는

따라서,  $x_1, x_2, x_3, x_4$ 에 대해서 zero error 를 가지는 linear classifier 가 존재하지 않는다.

이 문제를 해결하기 위해서는  $x_1, x_3$  이라는 feature의 값이 0이지만

$x_2$  와  $x_4$ 에서는 feature의 값이 0인 feature를 만들면 된다.

예를 들면, 'A 감독이 만들었다.' 또는 '조조로 봤다.' 와 같은 feature가 가능하다.

Problem 2a.

$$\text{Loss squared } (x, y, w) = (\sigma(w \cdot \phi(x)) - y)^2$$

Problem 2b

$$\nabla_w \text{Loss squared } (x, y, w) = 2 \left( (1 + e^{-w \cdot \phi(x)})^{-1} - y \right) \cdot \frac{e^{-w \cdot \phi(x)}}{(1 + e^{-w \cdot \phi(x)})^2} \cdot \phi(x)$$

Problem 2c

$$y=0 \text{ or } \nabla_w \text{Loss } (x, y, w) = \frac{2 \cdot e^{-w \phi(x)} \cdot \phi(x)}{(1 + e^{-w \phi(x)})^3} \text{ of smallest magnitude}$$

$$\lim_{w \rightarrow \infty} \nabla_w \text{Loss } (x, y, w) = 0$$

$$\therefore 0$$

Problem 2d

$$y=1 \text{ or } \nabla_w \text{Loss } (x, y, w) = \frac{2 \cdot e^{-w \phi(x)} \cdot \phi(x)}{(1 + e^{-w \phi(x)})^3} \text{ of largest magnitude}$$

$$\frac{\partial}{\partial w} (\nabla_w \text{Loss } (x, y, w)) = - \frac{2 \cdot \phi(x)^2 \cdot e^{2w \phi(x)} (e^{w \phi(x)} - 2)}{(e^{w \phi(x)} + 1)^4} = 0$$

$$\Rightarrow e^{\phi(x) \cdot w} = 2$$

$$\Rightarrow w = \frac{\ln 2}{\|\phi(x)\|}$$

Problem 2e.

### Problem 3d

a perfectly competent and often imaginative film that lacks what little lilo & stitch had in spades -- charisma .

설명 : had의 weight가 -0.373로 매우 낮다.

→ perfectly competent 또는 imaginative film 과 같은 구절은 positive를 확실하게 만드는 구절이므로 연달아 존재하는 단어 두 개를 하나의 feature로 고려한다면 error를 해결할 수 있을 것이다.

또한, have와 같이 positive와 negative로 여겨지던 것이 확실한 단어와 a 와 the 같은 경우도 제외하는 것을 좋을 것 같다.

### Problem 3f.

polarity.train on training을 하고, polarity.dev로 test를 한다.

Word feature를 사용했을 때보다 4~6 길이의 Character feature를 사용했을 때

error가 적거나 거의 비슷하게 작았다. 가장 test error가 적을 때는 5 길이의 character feature를 사용했을 때이다.

Character feature가 Word feature를 outperform할 수 있는 review는 예를 들면 'It is good before his acting.'

good이라는 긍정적인 단어가 나오지만 good before를 보면 이 리뷰가 부정적인 리뷰라는 것을 알 수 있다.

즉, 문맥의 흐름 또는 형용사나 부사의 사용이 단어의 뉘앙스를 바꿀 수 있으므로 단어 + 단어를 꾸며주는 부사 또는 형용사를 하나의 feature로 볼 때 정확성이 높아진다.

### Problem 4a.

$$1. \mu_1 = [-1, 0] \text{ and } \mu_2 = [3, 2]$$

$$\begin{aligned} \text{dist}(\emptyset(x_1), \mu_1) &= 1, \quad \text{dist}(\emptyset(x_1), \mu_2) = 13 & \because x_1 \rightarrow \mu_1 \\ \text{dist}(\emptyset(x_2), \mu_1) &= 2, \quad \text{dist}(\emptyset(x_2), \mu_2) = 10 & \because x_2 \rightarrow \mu_1 \\ \text{dist}(\emptyset(x_3), \mu_1) &= 9, \quad \text{dist}(\emptyset(x_3), \mu_2) = 5 & \because x_3 \rightarrow \mu_2 \\ \text{dist}(\emptyset(x_4), \mu_1) &= 13, \quad \text{dist}(\emptyset(x_4), \mu_2) = 1 & \because x_4 \rightarrow \mu_2 \\ \text{dist}(\emptyset(x_1), \mu_1) &= 0.25, \quad \text{dist}(\emptyset(x_1), \mu_2) = 5 & \because x_1 \rightarrow \mu_1 \\ \text{dist}(\emptyset(x_2), \mu_1) &= 0.25, \quad \text{dist}(\emptyset(x_2), \mu_2) = 4 & \because x_2 \rightarrow \mu_1 \\ \text{dist}(\emptyset(x_3), \mu_1) &= 4.25, \quad \text{dist}(\emptyset(x_3), \mu_2) = 1 & \because x_3 \rightarrow \mu_2 \\ \text{dist}(\emptyset(x_4), \mu_1) &= 6.25, \quad \text{dist}(\emptyset(x_4), \mu_2) = 1 & \because x_4 \rightarrow \mu_2 \end{aligned} \quad \left. \begin{array}{l} \mu_1 = [0, 0.5] \\ \mu_2 = [2, 1] \end{array} \right\}$$

$$2. \mu = [1, -1], \mu_2 = [0, 2]$$

$$\begin{aligned} \text{dist}(\emptyset(x_1), \mu_1) &= 2, \quad \text{dist}(\emptyset(x_1), \mu_2) = 4 & \because x_1 \rightarrow \mu_1 \\ \text{dist}(\emptyset(x_2), \mu_1) &= 5, \quad \text{dist}(\emptyset(x_2), \mu_2) = 1 & \because x_2 \rightarrow \mu_2 \\ \text{dist}(\emptyset(x_3), \mu_1) &= 2, \quad \text{dist}(\emptyset(x_3), \mu_2) = 8 & \because x_3 \rightarrow \mu_1 \\ \text{dist}(\emptyset(x_4), \mu_1) &= 10, \quad \text{dist}(\emptyset(x_4), \mu_2) = 4 & \because x_4 \rightarrow \mu_2 \\ \text{dist}(\emptyset(x_1), \mu_1) &= 1, \quad \text{dist}(\emptyset(x_1), \mu_2) = 3.25 & \because x_1 \rightarrow \mu_1 \\ \text{dist}(\emptyset(x_2), \mu_1) &= 2, \quad \text{dist}(\emptyset(x_2), \mu_2) = 1.25 & \because x_2 \rightarrow \mu_2 \\ \text{dist}(\emptyset(x_3), \mu_1) &= 1, \quad \text{dist}(\emptyset(x_3), \mu_2) = 3.25 & \because x_3 \rightarrow \mu_1 \\ \text{dist}(\emptyset(x_4), \mu_1) &= 5, \quad \text{dist}(\emptyset(x_4), \mu_2) = 1.25 & \because x_4 \rightarrow \mu_2 \end{aligned} \quad \left. \begin{array}{l} \mu_1 = [1, 0] \\ \mu_2 = [1, 1.5] \\ \mu_1 = [1, 0] \\ \mu_2 = [1, 1.5] \end{array} \right\}$$

## Problem 4c.

1. 미리 주어진 각각의 cluster에서 average를 구하여 center를 pick한다.

이때, cluster에 있던 points를 바로 center에 assign한다.

만약 미리 주어진 cluster의 개수가 K보다 작으면, cluster에 속한 example을 제외한 subset에서 random point를 ( $K -$  주어진 cluster의 개수) 만큼 pick하여 center로 지정한다.

## 2. alternate part

① 이미 assign된 것들을 제외하고 examples를 closest mean에 assign한다.

② assigned points의 평균으로 center를 update한다.

이때, 미리 clustering에 대한 정보가 주어졌던 points는 자신이 속한 cluster의 평균으로 바뀐 center에 바로 assign한다.