CS 416 HW3

Xianglu Peng

930990871

xpeng

1. $K(x,z) = (x^T z + 1)^3$

$$\phi(x) = \begin{bmatrix} 1 \\ \sqrt{3}x_1 \\ \sqrt{3}x_2 \\ \cdots \\ \sqrt{3}x_d \\ \sqrt{3}x_1x_1 \\ \sqrt{3}x_1x_2 \\ \cdots \\ \sqrt{3}x_1x_d \\ \cdots \\ \sqrt{3}x_2x_1 \\ \cdots \\ \sqrt{3}x_2x_d \\ \cdots \\ \sqrt{3}x_dx_d \\ x_1x_1x_1 \\ x_1x_1x_2 \\ \cdots \\ x_2x_1x_1 \\ \cdots \\ x_dx_1x_1 \\ \cdots \\ x_dx_dx_d \end{bmatrix}$$

$$K(x,z) = (x^T z+1)^3 = \left(\sum_{i=1}^{d} x_i z_i + 1\right)\left(\sum_{j=1}^{d} x_j z_j + 1\right)\left(\sum_{k=1}^{d} x_k z_k + 1\right)$$

$$= \left(\sum_{i=1}^{d}\sum_{j=1}^{d} x_i z_i x_j z_j + 2\left(\sum_{i=1}^{d} x_i z_i\right) + 1\right)\left(\sum_{k=1}^{d} x_k z_k + 1\right)$$

$$= \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} x_i x_j x_k z_i z_j z_k + 3\sum_{i=1}^{d}\sum_{j=1}^{d} x_i x_j z_i z_j + 3\sum_{i=1}^{d} x_i z_i + 1$$

$$= \sum_{i,j,k}^{d} (x_i x_j x_k)(z_i z_j z_k) + \sum_{i,j}^{d} (\sqrt{3} x_i z_i)(\sqrt{3} x_j z_j) + \sum_{i=1}^{d} (\sqrt{3} x_i)(\sqrt{3} z_i) + 1 \cdot 1$$

$$= \langle \phi(x), \phi(z) \rangle$$

Running time for computing kernel function: $O(n)$.

Since if using kernel trick, we can implement a non-linear feature expansion at no additional cost.

Running time for computing inner product: $O(n^3)$

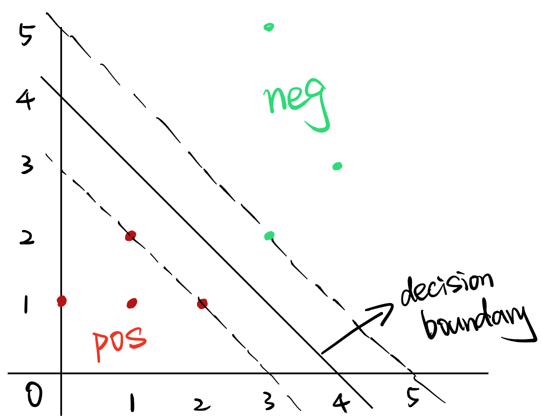
Because if we compute the inner product directly, the highest term we need to compute has the power of 3. Therefore, the time complexity is $O(n^3)$.

2. SVM

1. $\min \quad 0.5 w^T w = 0.5 (a_1^2 + a_2^2 + a_3^2)$

$$1(a_1 + 2a_2 + 3a_3 + b \geq 1$$
$$1(a_1 + a_2 + 4a_3 + b) \geq 1$$
$$(-1)(3a_1 + 2a_2 - a_3 + b) \geq 1$$
$$(-1)(4a_1 + 3a_2 - 2a_3 + b) \geq 1$$
$$(-1)(3a_1 + 5a_2 - 3a_3 + b) \geq 1$$

2. Support Vector: (2,1), (3,2), (1,2)



street: $y = -x+3$

$y = -x+5$

$\Rightarrow x_1 + x_2 - 3 = 0$

$x_1 + x_2 - 5 = 0$

$\Rightarrow$ decision boundary

$x_1 + x_2 - 4 = 0.$

3. Decision Tree.

- Entropy

$S = \{ -, +, +, -, + \}.$

$E(S) = -\frac{3}{5}\log(\frac{3}{5}) - \frac{2}{5}\log(\frac{2}{5}) = 0.971$

x_1

$+$ $-$

$2+,$ $1+, 2-$

$E(S_1) = 0$

$E(S_2) = -\frac{1}{3}\log(\frac{1}{3}) - \frac{2}{3}\log(\frac{2}{3})$

$= 0.918$

$\frac{2}{5}\times 0 + \frac{3}{5}\times 0.918 = 0.551$

Gain: $0.971 - 0.551 = 0.420$

x_2

$+$ $-$

$3+, 1-$ $1-$

$E(S_1) = -\frac{3}{4}\log(\frac{3}{4}) - \frac{1}{4}\log(\frac{1}{4})$

$= 0.811$

$E(S_2) = 0$

$0.881 \times \frac{4}{5} + \frac{1}{5}\times 0 = 0.649$

Gain $= 0.971 - 0.649 = 0.322$

x_3

$+$ $-$

$1+, 2-$ $2+$

$E(S_1) = -\frac{1}{3}\log(\frac{1}{3}) - \frac{2}{3}\log(\frac{2}{3})$

$= 0.918$

$E(S_2) = 0.$

$0.918 \times \frac{3}{5} + 0 \times \frac{2}{5} = 0.551$

Gain $= 0.971 - 0.551 = 0.420$

X_1

$+$ $-$

$\langle +,+,+\rangle : +$

$\langle +,+,-\rangle : +$

pos

$\langle -,+,+\rangle : -$

$\langle -,-,+\rangle : -$

$\langle -,+,-\rangle : +$

$S = \{-,-,+\}.$

$E(S) = -\frac{1}{3}\log(\frac{1}{3}) - \frac{2}{3}\log(\frac{2}{3}) = 0.918$

x_2

$+$ $-$

$1+, 1-$ $1-$

$E(S_1) = 1$

$E(S_2) = 0.$

$\frac{2}{3}\times 1 + 0 \times \frac{1}{3} = \frac{2}{3}.$

Gain $= 0.918 - \frac{2}{3} = 0.251$

x_3

$+$ $-$

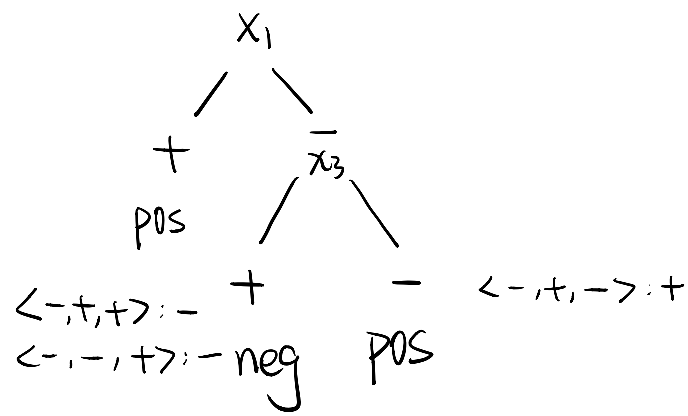
$2-$ $1+$

$E(S_1) = 0$

$E(S_2) = 0$

$\frac{2}{3}\times 0 + \frac{1}{3}\times 0 = 0$

Gain $= 0.918 - 0 = 0.918$



prediction: $<+,-,->$: +

$<-,-,->$: +

$<+,-,+>$: +

- Adaboost

$<-1,1,1>$: -1
$<1,1,1>$: 1
$<-1,1,-1>$: 1
$<-1,-1,1>$: -1
$<1,1,-1>$: 1

Round 1:

weight $= \frac{1}{n} = \frac{1}{5}$

if choose $x_1 = y$, $\varepsilon = 0.2$

if choose $x_2 = y$, $\varepsilon = 0.2$

if choose $x_3 = -y$, $\varepsilon = 0.2$

Choose $h_1 = (x_1 = +1)$:

3 is classified incorrectly

$\beta_1 = \frac{1-0.2}{0.2} = \frac{0.8}{0.2} = 4$

$\alpha_1 = \ln(4) = 1.386$

1	2	3	4	5
1/5	1/5	1/5	1/5	1/5
1/5	1/5	4/5	1/5	1/5
1/8	1/8	1/2	1/8	1/8

Round 2:

if choose $x_2 = y$, $\varepsilon = \frac{1}{8}$

if choose $x_3 = -y$, $\varepsilon = \frac{1}{8}$

Choose $h_2 = (x_2 = +1)$

① is classified incorrectly

$$\beta_2 = \frac{1-\frac{1}{8}}{\frac{1}{8}} = \frac{7/8}{1/8} = 7$$

$$\alpha_2 = \ln(7) = 1.946$$

1	2	3	4	5
1/8	1/8	1/2	1/8	1/8
7/8	1/8	1/2	1/8	1/8
1/2	1/14	2/7	1/14	1/14

Round 3:

$h_3 = (x_3 = -1)$

② is classified incorrectly.

$\varepsilon_3 = \frac{1}{14}$

$$\beta_3 = \frac{1-\frac{1}{14}}{\frac{1}{14}} = 13$$

$$\alpha_3 = \ln(13) = 2.565$$

1	2	3	4	5
1/2	1/14	2/7	1/14	1/14
1/2	13/14	2/7	1/14	1/14
7/26	1/2	2/13	1/26	1/26

$$H(x) = 1.386\,(x_1 = +1) + 1.946\,(x_2 = +1) + 2.565\,(x_3 = -1)$$

prediction:

<+, -, ->

$1.386 - 1.946 + 2.565 > 0 \Rightarrow +$

$\langle -, -, - \rangle$

$-1.386 - 1.946 + 2.565 < 0 \Rightarrow -$

$\langle +, -, + \rangle$

$1.386 - 1.946 - 2.565 < 0 \Rightarrow -$

4. Write up for 1.3 & 1.4

For the polynomial kernel, when I increase C , the decision boundary of the brown part would be curvy and the number of correctly classified points increased. When I increase d, the decision boundaries would be more curvy and better separate the data. If d becomes too large, the model is overfitting.

For the Gaussian kernel, when I increase C, the number of correctly classified points also increased and the area of the decision surface also changes. When I increase sigma, the decision boundaries become more linear and there are less misclassification.