HomeWork 4

1. (a) False. Random Forest average the Var
So it has smaller var

(b) False. More weak classifier makes Adaboost focus more efficiently.

(c) False. Variance is lowered, not bias

2. (a) $\epsilon_t = \sum_{i: y_i \neq h_t(x_i)} W_{t,i}$ $\quad 1 - \epsilon_t = \sum_{i: y_i = h_t(x_i)} W_{t,i}$

given binary classifier (weak), we get -1 or 1 from sign func

thus: if $y_i = h_t(x_i)$

$$W_{t+1} = W_t e^{-\beta} / Z_t$$

else $y_i \neq h_t(x_i)$

$$W_{t+1} = W_t e^{\beta} / Z_t$$

$$Z_t = \sum_{i=1}^{N} W_{t,i} \times e^{-\beta_t y_i h_t(x_i)}$$

$$= \sum_{\substack{i=1 \\ y_i = h_t(x_i)}}^{N} W_{t,i} e^{-\beta_t} + \sum_{\substack{i=1 \\ y_i \neq h_t(x_i)}}^{N} W_{t,i} e^{\beta_t}$$

$$= e^{-\beta_t}(1 - \epsilon_t) + e^{\beta_t} \epsilon_t$$

(b) $$\frac{\partial Z}{\partial \beta} = \epsilon_t e^{\beta} - (1 - \epsilon_t) e^{-\beta} = 0$$

$$\epsilon_t e^{2\beta} = (1 - \epsilon_t)$$

$$e^{2\beta} = \frac{1 - \epsilon_t}{\epsilon_t}$$

$$\beta = \frac{1}{2} \ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$$

(c) β is changed for each iteration to minimize the average exponential loss.

While the function is convex, the min point sit at where the gradient equals zero

3. (a) Calculate the mean of each x_i and center the matrix

$$\bar{x}_1 = \frac{4+2+5+1}{4} = \frac{12}{4} = 3$$

$$\bar{x}_2 = \frac{1+3+4+0}{4} = 2$$

$$X_c = \begin{bmatrix} 4-3 & 1-2 \\ 2-3 & 3-2 \\ 5-3 & 4-2 \\ 1-3 & 0-2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & +1 \\ +2 & 2 \\ -2 & -2 \end{bmatrix}$$

get the cov matrix:

$$Cov = X_c^T \cdot X = \begin{bmatrix} 1 & -1 & +2 & -2 \\ -1 & +1 & 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ -1 & +1 \\ +2 & 2 \\ -2 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix} \qquad \cancel{= \begin{bmatrix} 12 & 0 \\ 0 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}$$

$$= \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} \qquad \cancel{= \begin{bmatrix} 10 & -2 \\ -2 & 10 \end{bmatrix} = \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}}$$

$\det(C - \lambda I)v = \cancel{\det(Cov - \lambda I)v = 0}$

$(5-\lambda)^2 = 3^2$ $\quad \cancel{(5-\lambda)^2 = 2^2}$

$\lambda = 2$ or 8 $\quad \cancel{\lambda = 3 \text{ or } 7}$

~~for $\lambda = 7$:~~

$\cancel{\begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0}$

$\cancel{-2x_1 - x_2 = 0}$

$\cancel{-x_1 - 2x_2 = 0}$

$\lambda = 8$

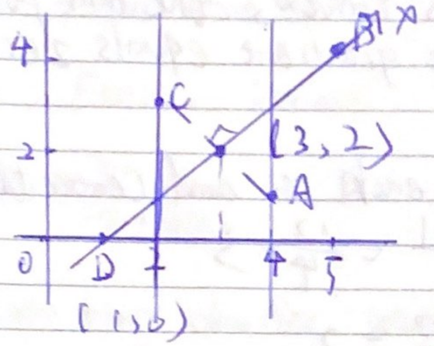
$$\begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

$x = y$

3 (a) We use $\lambda = 8$ since it is bigger => high variation

we get $v = (\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}})$

(b)

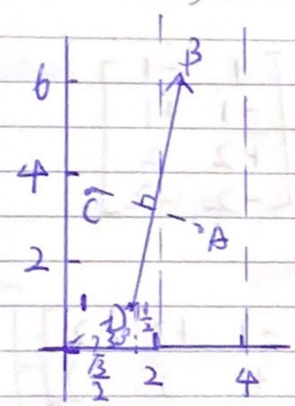


A C Project to 0 (center)

D project to $-2\sqrt{2}$

B -> $2\sqrt{2}$

(c)



A C ~ 0

D ~ $-2\sqrt{2}$

B ~ $2\sqrt{2}$

$x_{new} = \frac{1}{2} x$

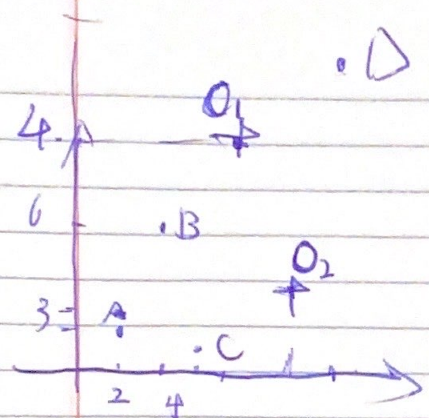
$y_{new} = \frac{\sqrt{3}}{2} y$

new $(x_B - x_D)^2 + (y_B - y_D)^2$

$= \frac{1}{4}(x_B - x_D)^2 + \frac{3}{4}(y_B - y_D)^2$

$= \frac{1}{4} \times 4^2 + \frac{3}{4} \times 4^2 = 2 \times 4^2$

= old length



k=1

$O_1 = (6,9)$ $O_2 = (8,4)$

$d(A,1) = 7.2$ $d(A,2) = 6.1$
$d(B,1) = 3.6$ $d(B,2) = 4.8$
$d(C,1) = 8.1$ $d(C,2) = 4.2$
$d(D,1) = 5$ $d(B,2) = 8.2$

$O_1 \{B, D\}$ $O_2 \{A, C\}$

$O_1 = (B+D)/2$ $O_2 = (A+C)/2$

$= (7,9)$ $= (3.5, 2)$

k=2

$d(A,1) = 7.8$ $d(A,2) = 1.8$
$d(B,1) = 4.2$ $d(B,2) = 4$
$d(C,1) = 8.2$ $d(C,2) = 1.8$
$d(D,1) = 4.2$ $d(D,2) = 11.9$

$O_1 \{D\}$ $O_2 (A, B, C)$

$O_1 = (10, 12)$ $O_2 = (3.6, 3.3)$