Michigan State University **April 17, 2024**
CMSE381 - Data Science

CMSE381 - Practice problems for the final exam

1. (SVM) For the following dataset, if I was given the information that the classification hyperplane has a slope of -1. (a) Could you help to figure out the equation of the MMC?

$4 - x_1 - x_2 = 0 \Rightarrow \frac{4}{\sqrt{2}} - \frac{x_1}{\sqrt{2}} - \frac{x_2}{\sqrt{2}} = 0$ MMC

$\beta_1 = -1$ $\beta_2 = -1$ $\|\beta\|_2 = \sqrt{1^2 + 1^2} = \sqrt{2}$

(b) what should the label be for the red class?

$y \cdot \left(\frac{4}{\sqrt{2}} - \frac{x_1}{\sqrt{2}} - \frac{x_2}{\sqrt{2}}\right) \geq M > 0$ $y\left(-\frac{3}{\sqrt{2}}\right) > 0$

$\Rightarrow y = -1$

(c) Please use the equation to find M, the value of the margin.

M = distance between $(0,3)$ and the center

$= \frac{4}{\sqrt{2}} - \frac{0}{\sqrt{2}} - \frac{3}{\sqrt{2}} = \frac{1}{\sqrt{2}}$

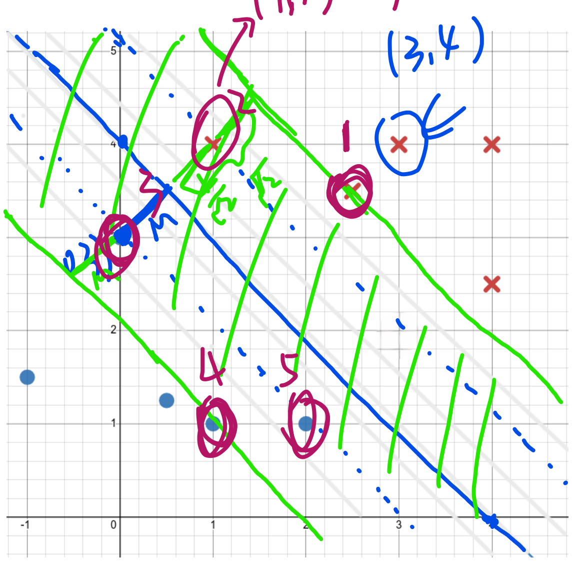
(d) What is the main difference between MMC and SVC?

soft margin

(e) If I want to let $M = \sqrt{2}$. Then how many support vectors would I have?

5

(f) For each of the support vectors, what is the smallest value for the associated slack variable?



$M = \sqrt{2}$

$\varepsilon_1 = \varepsilon_4 = 0$

$\varepsilon_2 = \frac{1}{2} = \varepsilon_3 = \varepsilon_5$

$y \cdot \left(\frac{4}{\sqrt{2}} - \frac{x_1}{\sqrt{2}} - \frac{x_2}{\sqrt{2}}\right) \geq \sqrt{2}(1 - \varepsilon)$

$-1 \cdot \left(-\frac{1}{\sqrt{2}}\right) \geq \sqrt{2} \cdot (1 - \varepsilon_2)$

$\varepsilon_2 \geq 1/2$

(g) Find the inner product representation of the SVC

$$\frac{4}{\sqrt{2}} - \frac{x_1}{\sqrt{2}} - \frac{x_2}{\sqrt{2}} = 0$$

$$\|$$

$$f(x) = \beta_0 + \sum_{i=1}^{2} \alpha_i \langle x, x_i \rangle$$

$$f(x) = \frac{4}{\sqrt{2}} + \alpha_1 \langle x, \begin{pmatrix} 0 \\ 3 \end{pmatrix} \rangle + \alpha_2 \langle x, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle$$

$$f(x) = \frac{4}{\sqrt{2}} + \alpha_1 (0 \cdot x_1 + 3 \cdot x_2) + \alpha_2 (x_1 + x_2)$$

$$f(x) = \frac{4}{\sqrt{2}} + \alpha_2 x_1 + (3\alpha_1 + \alpha_2) x_2$$

$$\alpha_2 = -1/\sqrt{2} \quad \alpha_1 = 0$$

(h) Using the kernel $K(x, y) = e^{-\frac{\|x-y\|_2^2}{2}}$ to replace the inner product found in the previous problem, what is prediction of SVM for the new input $x = [1, 5]$?

$$\rightarrow f(x) = \frac{4}{\sqrt{2}} - \frac{1}{\sqrt{2}} \cdot \underbrace{\langle x, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle}_{K(x, \begin{pmatrix} 1 \\ 1 \end{pmatrix})}$$

$$\left\| \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \right\|_2 = \sqrt{x_1^2 + x_2^2 \cdots x_p^2}$$

$$\rightarrow f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{4}{\sqrt{2}} - \frac{1}{\sqrt{2}} \cdot e^{-\frac{\left\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|_2^2}{2}}$$

$$f\left(\begin{bmatrix} 1 \\ 5 \end{bmatrix}\right) = \frac{4}{\sqrt{2}} - \frac{1}{\sqrt{2}} e^{-\frac{\left\| \begin{bmatrix} 1 \\ 5 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|_2^2}{2}}$$

$$(\beta_0 + \beta_1 x_1 \cdots + \beta_p x_p)$$

$$y \cdot f(x) \geq M > 0$$

$$= \frac{4}{\sqrt{2}} - \frac{1}{\sqrt{2}} e^{-8} > 0$$

$$\therefore y = 1$$

2. (Decision Tree for classification) (a) which of the following classification tree is preferred during pruning with an $\alpha = 0.1$?
Tree 1: three leaves (L_1,L_2,L_3), training samples classified to L_1 has labels $(-1, 1, 1)$, those classified to L_2 has labels $(-1, 1)$,those classified to L_3 has labels $(1, 1)$
Tree 2: two leaves (L_1,L_2), training samples classified to L_1 has labels $(-1, 1, 1, 1)$, those classified to L_2 has labels $(-1, 1, 1)$.

$P_{-1} = 1/4 \quad P_1 = 3/4$

$$obj = RSS + \alpha |T|$$
$$= \text{Gini index} + \alpha |T|$$

$$G = \sum_{i=1}^{K} p_i (1 - p_i)$$

p_i : prob of falling into class i

K : no. of classes

Tree 1 : L_1: $P_{-1} = 1/3, \quad P_1 = 2/3$

$$G_1 = P_{-1}(1 - P_{-1}) + P_1(1 - P_1) = \frac{1}{3} \cdot \frac{2}{3} + \frac{2}{3} \cdot \frac{1}{3} = 4/9$$

L_2: $P_{-1} = 1/2 \quad P_1 = 1/2$

$$G_2 = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

L_3 : $G_3 = 0$

$$obj_1 = G_1 + G_2 + G_3 + \alpha |T| = 4/9 + \frac{1}{2} + 0 + 0.1 \cdot 3$$

Tree 2

$$G_1 = \frac{1}{4} \cdot \frac{3}{4} + \frac{3}{4} \cdot \frac{1}{4} = 3/8$$

$$G_2 = 4/9$$

$$obj_2 = G_1 + G_2 + \alpha |T| = 3/8 + 4/9 + 0.1 \cdot 2$$

$$obj_1 - \underline{obj_2} = \frac{1}{2} + 0.1 - 3/8 > 0$$

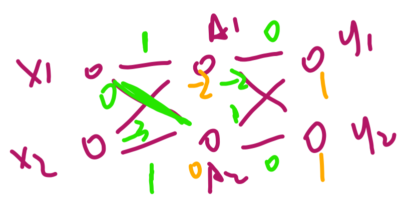
smaller

3. (Neural Network) (a) Draw the diagram for a neural net with input data points with $p = 2$ (i.e., (X_1, X_2)), two units in the hidden layer and two units in the output layer with the following β and $\beta^{(2)}$ matrices as weights for the first and second layer (last columns are the bias, β_{ij} is the weight associated with X_i and A_j),

$$\beta = \begin{array}{c} \quad\; \beta_{12} \quad \\ \begin{pmatrix} 1 & 0 & -2 \\ -3 & 1 & 0 \end{pmatrix} \end{array} \qquad \beta^{(2)} = \begin{pmatrix} 0 & -2 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

and using the activation function

$$g(z) = (z)_+ = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{else.} \end{cases}$$



(b) What is the output Y for the new point $(2, 1)$?

$A_1 = g(2 \cdot 1 + (-3) \cdot 1 + -2) = 0$

$A_2 = g(2 \cdot 0 + 1 \cdot 1 + 0) = 1$

$y_1 = 0 \cdot A_1 + 1 \cdot A_2 + 1 = 2$

$y_2 = 1$

(c) Add a softmax layer to the end of the network, what is the predicted Y value for the new point $(2, 1)$ now?

$$\begin{pmatrix}2\\1\end{pmatrix} \xrightarrow{\text{softmax}} \begin{pmatrix}\dfrac{e^2}{e^2+e^1}\\[2ex] \dfrac{e^1}{e^2+e^1}\end{pmatrix}$$

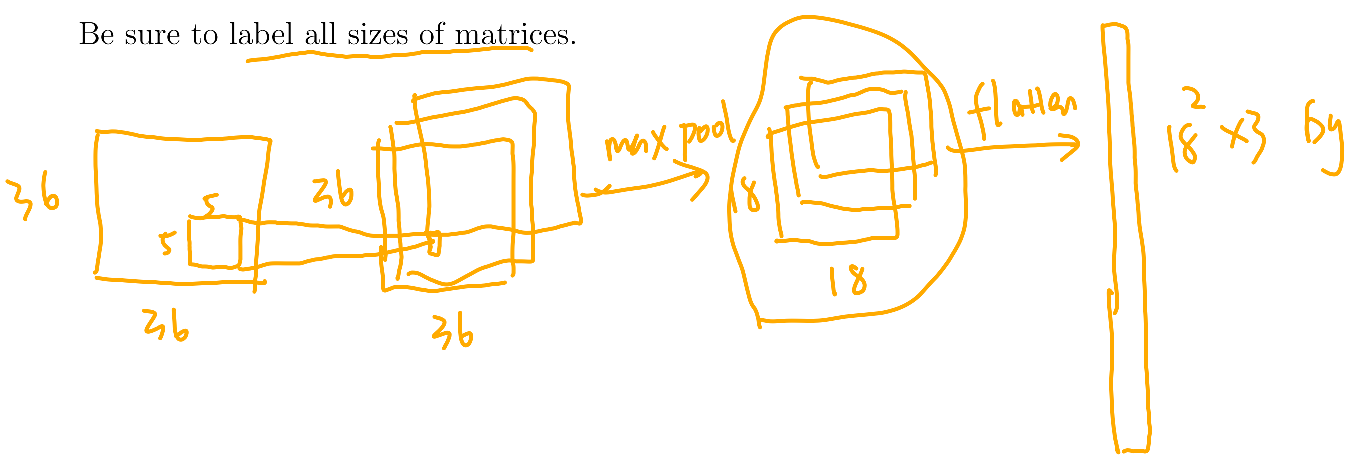
(d) How many parameters are trained by the computer in this setup?

12 param

4. (Convolutional Neural Network) (a) Draw a sketch of a CNN which

- takes as input a 36×36 black and white image,
- applies a convolution layer with three filter matrices of size 5×5,
- applies a 2×2 max pool layer,
- flattens the result into a vector.

Be sure to label all sizes of matrices.



5. For the following input "image" matrix X, we convolve with the matrix F resulting in the matrix A.

$$X = \begin{bmatrix} 6 & 5 & -1 & 5 & 4 \\ -6 & 2 & 4 & 0 & 5 \\ 5 & 2 & 1 & 4 & 3 \\ 3 & 2 & -2 & -1 & -6 \\ 5 & 0 & 2 & 2 & 1 \end{bmatrix}, \qquad F = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \qquad A = \begin{bmatrix} 18. & 7. & 9. & 18. \\ 0. & 11. & 8. & 13. \\ 11. & 2. & ???\ 8 & 4. \\ 7. & 0. & -2. & -12. \end{bmatrix}$$

(a) What value goes in the missing spot in matrix A?

$$\left\langle \begin{pmatrix} 1 & 4 \\ -2 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \right\rangle = 1\cdot 1 + 4\cdot 2 + (-2)\cdot 0 + (-1)\cdot 1$$

$$= 8$$

(b) If we apply a 2x2 max pooling layer to the matrix A, what would the resulting matrix be?

$$\begin{bmatrix} 18 & 18 \\ 11 & 8 \end{bmatrix}$$

(c) If we apply the ReLU function to every entry in the A matrix, what would the resulting matrix be?

$$\begin{bmatrix} 11 & 18 \\ 11 & 8 \end{bmatrix}$$