Name: Yuxuan Zhang yuxuanz8

University of Illinois Spring 2020

CS 446/ECE 449 Machine Learning
Homework 3: Support Vector Machine (SVM)

Due on Thursday February 20 2020, noon Central Time

1. **[30 points]** Max-Margin Support Vector Machine

We are given a dataset $\mathcal{D} = \left\{ \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, 1 \right), \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}, 1 \right), \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, -1 \right), \left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}, -1 \right) \right\}$ containing four pairs (x, y), where each $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$ denotes a 2-dimensional point and $y \in \{-1, +1\}$.

We want to train the parameters w and the bias b of a max-margin support vector machine (SVM) using (with hyperparameter $C > 0$)

$$\min_{w,b} \frac{C}{2}\|w\|_2^2 \quad \text{s.t.} \quad \forall (x^{(i)}, y^{(i)}) \in \mathcal{D} \quad y^{(i)}(w^\top x^{(i)} + b) \geq 1. \tag{1}$$

(a) (5 points) For the given data $\mathcal{D}$, how many constraints are part of the program in Eq. (1)? Specify all of them **explicitly**.

Your answer: 4 constraints $\quad w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$

$w^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \geq 1$

$w^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} + b \geq 1$

$-\left(w^T \begin{pmatrix} 0 \\ 0 \end{pmatrix} + b\right) \geq 1$

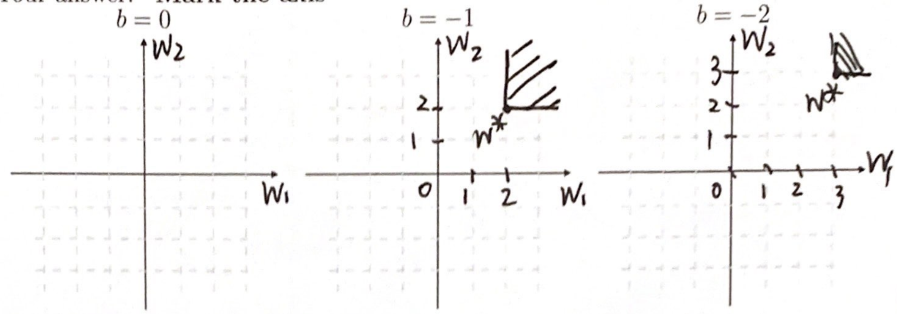
$-\left(w^T \begin{pmatrix} -1 \\ -1 \end{pmatrix} + b\right) \geq 1$

$$\Rightarrow \begin{cases} w_1 + b \geq 1 \\ w_2 + b \geq 1 \\ -b \geq 1 \\ w_1 + w_2 - b \geq 1 \end{cases}$$

Name: Yuxuan Zhang

(b) (8 points) Highlight the feasible set in w_1-w_2-space for $b = 0$, $b = -1$ and $b = -2$. For each of the three choices for b also highlight the optimal w. Given only the three options $b \in \{0, -1, -2\}$ what is the optimal solution? Does a better solution exist (reason)?

Your answer: **Mark the axis**



$b=0$. No feasible set

$b=-1$

$$\begin{cases} w_1 \geq 2 \\ w_2 \geq 2 \\ w_1 + w_2 + 1 \geq 1 \end{cases} \Rightarrow \begin{cases} w_1 \geq 2 \\ w_2 \geq 2 \end{cases}$$

$$w^* = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$b = -2$.

$$\begin{cases} w_1 \geq 3 \\ w_2 \geq 3. \end{cases}$$

$$w^* = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

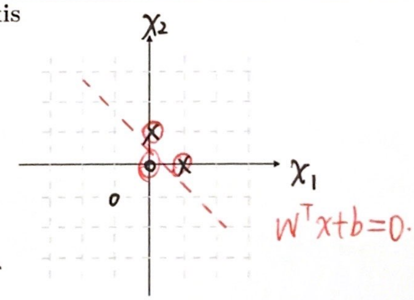
Optimal solution is $w \begin{pmatrix} 2 \\ 2 \end{pmatrix}$. $b=-1$. No better solution exists.
Because $b \geq -1$. $w_1 \geq 1-b$, $w_2 \geq 1-b$. So the cost will increase if b decrease

(c) (6 points) Draw the dataset in x_1-x_2-space using crosses for the points belonging to class 1 and circles for the points belonging to class -1. Find by inspection and highlight the support vectors, *i.e.*, those points for which the constraints hold with equality at the optimal solution. Solve the resulting linear system w.r.t. w and b and draw the solution into x_1-x_2-space.

Your answer: **Mark the axis**

$$\begin{cases} w_1 + b = 1 \\ w_2 + b = 1 \\ b = -1 \end{cases}$$

$\Rightarrow b = -1$, $w_1 = 2$, $w_2 = 2$



Name: Yuxuan Zhang

(d) (1 point) What conditions do the datapoints have to fulfill such that the program in Eq. (1) has a feasible solution?

Your answer:

They are linearly separable in the hyperplane.

(e) (6 points) In practice, for large datasets, it is hard to find the support vectors by inspection. A gradient based method is applicable. Use **general** notation, introduce slack variables into the program given in Eq. (1) and state the corresponding program (including all constraints). Subsequently, reformulate this program into an unconstrained program. Finally compute the gradient of this unconstrained program w.r.t. w (use $\frac{\partial}{\partial x}\max\{0, x\} = 1$ for $x > 0$, 0 otherwise). Evaluate the gradient at $w_1 = 2$, $w_2 = 2$ and $b = -1$. What can we conclude?

Your answer:

$$\min_{w,b,\xi^{(i)}\geq 0} \frac{C}{2}\|w\|_2^2 + \sum_{i\in D} \xi^{(i)}$$

$$\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi^{(i)} \quad i \in D.$$

$$\xi^{(i)} \geq 1 - y^{(i)}(w^T x^{(i)} + b)$$

$$\Rightarrow. \min_{w,b} \frac{C}{2}\|w\|_2^2 + \sum_{i\in D} \max\{0, 1 - y^{(i)}(w^T x^{(i)} + b)\}$$

let $C = 1$

$$\nabla_w f = w + \sum_{i\in D.} \delta(y^{(i)}(w^T x^{(i)} + b) < 1)\,(-y^{(i)} x^{(i)})$$

for $w_1 = w_2 = 2$. $b = -1$.

$y^{(1)}(w^T x^{(1)} + b) = 1 \quad \geq 1 \quad \delta(\ ) = 0$

$y^{(2)}(w^T x^{(2)} + b) = w_2 + b = 1 \quad \geq 1$

$y^{(3)}(w^T x^{(3)} + b) = 1 \quad \geq 1$

$y^{(4)}(w^T x^{(4)} + b) = -(-2-2-1) = 5 \quad \geq 1$

$\Rightarrow \nabla_w f = w = \begin{pmatrix}2\\2\end{pmatrix}$ The solution $w = \begin{pmatrix}2\\2\end{pmatrix}$ is not optimal. we also need to consider the gradient of b.

(f) (4 points) Complete A3_SVM.py and verify your reply for the previous answer. What is the optimal solution (w, b) that your program found and what's the corresponding loss? Explain the solution and what you observe when running the program, as well as how to fix this issue.

Your answer:

$(w,b) = \left(\begin{pmatrix}0.6671\\0.6671\end{pmatrix},\ 0.3330\right)$ Loss $= 1.779058$

The plot of loss is decreasing and converges to a number.

The C is fixed here and we need to change the value of C.