CS 7301
Problem Set 1
Student: TRI CAO
Date: September 9, 2017

$\boxed{\text{Warm-up}}$

1. Compute a subgradient

a) $f(x) = \max\left\{\frac{1}{2}x^2, |x|\right\}$

- at $x=0$, $f(x)$ can be either $\frac{1}{2}x^2$ ~~and~~ or $|x|$ and it is at the lowest point.

  $\Rightarrow$ Subgradient of $f(x)$ at $x=0$ is $\vec{0}$

- at $x = -2$

  we also have $\frac{1}{2}x^2 = |x|$

  At $x=-2$, ~~the~~ one subgradient of $f(x)$ is $x$

b) $g(x) = \max\{\exp(x), 10x\}$ at $x=2$ and $x=-1$

- at $x = 2$

  subgradient of $g(x)$ is $10x$

- at $x = -1$

  subgradient of $g(x)$ is $e^{-1}.x = \frac{x}{e}$

2. Show that $f(x) = \max\{ax+b,\ cx+d\}$ is a convex function for $x \in \mathbb{R}$ and any choice of constants $a, b, c, d \in \mathbb{R}$.

We have

- $\lambda f(x) + (1-\lambda) f(y)$ (call this LHS)

$$= \lambda \max\{ax+b,\ cx+d\} + (1-\lambda)\max\{ay+b,\ cy+d\}$$

- $f(\lambda x + (1-\lambda) y)$ (call this RHS)

$$= \max\{a(\lambda x + (1-\lambda)y) + b,\ c(\lambda x + (1-\lambda)y) + d\}$$

There are four cases:
$cx+d \geq ax+b$ and $cy+d \geq ay+b$
$cx+d \geq ax+b$ and $cy+d \leq ay+b$
$cx+d \leq ax+b$ and $cy+d \geq ay+b$
$cx+d \leq ax+b$ and $cy+d \leq ay+b$.

We will analyze the first two cases as the other two are very similar.

- Suppose $cx+d \geq ax+b$ and $cy+d \geq ay+b$

We have $LHS = \lambda(cx+d) + (1-\lambda)(cy+d)$

Now ~~as~~ we need to find the RHS.

$$RHS = \max\{\underbrace{a(\lambda x + (1-\lambda)y)+b}_{S},\ \underbrace{c(\lambda x + (1-\lambda)y)+d}_{T}\}$$

$$S = a\lambda x + a(1-\lambda)y + b$$
$$= a\lambda x + a(1-\lambda)y + b + \lambda b - \lambda b$$
$$= \lambda(ax+b) + (1-\lambda)(ay+b)$$

$$T = c\lambda x + c(1-\lambda)y + d$$
$$= c\lambda x + c(1-\lambda)y + d + \lambda d - \lambda d$$
$$= \lambda(cx+d) + (1-\lambda)(cy+d)$$

Since ~~$\lambda \geq 0$~~ $\lambda \geq 0$ and $1-\lambda \geq 0$ and $cx+d \geq ax+b$ and $cy+d \geq ay+d$

$\Rightarrow T \geq S \Rightarrow RHS = T = LHS$

$\Rightarrow LHS \geq RHS$ is true.

- Suppose $cx+d \geq ax+b$ and $cy+d \leq ax+b$.

We have $LHS = \lambda(cx+d) + (1-\lambda)(ay+b)$

Again, $RHS =$ ~~$\{S$~~ $\max\{S, T\}$ using the same notation in the previous case.

$$S = \lambda(ax+b) + (1-\lambda)(ay+b)$$
$$T = \lambda(cx+d) + (1-\lambda)(cy+d)$$

Because $\lambda \geq 0$, $(1-\lambda) \geq 0$, $cx+d \geq ax+b$ and $cy+d \leq ay+b$, ~~B~~ it's clear that $LHS \geq S$ and $LHS \geq T$.

therefore ~~$\&$~~ $LHS \geq RHS$ is true.

- the other two cases will have the similar analyses.

And we conclude that $LHS \geq RHS$ in all cases.

That means $f(x) = \max\{ax+b, cx+d\}$ is convex.

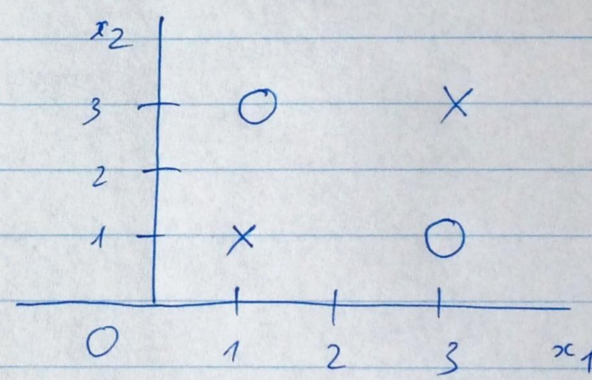
## Problem 1

1 and 2: run the code at 'problem-1.py'

3. For this ~~problem~~ particular data set, the rate of convergence does not change as I change the step size. In the code, I test 4 different step sizes in the set $\{10, 1, 0.1, 0.01\}$, but the rate of converge is the same.

4. The smallest number data points on which the algorithm fails to converge is 4.

An example is:




In this example, without using new feature vectors, we can not find a linear separator for this data set.
We can see that there will be cases where we cannot find a perfect linear classifier for the data set (even with feature engineering)

## $\boxed{\text{Problem 2}}$

~~Two~~ One feature vector~~s~~ will make the data linearly separable:

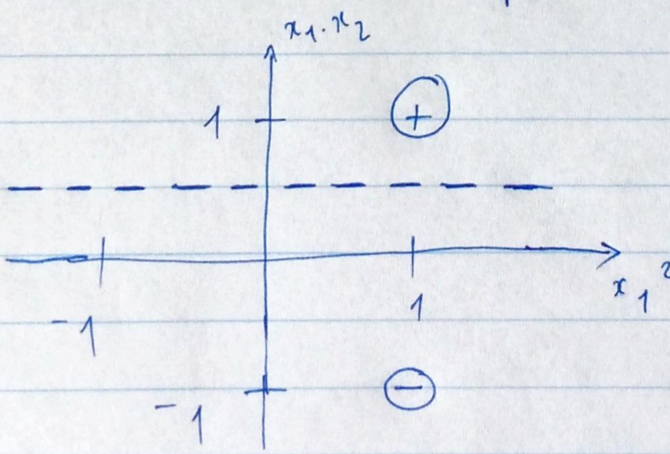
b) $\phi(x_1, x_2) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \end{bmatrix}$

Let's use only features: $x_1^2$ and $x_1 x_2$

We have the data set as:

| $x_1^2$ | $x_1 . x_2$ | $y$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | -1 | -1 |
| 1 | -1 | -1 |
| 1 | 1 | 1 |

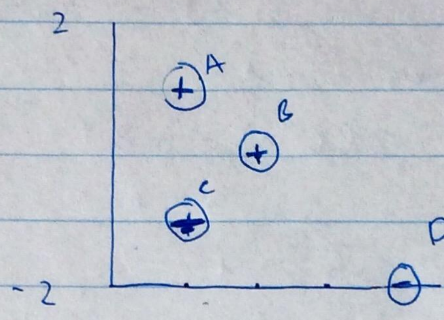
Let's draw the data points




It's clear that a linear separator exists, such as the x-axis, or the dashed line in the figure.

## Problem 5 Support Vector Machines, by Hand!

First, let's visualize the point




Point A and B are positive $(y=1)$
C and D are negative $(y=-1)$

Note that A, B and D are on a line.

1. Let formulate our problem:

$$\min_{w} \; w_1^2 + w_2^2$$

$$\begin{aligned}
\text{s.t.} \quad & 1 . b \geq 1 && (B)\\
& -1(-w_1 - w_2 + b) \geq 1 && (C)\\
& -1(2w_1 - 2w_2 + b) \geq 1 && (D)\\
& 1(-w_1 + w_2 + b) \geq 1 && (A)
\end{aligned}$$

Now we focus on two positive points A and B.
One of them must be a support vector.

- Suppose A is a support vector.

Then $-w_1 + w_2 + b = 1$

$$\Leftrightarrow \boxed{w_1 - w_2 = b - 1}$$

Consider point D, we need: $-2w_1 + 2w_2 - b \geq 1$

$$\Leftrightarrow -2(w_1 - w_2) - b \geq 1$$

Since $w_1 - w_2 = b - 1$

$$\Rightarrow -2(b-1) - b \geq 1$$

$$\Leftrightarrow -2b + 2 - b \geq 1$$

$$\Leftrightarrow -3b \geq -1$$

$$\Leftrightarrow b \leq \frac{1}{3}$$

But due to a constraint of point B, $b \geq 1 \Rightarrow$ contradict.

$\Rightarrow$ Point A cannot be a support vector.

And Point B is a support vector.

- From the figure and the fact that B is a support vector, we can guess that either negative point can be a support vector.

- Here, I will only solve one case. Let it be point C being a ~~vecto~~ support vector.

Because B is a support vector, we have $\boxed{b = 1}$

Because C is also a support vector,

$$w_1 + w_2 - b \cancel{\geq} = 1$$

$$\Leftrightarrow \boxed{w_1 + w_2 \cancel{\geq} 2}$$

$$\hookrightarrow w_1 + w_2 = 2$$

the other constrainsts are:

$$\begin{cases} -1(2w_1 - 2w_2 + b) \geq 1 & (1) \\ 1(-w_1 + w_2 + b) \geq 1 & (2) \end{cases}$$

$$(1) \Leftrightarrow -2(w_1 - w_2) - 1 \geq 1$$
$$\Leftrightarrow -2(w_1 - w_2) \geq 2$$
$$\Leftrightarrow -(w_1 - w_2) \geq 1$$
$$\Leftrightarrow \boxed{w_2 - w_1 \geq 1}$$

$$(2) \Leftrightarrow w_2 - w_1 \ngeq 0$$

Using $w_2 + w_1 = 2$ and $w_2 - w_1 \geq 1$ we can choose $\boxed{w_2 = 1.5 \,;\; w_1 = 0.5}$

let check the constraints:

$$1.b = 1.1 = 1 \geq 1 \quad (B)$$
$$-1(-w_1 - w_2 + b) = 1 \geq 1 \quad (C)$$
$$-1(2w_1 - 2w_2 + b) = 1 \geq 1 \quad (D)$$
$$1(-w_1 + w_2 + b) = 2 \geq 1 \quad (A)$$

$\Rightarrow$ we have a valid solution with $w_2 = 1.5$, $w_1 = 0.5$
$b = 1$

Note that we choose $w_2$ and $w_1$ to minimize $w_2^2 + w_1^2$. Intuitively, if we choose $w_2$ to be close to 2, then $w_2^2$ will be large, so we want to minimize $w_2$. That's why I choose $w_2 = \frac{3}{2}$ and $w_1 = \frac{1}{2}$.

3. Size of the margin:

The size of the margin is $\frac{1}{||w||} = \frac{1}{\sqrt{w_1^2 + w_2^2}} = \frac{1}{\sqrt{\frac{10}{4}}} = \sqrt{\frac{4}{10}}$

## Problem 4

- The optimization problem I solve is the basic SVM formulation.

$$\min_{w} \|w\|^2$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \text{ for all } i$$

- I add a number of features in addition to the original features.

  There are two types of features that are added:
  - Quadratic features; such as $x_1^2$ and $x_1 . x_2$
  - Cubic features, such as $x_1^3$ and $x_2 . x_3 . x_0$ and $x_2^2 . x_3$

- The calculated margin is

$$\frac{1}{\|w\|} = 157.185$$

- The weights and bias are shown in the output of the code at 'problem-4.py'