Jonathan Nusantara jan265 Assignment # 3

Problem 1 (10 points). Recall that a linear classifier is specified by $(\vec{w}, t)$, where $\vec{w} \in \mathbb{R}^d$, and $t \in \mathbb{R}$, and a decision rule $\vec{w} \cdot \vec{X} \lessgtr t$ for a feature vector $\vec{X} \in \mathbb{R}^d$.

Recall the perceptron algorithm from the lectures. Suppose $d = 2$, $n = 4$, and there are four training examples, given by:

i	feature $\vec{X}_i$	label y_i
1	$(-0.6, 1)$	-1
2	$(-3, -4)$	-1
3	$(3, -2)$	$+1$
4	$(0.5, 1)$	$+1$

1. In the class we started with the the initial $(\vec{w}, t) = (\vec{0}, 0)$, and derived the convergence of perceptron. In this problem:
   - Start with the initialization $(\vec{w}, t) = ((0, 1), 0)$ (the x-axis).
   - Implement the perceptron algorithm **by hand**. Go over the data-points **in order**. Output a table of the form below where each row works with one example in some iteration. We have filled in some entries in the first two rows. You need to add rows until no mistakes happen on any example.

starting $\vec{w}$	starting t	features $\vec{X}_i$	label y_i	predicted label	new $\vec{w}$	new t
$(0, 1)$	0	$(-0.6, 1)$	-1	...	...	...
...	...	$(-3, -4)$	-1	...	...	...
...	...	...	...	...	...	...

   - Draw a 2-d grid. On this grid, mark the four examples (like we do on the board). Draw the line you obtain as the final result.

(1)

$\hat{y} = \text{sign}(\vec{w} \cdot \vec{X}_i - t)$; $\vec{w} + y_i \cdot \vec{X}_i$; $t - y_i$

starting $\vec{w}$	starting t	features $\vec{X}_i$	label y_i	Predicted label $\hat{y}$	new $\vec{w}$	new t
(0,1)	0	(-0.6,1)	-1	sign(0×-0.6+1×1-0) = 1 -> +1	(0,1)+(0.6,-1)=(0.6,0)	0-(-1) = 1
(0.6,0)	1	(-3,-4)	-1	sign(0.6×-3+0×-4-1) = -2.8 -> -1	(0.6,0)	1
(0.6,0)	1	(3,-2)	+1	sign(0.6×3+0×-2-1) = 0.8 -> +1	(0.6,0)	1
(0.6,0)	1	(0.5,1)	+1	sign(0.6×0.5+0×1-1) = -0.7 -> -1	(0.6,0)+(0.5,1)=(1.1,1)	1-1=0
(1.1,1)	0	(-0.6,1)	-1	sign(1.1×-0.6+1×1-0) = 0.34 -> +1	(1.1,1)+(0.6,-1)=(1.7,0)	0-(-1)=1
(1.7,0)	1	(-3,-4)	-1	sign(1.7×-3+0×-4-1) = -6.1 -> -1	(1.7,0)	1
(1.7,0)	1	(3,-2)	+1	sign(1.7×3+0×-2-1) = 4.1 -> +1	(1.7,0)	1
(1.7,0)	1	(0.5,1)	+1	sign(1.7×0.5+0×1-1) = -0.15 -> -1	(1.7,0)+(0.5,1)=(2.2,1)	1-1=0
(2.2,1)	0	(-0.6,1)	-1	sign(2.2×-0.6+1×1-0) = -0.32 -> -1	(2.2,1)	0
(2.2,1)	0	(-3,-4)	-1	sign(2.2×-3+1×-4-0) = -10.6 -> -1	(2.2,1)	0
(2.2,1)	0	(3,-2)	+1	sign(2.2×3+1×-2-0) = 4.6 -> +1	(2.2,1)	0
(2.2,1)	0	(0.5,1)	+1	sign(2.2×0.5+1×1-0) = 2.1 -> +1	(2.2,1)	0

$\vec{w} = (2.2, 1)$ $t = 0$

② $\vec{w} = (2.2, 1)$, $t = 0$, $(\bar{w}, t) = ((2.2, 1), 0)$

$\bar{w} \cdot \bar{x} - t = 0 \rightarrow (2.2, 1) \cdot (x_1, x_2) - 0 = 0 \rightarrow 2.2x_1 + x_2 = 0$

when $x_1 = 1$

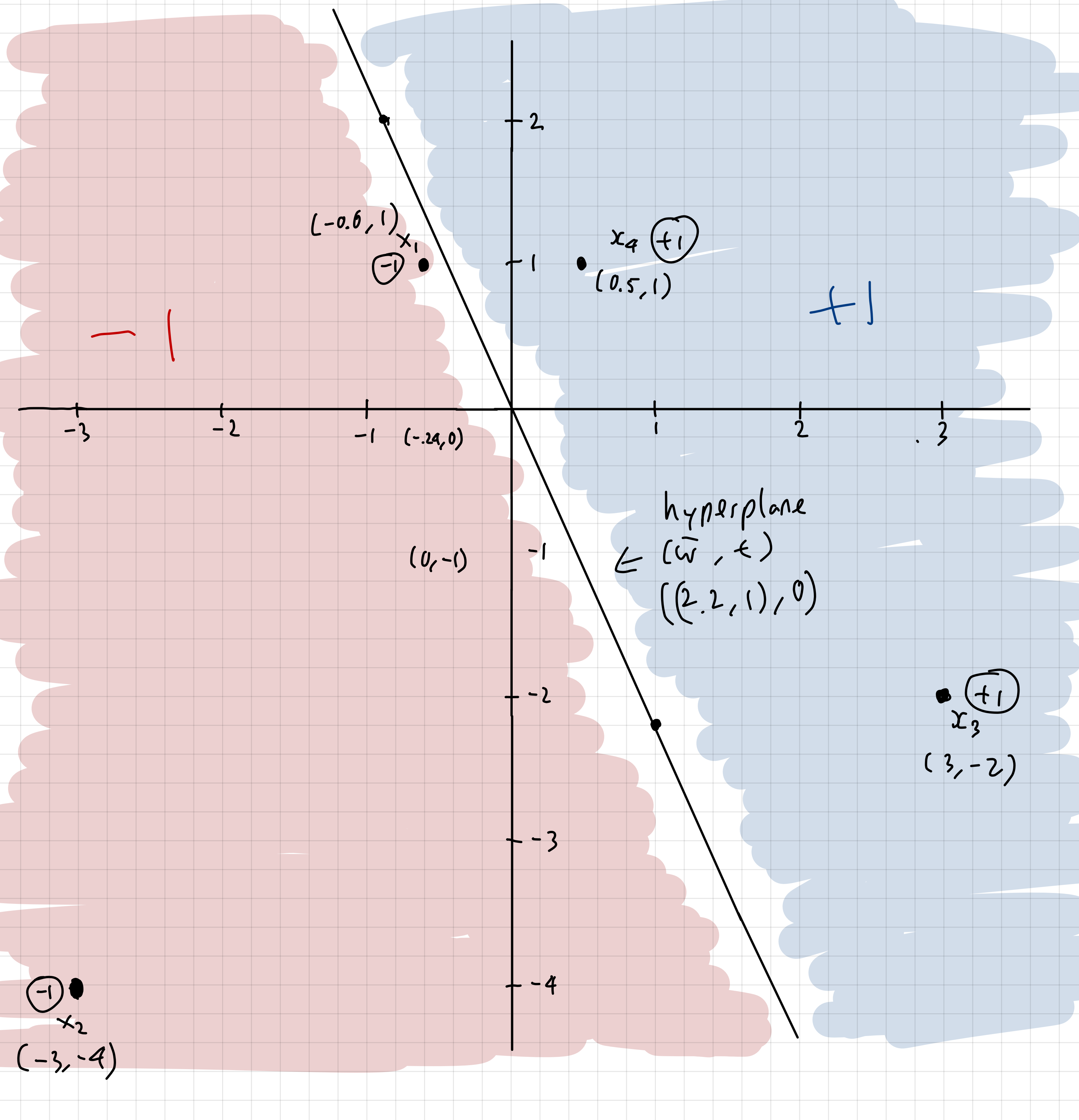
$2.2 + x_2 = 0$

$x_2 = -2.2$

when $x_2 = 2$

$2.2x_1 + 2 = 0$

$x_1 = -\frac{2}{2.2} = -0.91$



Problem 2 (10 points). Recall the log-likelihood function for logistic regression:

$$J(\overrightarrow{w}, t) = \sum_{i=1}^{n} \log Pr(y_i | \overrightarrow{X}_i, \overrightarrow{w}, t),$$

where $Pr(y_i|\overrightarrow{X}_i, \overrightarrow{w}, t)$ is the same as defined in the class for logistic regression, and $y_i \in \{-1, +1\}$. We will show that this function is concave as a function of $\overrightarrow{w}, t$.

1. Show that for two real numbers a, b, $\exp(a) + \exp(b) \geq 2\exp((a+b)/2)$. (Basically this says that exponential function is convex.)
2. Extending this, show that for any vectors $\overrightarrow{w}_1, \overrightarrow{w}_2, \overrightarrow{x} \in \mathbb{R}^d$,
$$\exp(\overrightarrow{w}_1 \cdot \overrightarrow{x}) + \exp(\overrightarrow{w}_2 \cdot \overrightarrow{x}) \geq 2\exp\left(\frac{(\overrightarrow{w}_1 + \overrightarrow{w}_2)}{2} \cdot \overrightarrow{x}\right).$$
3. Show that $J(\overrightarrow{w}, t)$ is concave (you can only show the concavity holds for $\lambda = 1/2$). You can show it any way you want. One way is to first show that for any $\overrightarrow{w}_1, \overrightarrow{w}_2 \in \mathbb{R}^d$, and $t_1, t_2 \in \mathbb{R}$,
$$\frac{1}{2}J(\overrightarrow{w}_1, t_1) + \frac{1}{2}J(\overrightarrow{w}_2, t_2) \leq J\left(\frac{\overrightarrow{w}_2 + \overrightarrow{w}_2}{2}, \frac{t_1 + t_2}{2}\right).$$
You can use that sum of concave functions are concave. A linear function is both concave and convex.

In this problem you can also work with the vector $\overrightarrow{w}^*$, which is the $d+1$ dimensional vector $(\overrightarrow{w}, t)$, and the $d+1$ dimensional feature vectors $\overrightarrow{X}_i^* = (\overrightarrow{X}_i, -1)$, which are the features appended with a -1. Just like in class, this can simplify some of the computations by using the fact that $\overrightarrow{w} \cdot \overrightarrow{X}_i - t = \overrightarrow{w}^* \cdot \overrightarrow{X}_i^*$.

1

$$e^a + e^b \geq 2e^{\frac{a+b}{2}}$$

$$e^a + e^b \geq 2e^{\frac{a+b}{2}}$$

$$e^a + e^b \geq 2e^{\frac{a}{2}} e^{\frac{b}{2}}$$

Let $x = e^{a/2}$ and $y = e^{b/2}$

$$x^2 + y^2 \geq 2xy$$

$$x^2 - 2xy + y^2 \geq 0$$

$$(x-y)^2 \geq 0$$

$$\left(e^{\frac{a}{2}} - e^{\frac{b}{2}}\right)^2 \geq 0 \quad \text{(proven)}$$

a and b can be positive or negative, because due to the square $()^2$, it will always ≥ 0

② Prove $\exp(\vec{w}_1 \cdot \vec{x}) + \exp(\vec{w}_2 \cdot \vec{x}) \geq 2\exp\left(\frac{(\vec{w}_1 + \vec{w}_2)}{2} \cdot \vec{x}\right).$

$$e^{(\bar{w}_1 \cdot \bar{x})} + e^{(\bar{w}_2 \cdot \bar{x})} \geq 2e^{\left(\frac{\bar{w}_1 + \bar{w}_2}{2} \cdot \bar{x}\right)}$$

$$e^{(\bar{w}_1 \cdot \bar{x})} + e^{(\bar{w}_2 \cdot \bar{x})} \geq 2e^{\left(\frac{\bar{w}_1}{2} \cdot \bar{x}\right)} e^{\left(\frac{\bar{w}_2}{2} \cdot \bar{x}\right)}$$

$$e^{(\bar{w}_1 \cdot \bar{x})} + e^{(\bar{w}_2 \cdot \bar{x})} \geq 2e^{\frac{1}{2}(\bar{w}_1 \cdot \bar{x})} e^{\frac{1}{2}(\bar{w}_2 \cdot \bar{x})}$$

$a^2 = e^{\frac{1}{2}(\bar{w}_1 \cdot \bar{x}) + \frac{1}{2}(\bar{w}_1 \cdot \bar{x})}$
$= e^{(\bar{w}_1 \cdot \bar{x})}$

Let $a = e^{\frac{1}{2}(\bar{w}_1 \cdot \bar{x})}$ and $b = e^{\frac{1}{2}(\bar{w}_2 \cdot \bar{x})}$

$$a^2 + b^2 \geq 2ab$$

$$a^2 - 2ab + b^2 \geq 0$$

$$(a-b)^2 \geq 0$$

$$\left(e^{\frac{1}{2}(\bar{w}_1 \cdot \bar{x})} - e^{\frac{1}{2}(\bar{w}_2 \cdot \bar{x})}\right)^2 \geq 0$$

(proven)

Since $\left(e^{\frac{1}{2}(\bar{w}_1 \cdot \bar{x})} - e^{\frac{1}{2}(\bar{w}_2 \cdot \bar{x})}\right)$ is squared, whatever $\left(e^{\frac{1}{2}(\bar{w}_1 \cdot \bar{x})} - e^{\frac{1}{2}(\bar{w}_2 \cdot \bar{x})}\right)$ is equal to, positive or negative, it will be squared and ≥ 0

this holds for any vector $\vec{w_1}, \vec{w_2}, \vec{x} \in \mathbb{R}^d$

③

3. Show that $J(\vec{w}, t)$ is concave (you can only show the concavity holds for $\lambda = 1/2$). You can show it any way you want. One way is to first show that for any $\vec{w}_1, \vec{w}_2 \in \mathbb{R}^d$, and $t_1, t_2 \in \mathbb{R}$,

$$\frac{1}{2}J(\vec{w}_1, t_1) + \frac{1}{2}J(\vec{w}_2, t_2) \leq J\left(\frac{\vec{w}_2 + \vec{w}_2}{2}, \frac{t_1 + t_2}{2}\right).$$

You can use that sum of concave functions are concave. A linear function is both concave and convex.

In this problem you can also work with the vector $\vec{w}^*$, which is the $d+1$ dimensional vector $(\vec{w}, t)$, and the $d+1$ dimensional feature vectors $\vec{X}_i^* = (\vec{X}_i, -1)$, which are the features appended with a -1. Just like in class, this can simplify some of the computations by using the fact that $\vec{w} \cdot \vec{X}_i - t = \vec{w}^* \cdot \vec{X}_i^*$.

Let

$$J_S(\bar{w},t) = \sum_{i=1}^{n} \log P(y_i \mid \bar{x}_i, \bar{w}, t) \quad (\text{a fn of } (\bar{w}, t)).$$

$$= \sum_{i=1}^{n} \left[\left(\frac{1+y_i}{2}\right) \cdot (\bar{w} \cdot \bar{x}_i - t) - \log\left(1 + \exp(\bar{x}_i \bar{w} - t)\right) \right]$$

(since $\log(\exp(x)) = x$).

Let's first prove it for one i. Let $i = 1$.

$$\frac{1}{2}\left(\frac{(1+Y_1)}{2} \cdot (\bar{w}_1 \cdot \bar{x}_1 - t_1) - \log(1 + \exp(\bar{x}_1 \cdot \bar{w}_1 - t_1))\right) + \frac{1}{2}\left(\frac{(1+Y_1)}{2} \cdot (\bar{w}_2 \cdot \bar{x}_1 - t_2) - \log(1 + \exp(\bar{x}_1 \bar{w}_2 - t_2))\right)$$

$$\leq \left(\frac{(1+Y_1)}{2} \cdot \left(\frac{\bar{w}_1 + \bar{w}_2}{2} \cdot \bar{x}_1 - \frac{t_1 + t_2}{2}\right) - \log\left(1 + \exp\left(\bar{x}_1 \cdot \frac{\bar{w}_1 + \bar{w}_2}{2} - \frac{t_1 + t_2}{2}\right)\right)\right)$$

Let $\bar{w}^*$ be $d+1$ dimension $(\bar{w}, t)$ and $\bar{x}_i^* = (\bar{x}_i, -1)$ also $d+1$ dimension.

So, $\bar{w} \cdot \bar{x}_i - t = \bar{w}^* \cdot \bar{x}_i^*$

$$\frac{1}{2}\left(\frac{(1+Y_1)}{2} \cdot (\bar{w}_1^* \cdot \bar{x}_1^*) - \log(1 + \exp(\bar{w}_1^* \cdot \bar{x}_1^*))\right) + \frac{1}{2}\left(\frac{(1+Y_1)}{2} \cdot (\bar{w}_2^* \cdot \bar{x}_1^*) - \log(1 + \exp(\bar{w}_2^* \cdot \bar{x}_1^*))\right)$$

$$\leq \left(\frac{(1+Y_1)}{2} \cdot \left(\frac{1}{2}(\bar{w}_1^* + \bar{w}_2^*) \cdot \bar{x}_1^*\right) - \log\left(1 + \exp\left(\frac{1}{2}(\bar{w}_1^* + \bar{w}_2^*) \cdot \bar{x}_1^*\right)\right)\right)$$

$$\frac{(1+Y_1)}{4} \cdot (\bar{w}_1^* \cdot \bar{x}_1^*) + \frac{(1+Y_1)}{4} \cdot (\bar{w}_2^* \cdot \bar{x}_1^*) - \frac{1}{2}\log(1 + \exp(\bar{w}_1^* \cdot \bar{x}_1^*)) - \frac{1}{2}\log(1 + \exp(\bar{w}_2^* \cdot \bar{x}_1^*))$$

$$\leq \frac{(1+Y_1)}{4} \cdot \left((\bar{w}_1^* + \bar{w}_2^*) \cdot \bar{x}_1^*\right) - \log\left(1 + \exp\left(\frac{1}{2}(\bar{w}_1^* + \bar{w}_2^*) \cdot \bar{x}_1^*\right)\right)$$

$$\frac{(1+y_1)}{4}\cdot\left(\bar{w}_1^*\cdot\bar{x}_1^* + \bar{w}_2^*\cdot\bar{x}_1^*\right) - \frac{1}{2}\log\left(1+\exp(\bar{w}_1^*\cdot\bar{x}_1^*)\right) - \frac{1}{2}\log\left(1+\exp(\bar{w}_2^*\cdot\bar{x}_1^*)\right)$$

$$\le \frac{(1+y_1)}{4}\cdot\left((\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right) - \log\left(1+\exp\left(\frac{1}{2}(\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right)\right)$$

$$\underbrace{\frac{(1+y_1)}{4}\cdot\left((w_1^*+w_2^*)\cdot\bar{x}_1^*\right)}_{=\ \text{equal}} - \frac{1}{2}\log\left(1+\exp(\bar{w}_1^*\cdot\bar{x}_1^*)\right) - \frac{1}{2}\log\left(1+\exp(\bar{w}_2^*\cdot\bar{x}_1^*)\right)$$

$$\le \frac{(1+y_1)}{4}\cdot\left((\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right) - \log\left(1+\exp\left(\frac{1}{2}(\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right)\right)$$

$$-\frac{1}{2}\log\left(1+\exp(\bar{w}_1^*\cdot\bar{x}_1^*)\right) - \frac{1}{2}\log\left(1+\exp(\bar{w}_2^*\cdot\bar{x}_1^*)\right) \le -\log\left(1+\exp\left(\frac{1}{2}(\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right)\right)$$

$$-\frac{1}{2}\left(\log\left(1+\exp(\bar{w}_1^*\cdot\bar{x}_1^*)\right) + \log\left(1+\exp(\bar{w}_2^*\cdot\bar{x}_1^*)\right)\right) \le -\log\left(1+\exp\left(\frac{1}{2}(\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right)\right)$$

$$\log\left(1+\exp(\bar{w}_1^*\cdot\bar{x}_1^*)\right) + \log\left(1+\exp(\bar{w}_2^*\cdot\bar{x}_1^*)\right) \ge 2\log\left(1+\exp\left(\frac{1}{2}(\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right)\right)$$

$$\log\left(\left(1+\exp(\bar{w}_1^*\cdot x_1^*)\right)\left(1+\exp(\bar{w}_2^*\cdot x_1^*)\right)\right) \ge 2\log\left(1+\exp\left(\frac{1}{2}(\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right)\right)$$

$$\left(1+\exp(\bar{w}_1^*\cdot x_1^*)\right)\left(1+\exp(\bar{w}_2^*\cdot x_1^*)\right) \ge \left(1+\exp\left(\frac{1}{2}(\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right)\right)^2$$

$$1+\exp(\bar{w}_2^*\cdot x_1^*) + \exp(\bar{w}_1^*\cdot x_1^*) + \exp\left((\bar{w}_1^*+\bar{w}_2^*)\cdot x_1^*\right)$$
$$\ge 1 + 2\exp\left(\frac{1}{2}\,\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right) + \exp\left((\bar{w}_1^*+\bar{w}_2^*)\cdot x_1^*\right)$$

$$\exp(\bar{w}_2^*\cdot x_1^*) + \exp(\bar{w}_1^*\cdot x_1^*) \ge 2\exp\left(\frac{1}{2}\,\bar{w}_1^*+\bar{w}_2^*)\cdot\bar{x}_1^*\right)$$

In 2.2, we have proven this: $\exp(\vec{w}_1\cdot\vec{x}) + \exp(\vec{w}_2\cdot\vec{x}) \ge 2\exp\left(\frac{(\vec{w}_1+\vec{w}_2)}{2}\cdot\vec{x}\right).$

So, we have proven the concavity when $i=1$.

$\Downarrow$ next page

Since it is proven that for one i, $i=1$ is concave and we know that the $\underbrace{\text{sum of concave function is concave.}}_{\text{know from lecture}}$

So, when $J(\bar{w},t) = \sum_{i=1}^{n} \left[\frac{(1+y_i)}{2} \cdot (\bar{w} \cdot \bar{x}_i - t) - \log\left(1+\exp(\bar{x}_i \bar{w} - t)\right) \right]$

we've proven that $\frac{1}{2}J(\overrightarrow{w}_1, t_1) + \frac{1}{2}J(\overrightarrow{w}_2, t_2) \leq J\left(\frac{\overrightarrow{w}_2 + \overrightarrow{w}_2}{2}, \frac{t_1 + t_2}{2}\right).$

and that $J(\bar{w},t)$ is concave.

Problem 3. (15 points). Consider the same set-up as in perceptron, where the features all satisfy $|\vec{X}_i| \le 1$, and the training examples are separable with margin γ. We showed in class that perceptron converges with at most $4/\gamma^2$ updates when initialized to the all all zero vectors, namely to $(\vec{0}, 0)$. Suppose instead the initial start with some initial $(\vec{w_0}, t_0)$ with $||\vec{w_0}|| \le R$, and $|t_0| \le R$. We run the perceptron algorithm with this initialization. Suppose, $(\vec{w_j}, t_j)$ is the hyperplane after jth update. Let $(\vec{w}_{opt}, t_{opt})$ be the optimal hyperplane. Then,

Similar to what we did in class, assume that the $d+1$ dimensional vector $(\vec{w}_{opt}, t_{opt})$ satisfies

$$||(\vec{w}_{opt}, t_{opt})||^2 \le 2. \quad \leftarrow \text{claim 2 in lecture}$$

1. Show that
$$(\vec{w}_j, t_j) \cdot (\vec{w}_{opt}, t_{opt}) \ge j\gamma - 2R.$$

2. Show that
$$(\vec{w}_j, t_j) \cdot (\vec{w}_j, t_j) \le 2j + 2R^2.$$

3. Using these conclude that the number of updates before perceptron converges is at most
$$\frac{4 + 4R\gamma}{\gamma^2}$$
updates.

Just like in lecture, we want to proof by induction

(1) Suppose the jth update happens at $(\bar{x}_i, y_i)$, i.e.,

$$\text{sign}(\bar{w}_{j-1} \cdot \bar{x}_i - t_{j-1}) \ne y_i$$

$$\bar{w}_j = \bar{w}_{j-1} + y_i \bar{x}_i, \quad t_j = t_{j-1} - y_i$$

$$(\bar{w}_j, t_j) \cdot (\bar{w}_{opt}, t_{opt}) = (\bar{w}_{j-1} + y_i \bar{x}_i, t_{j-1} - y_i) \cdot (\bar{w}_{opt}, t_{opt})$$

$$= \bar{w}_{j-1} \cdot \bar{w}_{opt} + y_i \bar{x}_i \cdot \bar{w}_{opt} + t_{j-1} t_{opt} - y_i t_{opt}$$

we know that

$y_i(\bar{w}_{opt} \cdot \bar{x}_i - t_{opt}) \ge \gamma$

(claim 1 in lecture)

$$= (\bar{w}_{j-1}, t_{j-1}) \cdot (\bar{w}_{opt}, t_{opt}) + y_i(\bar{w}_{opt} \cdot x_i - t_{opt})$$

$$\ge (\bar{w}_{j-1}, t_{j-1}) \cdot (\bar{w}_{opt}, t_{opt}) + \gamma$$

Since when $j = j-1 \rightarrow (\bar{w}_{j-1}, t_{j-1}) \cdot (\bar{w}_{opt}, t_{opt}) \ge (j-1)\gamma - 2R$

$$\ge (j-1)\gamma - 2R + \gamma$$

$$\ge j\gamma - \gamma - 2R + \gamma$$

$$(\bar{w}_j, t_j) \cdot (\bar{w}_{opt}, t_{opt}) \ge j\gamma - 2R \quad \text{(proven)}$$

(2) Suppose the jth update happens at $(\bar{x}_i, y_i)$, i.e.,

$$\text{sign}(\bar{w}_{j-1} \cdot \bar{x}_i - t_{j-1}) \neq y_i$$

$$\bar{w}_j = \bar{w}_{j-1} + y_i \bar{x}_i \quad , \quad t_j = t_{j-1} - y_i$$

$$(\bar{w}_j, t_j) \cdot (\bar{w}_j, t_j) = (\bar{w}_{j-1} + y_i \bar{x}_i, t_{j-1} - y_i) \cdot (\bar{w}_{j-1} + y_i \bar{x}_i, t_{j-1} - y_i)$$

$$= (\bar{w}_{j-1} + y_i \bar{x}_i)(\bar{w}_{j-1} + y_i \bar{x}_i) + (t_{j-1} - y_i)(t_{j-1} - y_i)$$

$$= \bar{w}_{j-1} \cdot \bar{w}_{j-1} + y_i(\bar{w}_{j-1} \cdot \bar{x}_i) + y_i(\bar{w}_{j-1} \cdot \bar{x}_i) + \overset{1}{y_i^2}(\bar{x}_i \cdot \bar{x}_i) + t_{j-1} \cdot t_{j-1} - y_i t_{j-1} - y_i t_{j-1} + \overset{1}{y_i^2}$$

since it is wrong prediction

$$2y_i(\bar{w}_{j-1} \cdot \bar{x}_i - t_{j-1})$$

if y_i	$(\bar{w}_{j-1} \cdot \bar{x}_i - t_{j-1})$
>0	<0
<0	>0

so always ≤ 0

$$= (\bar{w}_{j-1}, t_{j-1}) \cdot (\bar{w}_{j-1}, t_{j-1}) + \underbrace{\|\bar{x}_i\|_2^2}_{1^2=1} + 1 + \underbrace{2y_i(\bar{w}_{j-1} \cdot \bar{x}_i - t_{j-1})}_{\text{always } \leq 0}$$

$$\leq (2(j-1) + 2R^2) + 1 + 1$$

$$\leq 2j - \cancel{2} + 2R^2 + \cancel{2}$$

$$(\bar{w}_j, t_j) \cdot (\bar{w}_j, t_j) \leq 2j + 2R^2 \quad \text{(proven)}$$

③ Proof that $j \leq \frac{4+4R\gamma}{\gamma^2}$

$$|\text{cosine}(\theta)| \leq 1$$

$$\frac{((\bar{w}_j, t_j) \cdot (\bar{w}_{opt}, t_{opt}))^2}{\|\bar{w}_j, t_j\|_2^2 \ \|w_{opt}, t_{opt}\|_2^2} \leq 1$$

Using the proof $(\overrightarrow{w}_j, t_j) \cdot (\overrightarrow{w}_{opt}, t_{opt}) \geq j\gamma - 2R.$ and $(\overrightarrow{w}_j, t_j) \cdot (\overrightarrow{w}_j, t_j) \leq 2j + 2R^2.$ and $\|(\overrightarrow{w}_{opt}, t_{opt})\|^2 \leq 2.$

$$\frac{(j\gamma - 2R)^2}{(2j+2R^2)(2)} \leq 1$$

$$\frac{j^2\gamma^2 - 2jR\gamma - 2jR\gamma + 4R^2}{4j + 4R^2} \leq 1$$

$$j^2\gamma^2 - 2jR\gamma - 2jR\gamma + \cancel{4R^2} \leq 4j + \cancel{4R^2}$$

$$j^2\gamma^2 - 4jR\gamma - 4j \leq 0$$

$$j(j\gamma^2 - 4R\gamma - 4) \leq 0$$

$$\underline{j \leq 0}$$

$$j\gamma^2 - 4R\gamma - 4 \leq 0$$

$$\underline{j = \frac{4+4R\gamma}{\gamma^2}}$$

(proven)

$$j \leq \frac{4+4R\gamma}{\gamma^2}$$