Deep Learning - Theory and Practice

IE 643
Lecture 4

August 9, 2024.

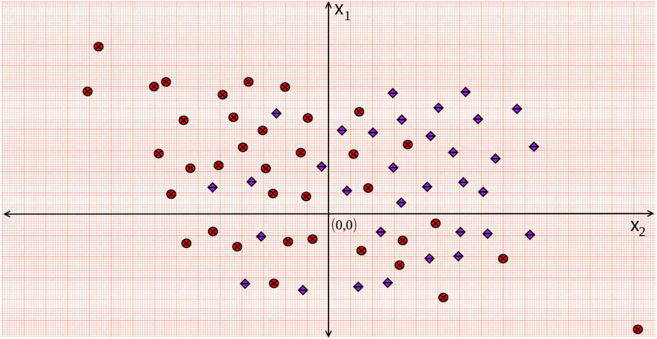
1. Perceptron Convergence

Convergence of Perceptron Training

Perceptron Convergence - Geometric Intuition

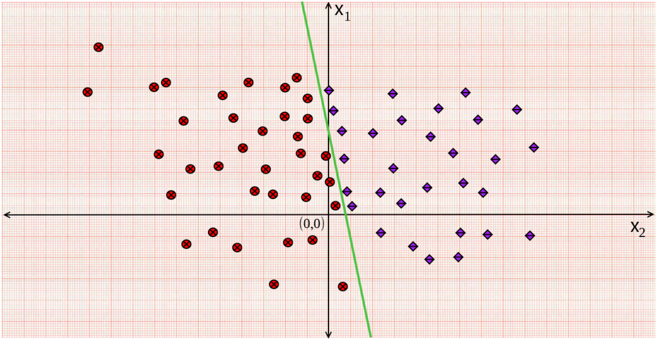
- What are some natural assumptions to expect the perceptron training to converge?
- Let us first motivate such assumptions through geometric intuition.

Perceptron Convergence - Geometric Intuition



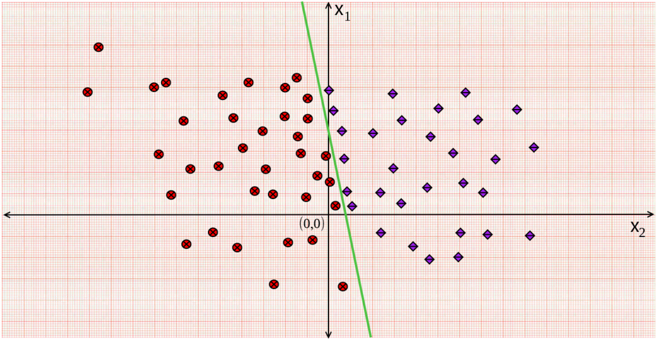
- Can the data be separated by a hyperplane?

Perceptron Convergence - Geometric Intuition



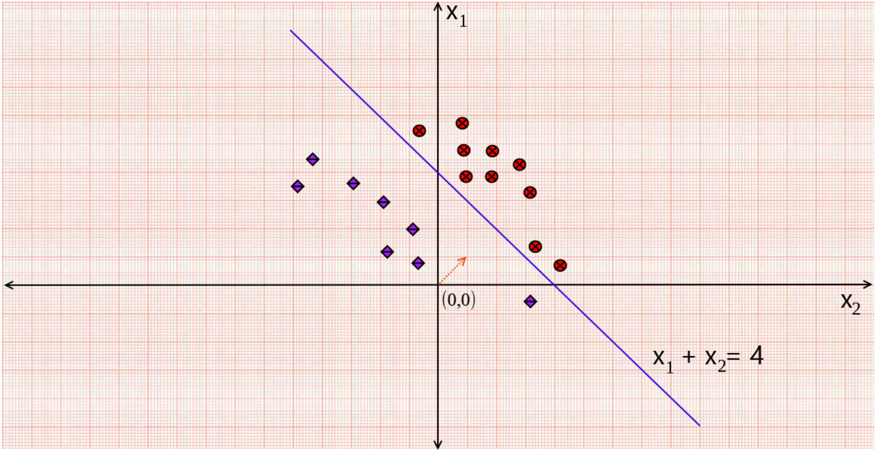
- **First assumption:** At least the data should be such that the samples with label 1 can be separated by a hyperplane from samples with label -1.

Perceptron Convergence - Geometric Intuition

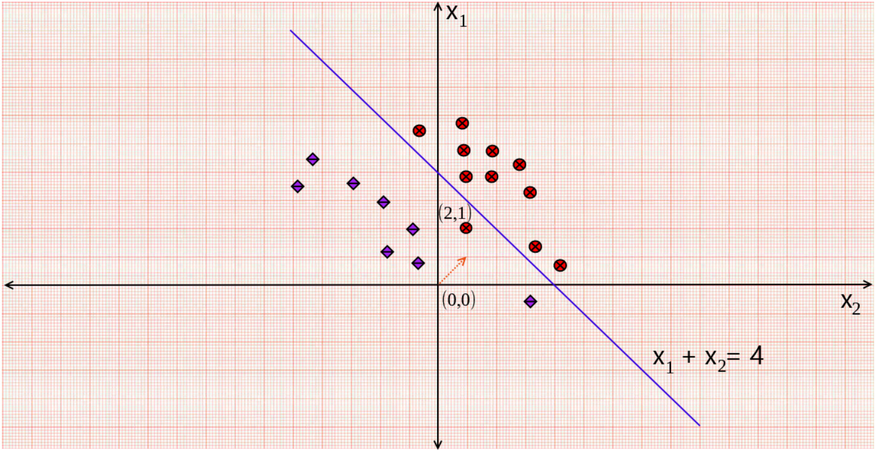


- **First assumption:** At least the data should be such that the samples with label 1 can to be separated by a hyperplane from samples with label -1.
- Is this assumption sufficient?

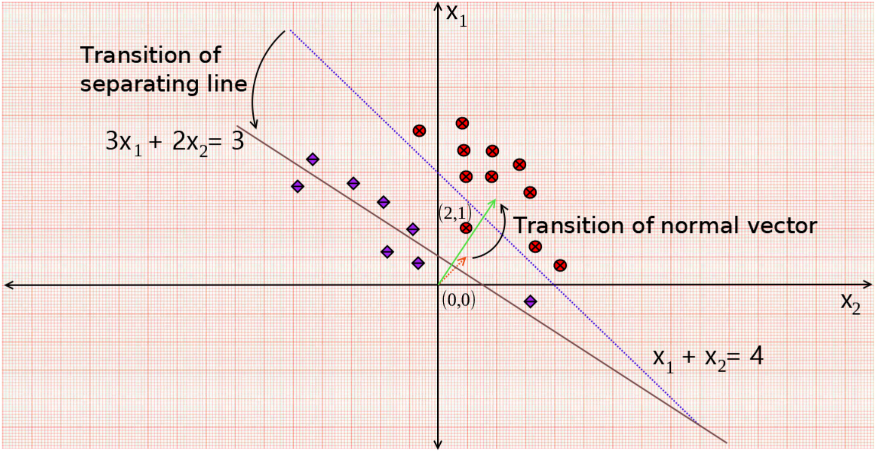
Perceptron Convergence - Geometric Intuition



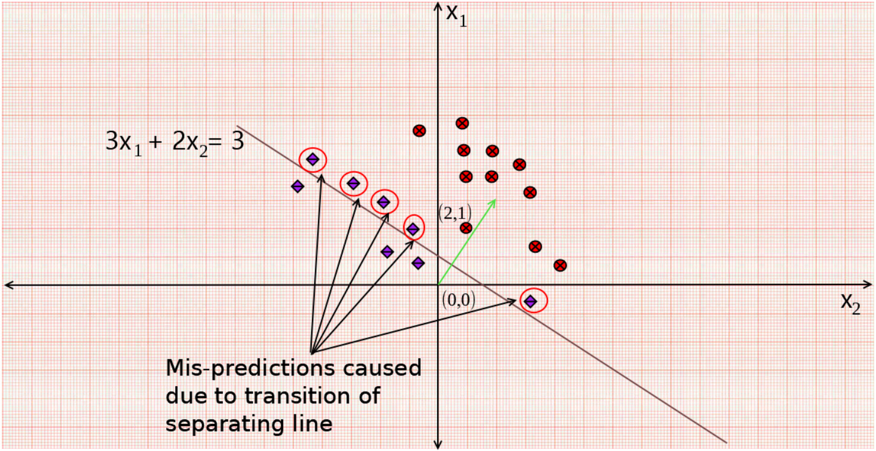
Perceptron Convergence - Geometric Intuition



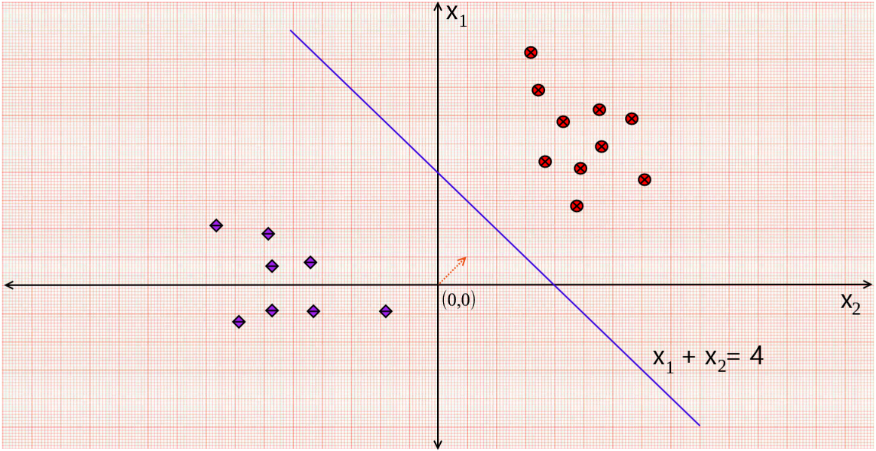
Perceptron Convergence - Geometric Intuition



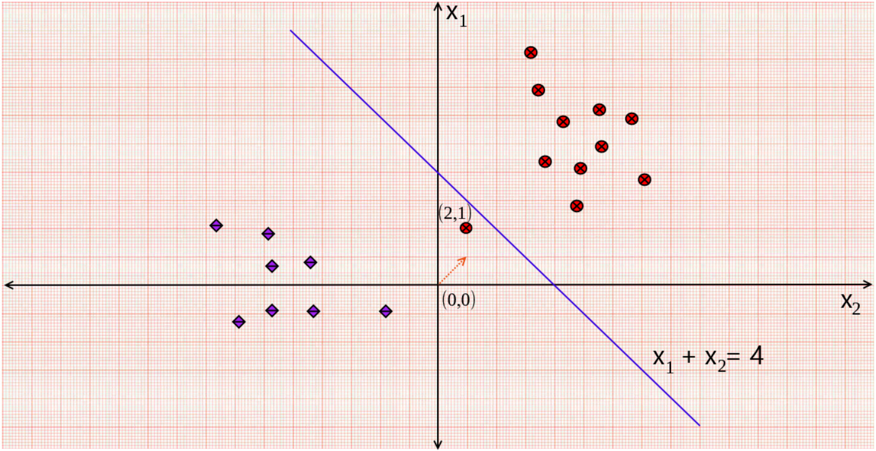
Perceptron Convergence - Geometric Intuition



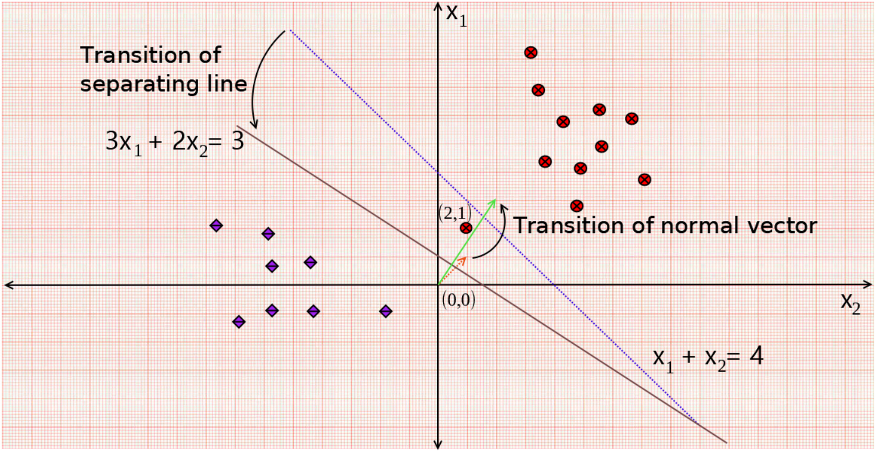
Perceptron Convergence - Geometric Intuition



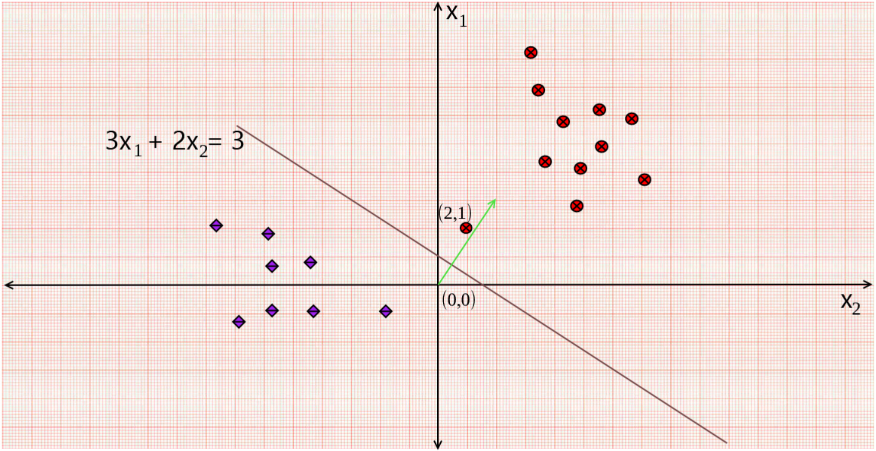
Perceptron Convergence - Geometric Intuition



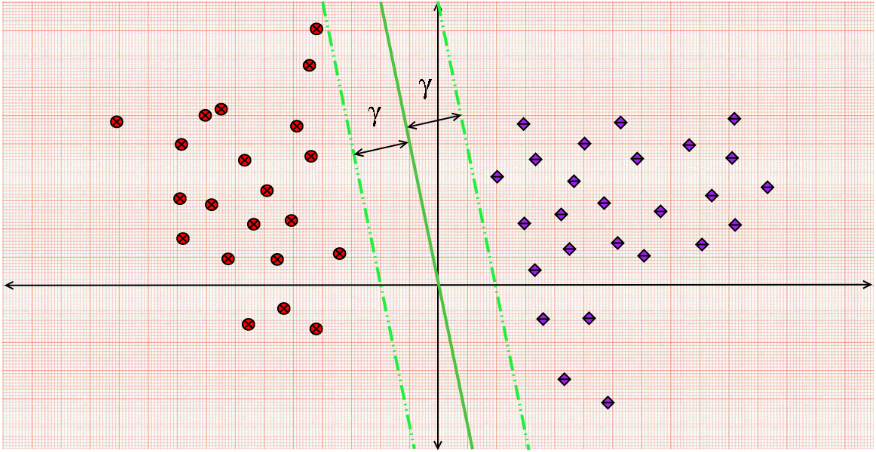
Perceptron Convergence - Geometric Intuition



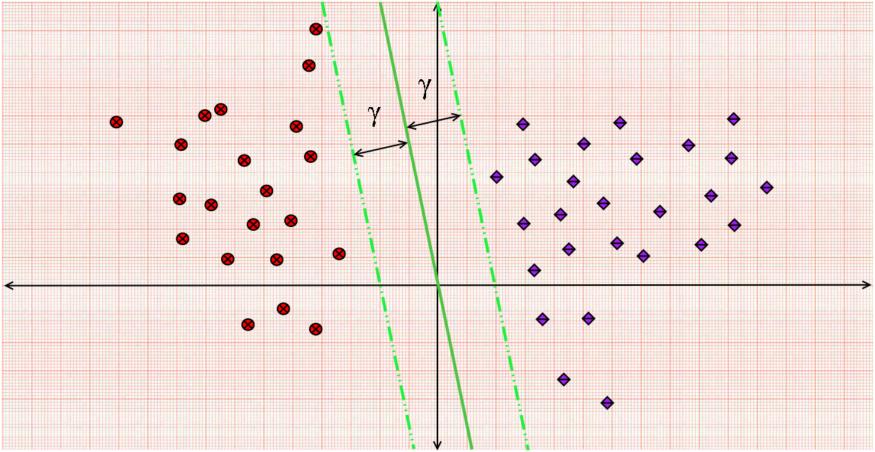
Perceptron Convergence - Geometric Intuition



Perceptron Convergence - Geometric Intuition



Perceptron Convergence - Geometric Intuition



- **Refined assumption:** We not only want the data to be separated but the separation should be good enough!

Perceptron Convergence - Separability Assumption

Linear Separability Assumption

Let $D = \{(x^t, y^t)\}_{t=1}^{\infty}$ denote the training data where $x^t \in \mathbb{R}^d$, $y^t \in \{+1, -1\}$, $\forall t = 1, 2, \ldots$. Then there exist $\mathbb{R}^d \ni w^* \neq 0,\ \gamma > 0$, such that:

$$\langle w^*, x^t \rangle > \gamma \text{ where } y^t = 1,$$
$$\langle w^*, x^t \rangle < -\gamma \text{ where } y^t = -1.$$

Perceptron Convergence - Separability Assumption

Linear Separability Assumption

Let $D = \{(x^t, y^t)\}_{t=1}^{\infty}$ denote the training data where $x^t \in \mathbb{R}^d$, $y^t \in \{+1, -1\}$, $\forall t = 1, 2, \ldots$. Then there exist $\mathbb{R}^d \ni w^* \neq 0,\ \gamma > 0$, such that:

$$y^t \langle w^*, x^t \rangle > \gamma.$$

Perceptron Convergence - Mistake Bound

- We will try to derive useful bounds on the number of mistakes that a perceptron can commit during its training.
- **Assumption on data:** Linear Separability
- Assume that T rounds of training have been completed in perceptron training. Assume T to be some large number.
- Assume that M mistakes are made by the perceptron in these T rounds. (Obviously, $M \leq T$.)
- We ask if the number of mistakes M can be bounded by some suitable quantity.

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.
- Now from linear separability assumption, we have $w^* \neq 0$ such that $y^t \langle w^*, x^t \rangle > \gamma$.

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.
- Now from linear separability assumption, we have $w^* \neq 0$ such that $y^t \langle w^*, x^t \rangle > \gamma$.
- **First step:** To bound the difference $\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$.

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.
- Now from linear separability assumption, we have $w^* \neq 0$ such that $y^t \langle w^*, x^t \rangle > \gamma$.
- Now we can write

$$\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle = \langle w^*, w^t + y^t x^t \rangle - \langle w^*, w^t \rangle$$

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.
- Now from linear separability assumption, we have $w^* \neq 0$ such that $y^t \langle w^*, x^t \rangle > \gamma$.
- Now we can write

$$\begin{aligned}\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle &= \langle w^*, w^t + y^t x^t \rangle - \langle w^*, w^t \rangle \\ &= \langle w^*, w^t \rangle + \langle w^*, y^t x^t \rangle - \langle w^*, w^t \rangle \ \text{(how?)}\end{aligned}$$

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.
- Now from linear separability assumption, we have $w^* \neq 0$ such that $y^t \langle w^*, x^t \rangle > \gamma$.
- Now we can write

$$
\begin{aligned}
\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle &= \langle w^*, w^t + y^t x^t \rangle - \langle w^*, w^t \rangle \\
&= \langle w^*, w^t \rangle + \langle w^*, y^t x^t \rangle - \langle w^*, w^t \rangle \\
&= \langle w^*, y^t x^t \rangle
\end{aligned}
$$

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \dots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.
- Now from linear separability assumption, we have $w^* \neq 0$ such that $y^t \langle w^*, x^t \rangle > \gamma$.
- Now we can write

$$\begin{aligned}
\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle &= \langle w^*, w^t + y^t x^t \rangle - \langle w^*, w^t \rangle \\
&= \langle w^*, w^t \rangle + \langle w^*, y^t x^t \rangle - \langle w^*, w^t \rangle \\
&= \langle w^*, y^t x^t \rangle \\
&= y^t \langle w^*, x^t \rangle \text{ (how?)}
\end{aligned}$$

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \dots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.
- Now from linear separability assumption, we have $w^* \neq 0$ such that $y^t \langle w^*, x^t \rangle > \gamma$.
- Now we can write

$$
\begin{aligned}
\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle &= \langle w^*, w^t + y^t x^t \rangle - \langle w^*, w^t \rangle \\
&= \langle w^*, w^t \rangle + \langle w^*, y^t x^t \rangle - \langle w^*, w^t \rangle \\
&= \langle w^*, y^t x^t \rangle \\
&= y^t \langle w^*, x^t \rangle > \gamma
\end{aligned}
$$

Perceptron Convergence - Mistake Bound

- We begin the analysis by considering an arbitrary round $t \in \{1, 2, \dots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.
- Now from linear separability assumption, we have $w^* \neq 0$ such that $y^t \langle w^*, x^t \rangle > \gamma$.
- Now we can write

$$\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle > \gamma.$$

Perceptron Convergence - Mistake Bound

- Now when no mistake is made in round t, we have $w^{t+1} = w^t$.
- Hence $\langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle = 0$.

Perceptron Convergence - Mistake Bound

Recall our assumptions:

- Assume that the T rounds of training have been completed in perceptron training. Assume T to be some large number.
- Assume that M mistakes are made by the perceptron in these T rounds. (Obviously, $M \leq T$.)

Perceptron Convergence - Mistake Bound

$$\sum_{t=1}^{T}\langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle = \sum_{\substack{t\in\{1,\ldots,T\},\\ t:\text{mistake is made}\\ \text{at round } t}} \langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle + \sum_{\substack{t\in\{1,\ldots,T\},\\ t:\text{no mistake is made}\\ \text{at round } t}} \langle w^*, w^{t+1}\rangle - \langle w^*, w^t\rangle$$

Perceptron Convergence - Mistake Bound

$$\sum_{t=1}^{T} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle = \sum_{\substack{t \in \{1,\ldots,T\}, \\ t\text{:mistake is made} \\ \text{at round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle +$$

$$\sum_{\substack{t \in \{1,\ldots,T\}, \\ t\text{:no mistake is made} \\ \text{at round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

$$= \sum_{\substack{t \in \{1,\ldots,T\}, \\ t\text{:mistake is made} \\ \text{at round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle + 0 \quad \text{(how?)}$$

Perceptron Convergence - Mistake Bound

$$\sum_{t=1}^{T} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle = \sum_{\substack{t \in \{1,\ldots,T\}, \\ t:\text{mistake is made} \\ \text{at round } t}} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle$$

$$> M\gamma \text{ (how?)}$$

Perceptron Convergence - Mistake Bound

Also note:

$$\sum_{t=1}^{T} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle = \langle w^*, w^{T+1} \rangle \text{ (homework!)}$$

Perceptron Convergence - Mistake Bound

Hence we have:

$$\sum_{t=1}^{T} \langle w^*, w^{t+1} \rangle - \langle w^*, w^t \rangle > M\gamma$$
$$\implies \langle w^*, w^{T+1} \rangle > M\gamma$$

Perceptron Mistake Bound - An upper bound

Perceptron Convergence - Mistake Bound

Now we will handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

Perceptron Convergence - Mistake Bound

Now we will handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

- From Cauchy-Schwarz inequality we have, $\langle w^*, w^{T+1} \rangle \leq \|w^*\|_2 \|w^{T+1}\|_2$. (Homework: Prove this inequality!)
- **Note:** $\|w^{T+1}\|_2$ denotes the Euclidean ℓ_2 norm of w^{T+1}.

Perceptron Convergence - Mistake Bound

Now we will handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

- From Cauchy-Schwarz inequality we have, $\langle w^*, w^{T+1} \rangle \leq \|w^*\|_2 \|w^{T+1}\|_2$. (Homework: Prove this inequality!)
- **Note:** $\|w^{T+1}\|_2$ denotes the Euclidean ℓ_2 norm of w^{T+1}.
- We will now see how to bound $\|w^{T+1}\|_2$.

Perceptron Convergence - Mistake Bound

- Again we consider an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.

Perceptron Convergence - Mistake Bound

- Again we conside an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.

Perceptron Convergence - Mistake Bound

- Again we conside an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.

Perceptron Convergence - Mistake Bound

- Again we conside an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.

Perceptron Convergence - Mistake Bound

- Again we conside an arbitrary round $t \in \{1, 2, \dots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.

Now, we have

$$\|w^{t+1}\|_2^2 = \|w^t + y^t x^t\|_2^2$$

Perceptron Convergence - Mistake Bound

- Again we conside an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.

Now, we have

$$\begin{aligned} \|w^{t+1}\|_2^2 &= \|w^t + y^t x^t\|_2^2 \\ &= \|w^t\|_2^2 + \|y^t x^t\|_2^2 + 2\langle w^t, y^t x^t \rangle \end{aligned}$$

Perceptron Convergence - Mistake Bound

- Again we conside an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.

Now, we have

$$\begin{aligned} \|w^{t+1}\|_2^2 &= \|w^t + y^t x^t\|_2^2 \\ &= \|w^t\|_2^2 + \|y^t x^t\|_2^2 + 2\langle w^t, y^t x^t \rangle \\ &= \|w^t\|_2^2 + \|x^t\|_2^2 + 2y^t \langle w^t, x^t \rangle \end{aligned}$$

Perceptron Convergence - Mistake Bound

- Again we conside an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.

Now, we have

$$\begin{aligned} \|w^{t+1}\|_2^2 &= \|w^t + y^t x^t\|_2^2 \\ &= \|w^t\|_2^2 + \|y^t x^t\|_2^2 + 2\langle w^t, y^t x^t \rangle \\ &= \|w^t\|_2^2 + \|x^t\|_2^2 + 2y^t \langle w^t, x^t \rangle \\ \implies \|w^{t+1}\|_2^2 &\leq \|w^t\|_2^2 + \|x^t\|_2^2 \text{ (How?)} \end{aligned}$$

Perceptron Convergence - Mistake Bound

- Again we conside an arbitrary round $t \in \{1, 2, \ldots, T\}$ where a mistake is made by the perceptron.
- **Recall:** During this t-th round:
  - Perceptron computes $\hat{y}^t = \text{sign}(\langle w^t, x^t \rangle)$.
  - $\hat{y}^t \neq y^t$.
  - Perceptron update: $w^{t+1} = w^t + y^t x^t$.

Now, we have

$$\begin{aligned}
\|w^{t+1}\|_2^2 &= \|w^t + y^t x^t\|_2^2 \\
&= \|w^t\|_2^2 + \|y^t x^t\|_2^2 + 2\langle w^t, y^t x^t \rangle \\
&= \|w^t\|_2^2 + \|x^t\|_2^2 + 2 y^t \langle w^t, x^t \rangle \\
\implies \|w^{t+1}\|_2^2 &\leq \|w^t\|_2^2 + \|x^t\|_2^2 \text{ (How?)}
\end{aligned}$$

Thus $\|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq \|x^t\|_2^2$.

Perceptron Convergence - Mistake Bound

Assumption on boundedness of $\|x^t\|_2$

We shall assume further that $\forall t = 1, 2, \ldots,$ the ℓ_2 norm (or length) of x^t is bounded, which is denoted as:

$$\|x^t\|_2 \leq R \quad \forall t = 1, 2, \ldots$$

Perceptron Convergence - Mistake Bound

Assumption on boundedness of $\|x^t\|_2$

We shall assume further that $\forall t = 1, 2, \ldots,$ the ℓ_2 norm (or length) of x^t is bounded, which is denoted as:

$$\|x^t\|_2 \leq R \quad \forall t = 1, 2, \ldots$$

- This is yet another assumption to help our analysis.
- Bounded $\|x^t\|_2$ is not very unrealistic, however finding a suitable value for R might be difficult.

Perceptron Convergence - Mistake Bound

Assumption on boundedness of $\|x^t\|_2$

We shall assume further that $\forall t = 1, 2, \ldots,$ the ℓ_2 norm (or length) of x^t is bounded, which is denoted as:

$$\|x^t\|_2 \leq R \quad \forall t = 1, 2, \ldots$$

- This is yet another assumption to help our analysis.
- Bounded $\|x^t\|_2$ is not very unrealistic, however finding a suitable value for R might be difficult.
- This is where normalizing all x^t might help, so that $\|x^t\|_2 \leq 1$ can be assumed.
- **Note:** The set $\{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$ is called a unit ball in $\mathbb{R}^d$.

Perceptron Convergence - Mistake Bound

We thus have

$$\|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq \|x^t\|_2^2 \implies \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq R^2.$$

Perceptron Convergence - Mistake Bound

We thus have

$$\|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq \|x^t\|_2^2 \implies \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq R^2.$$

Again, summing $\|w^{t+1}\|_2^2 - \|w^t\|_2^2$ over $t = 1, \ldots, T$ we get

Perceptron Convergence - Mistake Bound

We thus have

$$\|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq \|x^t\|_2^2 \implies \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq R^2.$$

Again, summing $\|w^{t+1}\|_2^2 - \|w^t\|_2^2$ over $t = 1, \dots, T$ we get

$$\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 = \sum_{\substack{t \in \{1,\dots,T\}, \\ t:\text{mistake is made} \\ \text{at round } t}} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 + \sum_{\substack{t \in \{1,\dots,T\}, \\ t:\text{no mistake is made} \\ \text{at round } t}} \|w^{t+1}\|_2^2 - \|w^t\|_2^2$$

Perceptron Convergence - Mistake Bound

We thus have

$$\|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq \|x^t\|_2^2 \implies \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq R^2.$$

Again, summing $\|w^{t+1}\|_2^2 - \|w^t\|_2^2$ over $t = 1, \dots, T$ we get

$$\begin{aligned}\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 &= \sum_{\substack{t\in\{1,\dots,T\},\\ t\text{:mistake is made}\\ \text{at round } t}} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 + \\ &\quad \sum_{\substack{t\in\{1,\dots,T\},\\ t\text{:no mistake is made}\\ \text{at round } t}} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \\ &= \sum_{\substack{t\in\{1,\dots,T\},\\ t\text{:mistake is made}\\ \text{at round } t}} \|w^{t+1}\|_2^2 - \|w^t\|_2^2\end{aligned}$$

Perceptron Convergence - Mistake Bound

We thus have

$$\|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq \|x^t\|_2^2 \implies \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq R^2.$$

Again, summing $\|w^{t+1}\|_2^2 - \|w^t\|_2^2$ over $t = 1, \dots, T$ we get

$$\begin{aligned}
\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 &= \sum_{\substack{t\in\{1,\dots,T\},\\ t\text{:mistake is made}\\ \text{at round } t}} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 + \\
&\quad \sum_{\substack{t\in\{1,\dots,T\},\\ t\text{:no mistake is made}\\ \text{at round } t}} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \\
&= \sum_{\substack{t\in\{1,\dots,T\},\\ t\text{:mistake is made}\\ \text{at round } t}} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq MR^2 \text{ (How?)}
\end{aligned}$$

Perceptron Convergence - Mistake Bound

Thus we have

$$\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq MR^2.$$

Perceptron Convergence - Mistake Bound

Thus we have

$$\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq MR^2.$$

On the other hand we get:

$$\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 = \|w^{T+1}\|_2^2.$$ (Homework!)

Perceptron Convergence - Mistake Bound

Thus we have

$$\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq MR^2.$$

On the other hand we get:

$$\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 = \|w^{T+1}\|_2^2. \text{ (Homework!)}$$

Combining both, we get

$$\|w^{T+1}\|_2^2 \leq MR^2.$$

Perceptron Convergence - Mistake Bound

Thus we have

$$\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 \leq MR^2.$$

On the other hand we get:

$$\sum_{i=1}^{T} \|w^{t+1}\|_2^2 - \|w^t\|_2^2 = \|w^{T+1}\|_2^2. \text{ (Homework!)}$$

Combining both, we get

$$\|w^{T+1}\|_2^2 \leq MR^2.$$

Thus we have bounded $\|w^{T+1}\|_2$.

Perceptron Convergence - Mistake Bound

Recall: We wanted to handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

Perceptron Convergence - Mistake Bound

Recall: We wanted to handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

Then using Cauchy-Schwarz inequality we had

$$M\gamma < \langle w^*, w^{T+1} \rangle \leq \|w^*\|_2 \|w^{T+1}\|_2$$

Perceptron Convergence - Mistake Bound

Recall: We wanted to handle the inner product term:

$$\langle w^*, w^{T+1}\rangle > M\gamma$$

Then using Cauchy-Schwarz inequality we had

$$\begin{aligned} M\gamma < \langle w^*, w^{T+1}\rangle &\leq \|w^*\|_2 \|w^{T+1}\|_2 \\ \implies M^2\gamma^2 &< \|w^*\|_2^2 \|w^{T+1}\|_2^2 \end{aligned}$$

Perceptron Convergence - Mistake Bound

Recall: We wanted to handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

Then using Cauchy-Schwarz inequality we had

$$\begin{aligned} M\gamma < \langle w^*, w^{T+1} \rangle &\leq \|w^*\|_2 \|w^{T+1}\|_2 \\ \implies M^2\gamma^2 &< \|w^*\|_2^2 \|w^{T+1}\|_2^2 \end{aligned}$$

Using the bound $\|w^{T+1}\|_2^2 \leq MR^2$ we obtain:

Perceptron Convergence - Mistake Bound

Recall: We wanted to handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

Then using Cauchy-Schwarz inequality we had

$$\begin{aligned} M\gamma < \langle w^*, w^{T+1} \rangle \leq \|w^*\|_2 \|w^{T+1}\|_2 \\ \implies M^2\gamma^2 < \|w^*\|_2^2 \|w^{T+1}\|_2^2 \end{aligned}$$

Using the bound $\|w^{T+1}\|_2^2 \leq MR^2$ we obtain:

$$M^2\gamma^2 < \|w^*\|_2^2 MR^2$$

Perceptron Convergence - Mistake Bound

Recall: We wanted to handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

Then using Cauchy-Schwarz inequality we had

$$\begin{aligned} M\gamma < \langle w^*, w^{T+1} \rangle &\leq \|w^*\|_2 \|w^{T+1}\|_2 \\ \implies M^2\gamma^2 &< \|w^*\|_2^2 \|w^{T+1}\|_2^2 \end{aligned}$$

Using the bound $\|w^{T+1}\|_2^2 \leq MR^2$ we obtain:

$$\begin{aligned} M^2\gamma^2 &< \|w^*\|_2^2 MR^2 \\ \implies M &< \frac{\|w^*\|_2^2 R^2}{\gamma^2} \end{aligned}$$

Perceptron Convergence - Mistake Bound

Recall: We wanted to handle the inner product term:

$$\langle w^*, w^{T+1} \rangle > M\gamma$$

Then using Cauchy-Schwarz inequality we had

$$M\gamma < \langle w^*, w^{T+1} \rangle \leq \|w^*\|_2 \|w^{T+1}\|_2$$
$$\implies M^2\gamma^2 < \|w^*\|_2^2 \|w^{T+1}\|_2^2$$

Using the bound $\|w^{T+1}\|_2^2 \leq MR^2$ we obtain:

$$M^2\gamma^2 < \|w^*\|_2^2 MR^2$$
$$\implies M < \frac{\|w^*\|_2^2 R^2}{\gamma^2}$$

Thus, assuming that $\|w^*\|_2$ and R can be controlled, the number of mistakes M is inversely proportional to γ, which determines the closeness of the data points to the separating hyperplane.

Perceptron Convergence - Mistake Bound

References:

- **H.D. Block**: The perceptron: A model for brain functioning. *Reviews of Modern Physics* 34, 123-135 (1962).
- **A.B.J. Novikoff**: On convergence proofs on perceptrons. In: *Proceedings of the Symposium on the Mathematical Theory of Automata*, vol. XII, pp. 615-622 (1962).

Perceptron Convergence - Mistake Bound

Two questions remain:

Perceptron Convergence - Mistake Bound

Two questions remain:

- How do we compute w^* and γ?
- What is the intuition behind the Perceptron update rule?

Perceptron Convergence - Mistake Bound

Two questions remain:

- How do we compute w^* and γ?
- What is the intuition behind the Perceptron update rule?

Try to find answers to these questions!