Deep Learning: Class 1

Binary Perceptrons

Deep Learning

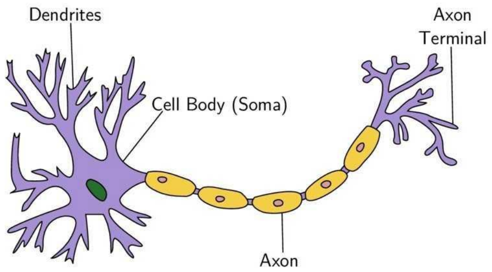
- The model does not require handcrafted features or attributes.
- Given the samples $\{x_1, x_2, .., x_n\}$ and the labels $\{y_1, y_2, .., y_n\}$ The algorithm learns to how to classify the input data correctly.
- No decision to make such as coding algorithmic states for the bayesian networks or reinforcement learning.

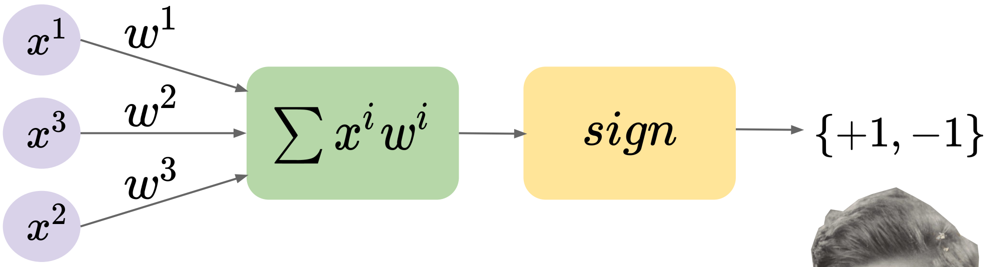
The Perceptron

- Invented by Frank Rosenblatt in the last 50s. Frank studied and later taught cognitive psychology in Cornell.

The Perceptron

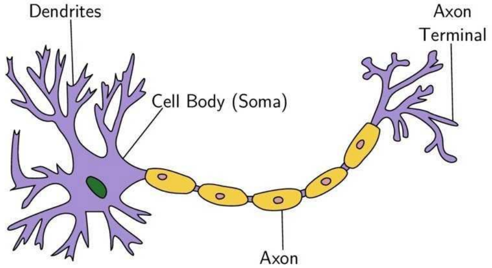
- Invented by Frank Rosenblatt in the last 50s. Frank studied and later taught cognitive psychology in Cornell.
- A model of the human brain and not simply a classification algorithm.

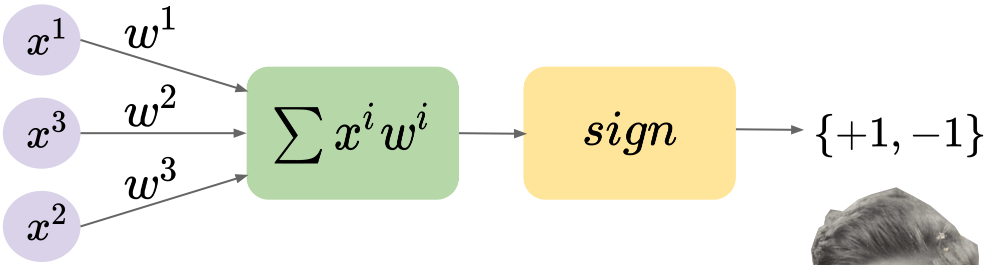




The Perceptron

- Invented by Frank Rosenblatt in the last 50s. Frank studied and later taught cognitive psychology in Cornell.
- A model of the human brain and not simply a classification algorithm.

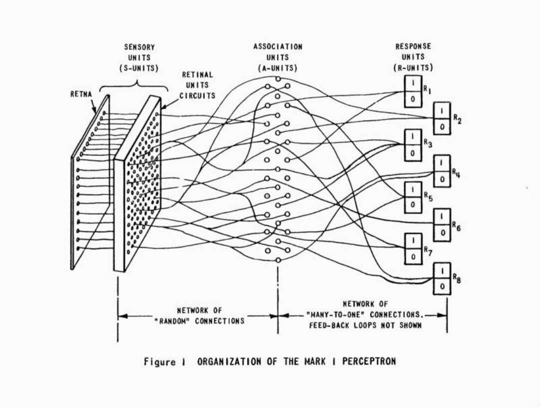


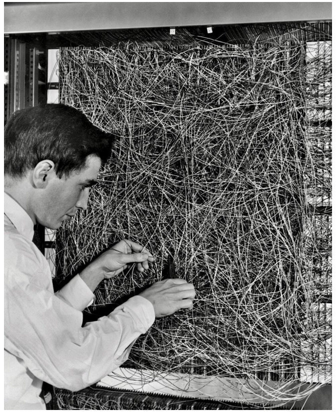


- NYT 7 JUL 1958: “The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence”

The Perceptron

- Invented by Frank Rosenblatt in the last 50s. Frank studied and later taught cognitive psychology in Cornell.
- A model of the human brain and not simply a classification algorithm.
- NYT 7 JUL 1958: “The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence”






Books of The Times
ALGERIA
by Germaine Tillion
ELIZABETH CADELL
SHADOWS ON THE WATER
An END to VALOR
by Philip Van Doren Stern
ONCE TO SINAI
CROSSWORD PUZZLE
Ask Any Girl

NYT, Jul 7, 1958


DECEMBER 6, 1958

New Yorker, Dec 6, 1958

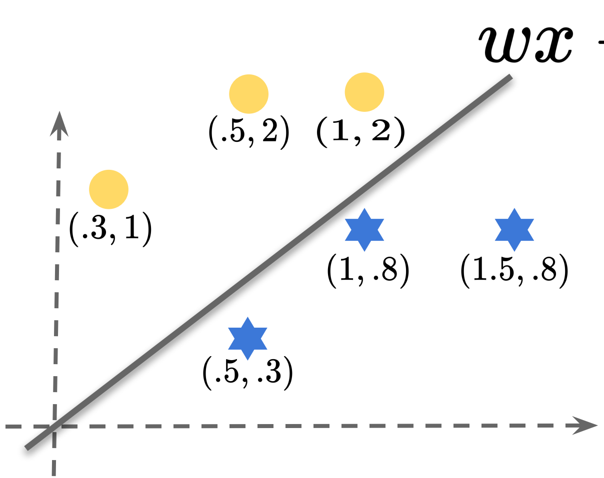
The Perceptron: Algebra

- Given class (yellow circle / blue star) find w, b that define the hyperplane to separate the two classes.
- The hyperplane is defined by as all the points such that:

$$wx + b = [\sum x^i w^i] + b = 0$$

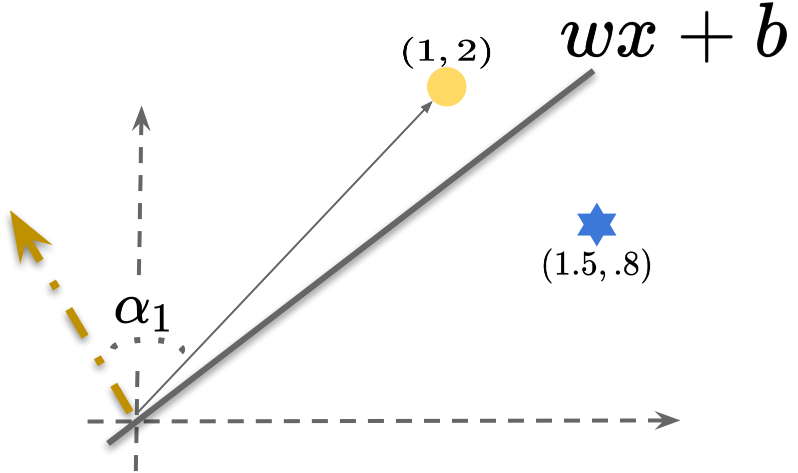
- In this example:

$$w = (1, -1)$$
$$b = 0$$



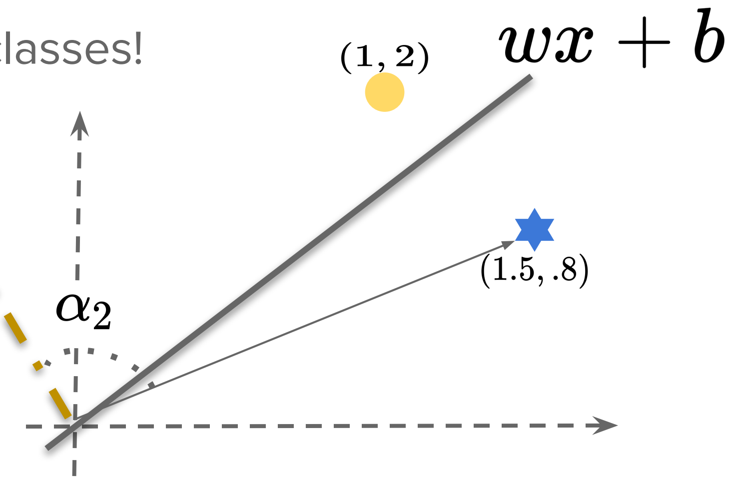
The Perceptron: Algebra

- Remember $x \cdot w = \sum x^i w^i = |x| \cdot |w| \cdot cos(\alpha)$
- W is orthogonal (90 degrees) to the hyperplane $w = (1, -1)$
- So for everything above the hyperplane (yellow dots) the angle is smaller than 90 degrees!
- The dot product is positive!



The Perceptron: Algebra

- Remember $x \cdot w = \sum x^i w^i = |x| \cdot |w| \cdot cos(\alpha)$
- W is orthogonal (90 degrees) to the hyperplane $w = (1, -1)$
- So for everything below the hyperplane (yellow dots) the angle is bigger than 90 degrees!
- The dot product is negative!
- Taking the sign separates the classes!

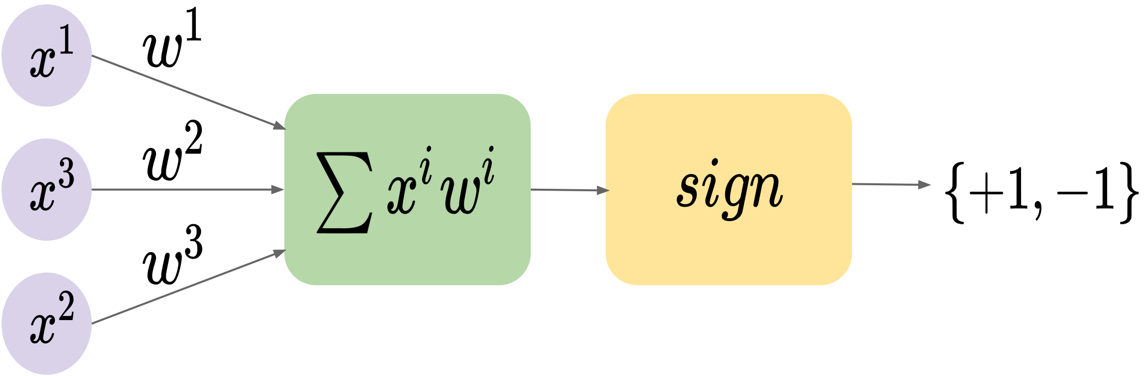


The Perceptron: W?

- According to Rosenblatt the perceptron can learn the right weights by initializing them to 0 and iteratively updating w when a sample is misclassified.

$$w_{k+1} = w_k + y \cdot x$$

$$y \in \{-1, 1\}$$

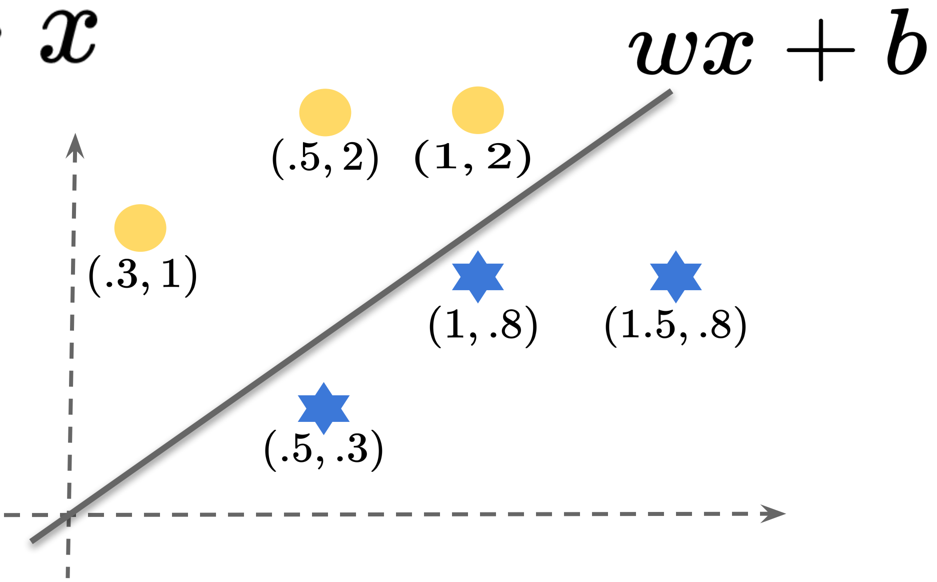


The Perceptron: W?

- According to Rosenblatt the perceptron can learn the right weights by initializing them to 0 and iteratively updating w when a sample is misclassified.

$$w_{k+1} = w_k + y \cdot x$$

$$y \in \{-1, 1\}$$



The Perceptron: b?

- In the next few slides we will forget about b and our equation becomes $sign(w \cdot x)$
- This is achieved via a simple transformation, for each $x = (x^1, x^2)$ we define $\hat{x} = (x^1, x^2, 1)$ and for $w = (w^1, w^2)$ we define $\hat{w} = (w^1, w^2, b)$
- Now instead of looking for a 2 dimensional hyperplane $wx + b$ we are looking for a 3 dimensional hyperplane $\hat{w}\hat{x}$.

The Perceptron: But does it work?

- Assume that the points are actually separable by a hyperplane. Meaning that there is a w^* for which $\forall j$ it holds that $w^* \cdot x_j = y_j$. How can we show the perceptron actually learns to classify the input correctly?

- Let k be an iteration in which some x_i was misclassified, meaning

$$w \cdot x_i \cdot y_i = -1$$

we want to show that is k bounded!

The Perceptron: But does it work?

- Let k be an iteration in which some x_i was misclassified:

$$w_{k+1} = w_k + y \cdot x$$
$$w^* w_{k+1} = w^* w_k + y \cdot w^* x \geq w^* w_k + 1$$
$$w^* \cdot w_{k+1} \geq w^* w_k + 1$$

Remember that $w_0 = \hat{0}$ hence:

$$w^* w_1 \geq 1$$
$$w^* w_2 \geq 2 \quad \Rightarrow |w_k| \geq \frac{k}{|w^*|}$$
$$\ldots$$
$$w^* w_k \geq k$$

The Perceptron: But does it work?

- Let k be an iteration in which some x_i was misclassified:

$$|w_{k+1}|^2 = |w_k + y_i x_i|^2 = |w_k|^2 + |x_i y_i|^2 + 2 w_k x_i y_i$$

$$|w_{k+1}|^2 = |w_k|^2 + |x_i|^2 + 2(w_k x_i) y_i$$

$$|w_{k+1}|^2 = |w_k|^2 + |x_i|^2 - 2 \leq |w_k|^2 + |x_i|^2$$

But x_i is from our data and is always bounded (for instance we can normalize)

Thus we get

$$|w_{k+1}|^2 \leq |w_k|^2 + R^2$$

$$|w_1|^2 \leq R^2$$

$$|w_2|^2 \leq 2 \cdot R^2$$

$$\dots$$

$$\Rightarrow |w_k|^2 \leq k \cdot R^2$$

The Perceptron: But does it work?

- Let k be an iteration in which some x_i was misclassified. We get that:

$$\frac{k^2}{|w^*|^2} \leq |w_k|^2 \leq k \cdot R^2$$

And therefore:

$$k \leq R^2 \cdot |w^*|^2$$

Meaning the perceptron converges after a bounded number of updates!

The Perceptron: Assignment!

- Before next class code the update in python. Pull from the github repository for this class and finish the perceptron class.
- The data is from the sklearn package we previously used. it is a breast cancer data from the Wisconsin med school, so every sample is a vector of 30 measurements concerning the shape and size of a tumor, and the label tells us if it's cancerous or not.
- Each of the 569 patients in this data set actually recovered so don't let it get you down.
- My implementation gets over 93% accuracy.