Linear Methods 1

Shivaram Kalyanakrishnan

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

February 2023

This Lecture

- Welcome, introduction
- The supervised learning problem
- Perceptron: classification with a linear model

This Lecture

- Welcome, introduction
- The supervised learning problem
- Perceptron: classification with a linear model

This Lecture

- Welcome, introduction
- The supervised learning problem
- Perceptron: classification with a linear model

Examples

- A bank has the following table.

Loan amount	Duration	Annual income	Age	Employment sector	Paid back?

Examples

- A bank has the following table.

Loan amount	Duration	Annual income	Age	Employment sector	Paid back?

For an ongoing loan, the bank wants to predict whether it will be paid back.

Examples

- A bank has the following table.

Loan amount	Duration	Annual income	Age	Employment sector	Paid back?

  For an ongoing loan, the bank wants to predict whether it will be paid back.

- A trading firm has recorded daily opening and closing prices of stocks, along with other relevant information/news, from the last 5 years.

Examples

- A bank has the following table.

Loan amount	Duration	Annual income	Age	Employment sector	Paid back?

For an ongoing loan, the bank wants to predict whether it will be paid back.

- A trading firm has recorded daily opening and closing prices of stocks, along with other relevant information/news, from the last 5 years.

The firm would like to get a good estimate of . . . tomorrow's opening stock price!

Examples

- A bank has the following table.

Loan amount	Duration	Annual income	Age	Employment sector	Paid back?

  For an ongoing loan, the bank wants to predict whether it will be paid back.

- A trading firm has recorded daily opening and closing prices of stocks, along with other relevant information/news, from the last 5 years.

  The firm would like to get a good estimate of . . . tomorrow's opening stock price!

- A website has millions of images in which objects have been labeled.

Examples

- A bank has the following table.

Loan amount	Duration	Annual income	Age	Employment sector	Paid back?

  For an ongoing loan, the bank wants to predict whether it will be paid back.

- A trading firm has recorded daily opening and closing prices of stocks, along with other relevant information/news, from the last 5 years.

  The firm would like to get a good estimate of . . . tomorrow's opening stock price!

- A website has millions of images in which objects have been labeled.

  When a new image is uploaded, the website want to know what objects it contains.

Examples

- A bank has the following table.

Loan amount	Duration	Annual income	Age	Employment sector	Paid back?

  For an ongoing loan, the bank wants to predict whether it will be paid back.

- A trading firm has recorded daily opening and closing prices of stocks, along with other relevant information/news, from the last 5 years.

  The firm would like to get a good estimate of . . . tomorrow's opening stock price!

- A website has millions of images in which objects have been labeled.

  When a new image is uploaded, the website want to know what objects it contains.

  What should the bank, trading firm, and website do?

Examples

- A bank has the following table.

Loan amount	Duration	Annual income	Age	Employment sector	Paid back?

  For an ongoing loan, the bank wants to predict whether it will be paid back.

- A trading firm has recorded daily opening and closing prices of stocks, along with other relevant information/news, from the last 5 years.

  The firm would like to get a good estimate of . . . tomorrow's opening stock price!

- A website has millions of images in which objects have been labeled.

  When a new image is uploaded, the website want to know what objects it contains.

  What should the bank, trading firm, and website do?

  Learn a predictive model.

Supervised Learning Problem

- **Input**: Data set $D = \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\}$.

Supervised Learning Problem

- **Input**: Data set $D = \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\}$.
  Each x^i is an array (or vector) of features $x^i_1, x^i_2, \ldots, x^i_d$.

Supervised Learning Problem

- **Input**: Data set $D = \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\}$.
  Each x^i is an array (or vector) of features $x_1^i, x_2^i, \ldots, x_d^i$.
  y^i is the label (or supervision) corresponding to x^i.

Supervised Learning Problem

- **Input**: Data set $D = \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\}$.
  Each x^i is an array (or vector) of features $x^i_1, x^i_2, \ldots, x^i_d$.
  y^i is the label (or supervision) corresponding to x^i.
  Output: Model.

$$D \longrightarrow \boxed{\text{Supervised Learning Algorithm}} \longrightarrow \text{Model.}$$

Supervised Learning Problem

- **Input**: Data set $D = \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\}$.
  Each x^i is an array (or vector) of features $x^i_1, x^i_2, \ldots, x^i_d$.
  y^i is the label (or supervision) corresponding to x^i.
  Output: Model.

$$D \longrightarrow \boxed{\text{Supervised Learning Algorithm}} \longrightarrow \text{Model.}$$

- Using the model on new x (not necessarily present in D) to predict y.

$$x \longrightarrow \boxed{\text{Model}} \longrightarrow y.$$

Feature Types

Features and labels are usually of three main types.

Feature Types

Features and labels are usually of three main types.

- Categorical (or nominal) features take discrete values with no particular ordering.
  Employment sector: { Healthcare, Business, Sports, Transportation, . . . }.
  Colour choice: { Yellow, White, Magenta}.

Feature Types

Features and labels are usually of three main types.

- Categorical (or nominal) features take discrete values with no particular ordering.
  Employment sector: { Healthcare, Business, Sports, Transportation, . . . }.
  Colour choice: { Yellow, White, Magenta}.

- Numeric (or real-valued) features take numeric values.
  Age
  Loan amount

Feature Types

Features and labels are usually of three main types.

- Categorical (or nominal) features take discrete values with no particular ordering.
  Employment sector: { Healthcare, Business, Sports, Transportation, ... }.
  Colour choice: { Yellow, White, Magenta}.

- Numeric (or real-valued) features take numeric values.
  Age
  Loan amount

- Ordinal features take discrete values that can be ordered.
  Rate your call quality!: $\{\star, \quad \star\star, \quad \star\star\star, \quad \star\star\star\star, \quad \star\star\star\star\star\}$
  Day of this course: $\{1, 2, 3, 4, 5\}$.

Feature Types

Features and labels are usually of three main types.

- Categorical (or nominal) features take discrete values with no particular ordering.
  Employment sector: { Healthcare, Business, Sports, Transportation, . . . }.
  Colour choice: { Yellow, White, Magenta}.

- Numeric (or real-valued) features take numeric values.
  Age
  Loan amount

- Ordinal features take discrete values that can be ordered.
  Rate your call quality!: $\{\star, \quad \star\star, \quad \star\star\star, \quad \star\star\star\star, \quad \star\star\star\star\star\}$
  Day of this course: $\{1, 2, 3, 4, 5\}$.

- If the class label is categorical, the task is a classification task.
  If the class label is numeric, the task is a regression task.
  For now, we'll assume all features are numeric. Reasonable assumption?

Questions

- What are some other examples of supervised learning problems?

Questions

- What are some other examples of supervised learning problems?
  - On-line ads: prediction of click-through rate.
  - Tagging people using face recognition in pictures.
  - Credit card fraud discovery.
  - Chances of rain tomorrow.

Questions

- What are some other examples of supervised learning problems?
  - On-line ads: prediction of click-through rate.
  - Tagging people using face recognition in pictures.
  - Credit card fraud discovery.
  - Chances of rain tomorrow.
- Does it take a lot of coding to implement a supervised learning algorithm?

Questions

- What are some other examples of supervised learning problems?
  - On-line ads: prediction of click-through rate.
  - Tagging people using face recognition in pictures.
  - Credit card fraud discovery.
  - Chances of rain tomorrow.

- Does it take a lot of coding to implement a supervised learning algorithm?

  No. There are many good, open-source implementations to choose from. But it will still take time to analyse and iterate and refine.

Questions

- What are some other examples of supervised learning problems?
  - On-line ads: prediction of click-through rate.
  - Tagging people using face recognition in pictures.
  - Credit card fraud discovery.
  - Chances of rain tomorrow.

- Does it take a lot of coding to implement a supervised learning algorithm?

  No. There are many good, open-source implementations to choose from. But it will still take time to analyse and iterate and refine.

- What are some challenges involved in supervised learning?

Questions

- What are some other examples of supervised learning problems?
  - On-line ads: prediction of click-through rate.
  - Tagging people using face recognition in pictures.
  - Credit card fraud discovery.
  - Chances of rain tomorrow.

- Does it take a lot of coding to implement a supervised learning algorithm?

  No. There are many good, open-source implementations to choose from. But it will still take time to analyse and iterate and refine.

- What are some challenges involved in supervised learning?

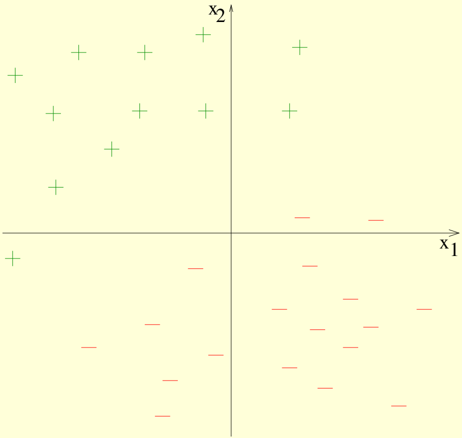
  Preparing the data set, handling missing/noisy fields, class imbalance, etc.

This Lecture

- Welcome, introduction
- The supervised learning problem
- Perceptron: classification with a linear model

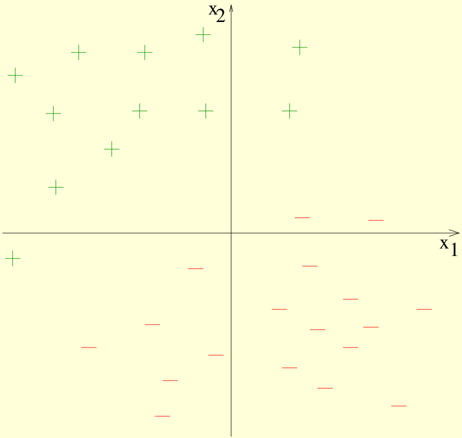
Perceptron Learning Algorithm

- Illustration with $d = 2$ features (x_1, x_2), 2 classes ("$+$" or 1, "$-$" or -1).



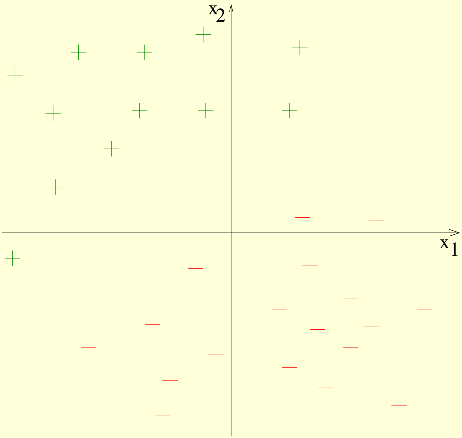
Perceptron Learning Algorithm

- Illustration with $d = 2$ features (x_1, x_2), 2 classes ("$+$" or 1, "$-$" or -1).
- Assumption: Data is linearly separable, by a line passing through the origin.

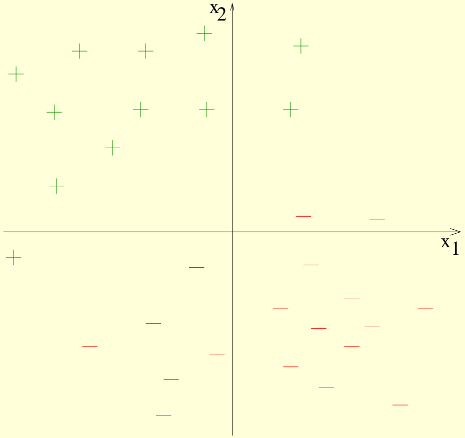


Perceptron Learning Algorithm

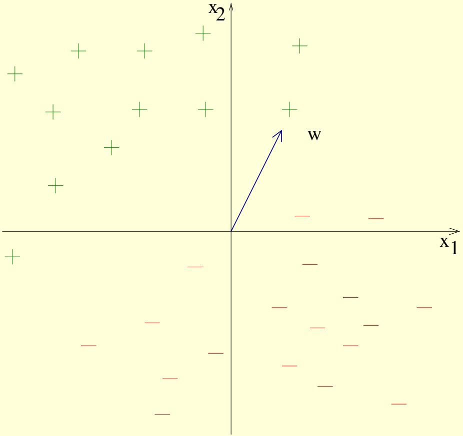
- Illustration with $d = 2$ features (x_1, x_2), 2 classes ("$+$" or 1, "$-$" or -1).
- Assumption: Data is linearly separable, by a line passing through the origin.
- How to find such a line (in higher dimensions, a hyperplane)?



First, How to *Represent* an Origin-passing Hyperplane?

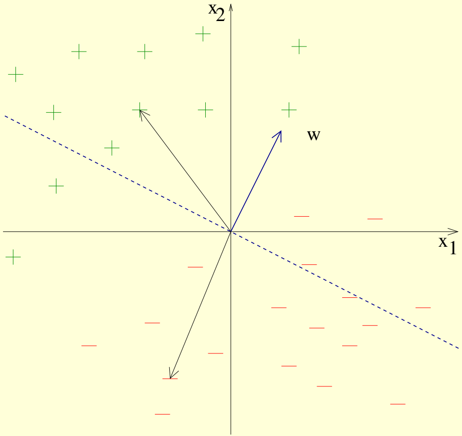


First, How to *Represent* an Origin-passing Hyperplane?



By a vector $w = (w_1, w_2)$.

First, How to *Represent* an Origin-passing Hyperplane?

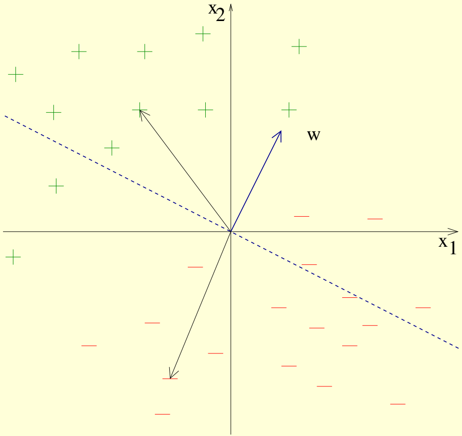


By a vector $w = (w_1, w_2)$.

For point x,

- assign label "+" if $w \cdot x \geq 0$;
- assign label "$-$" otherwise.

First, How to *Represent* an Origin-passing Hyperplane?



By a vector $w = (w_1, w_2)$.

For point x,

- assign label "+" if $w \cdot x \geq 0$;
- assign label "−" otherwise.

Now how to find a satisfying w?

Perceptron Learning Algorithm

Initialise w arbitrarily. (Recall that it is a d-dimensional vector.)

While there is some misclassified point:

Select an arbitrary misclassified point (x, y).
//That means $y = 1$ but $w \cdot x < 0$, or $y = -1$ but $w \cdot x \geq 0$.

Set $w \leftarrow w + yx$.

Return w.

Perceptron Learning Algorithm

Initialise w arbitrarily. (Recall that it is a d-dimensional vector.)

While there is some misclassified point:

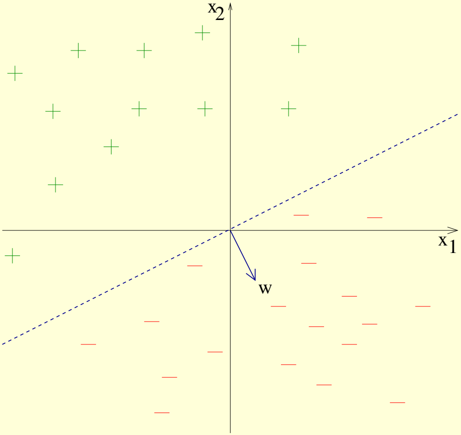
Select an arbitrary misclassified point (x, y).
//That means $y = 1$ but $w \cdot x < 0$, or $y = -1$ but $w \cdot x \geq 0$.

Set $w \leftarrow w + yx$.

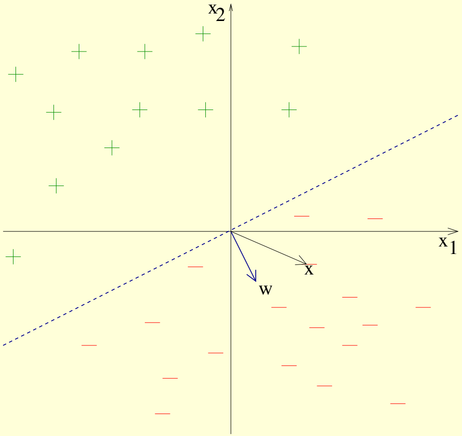
Return w.

That's it!

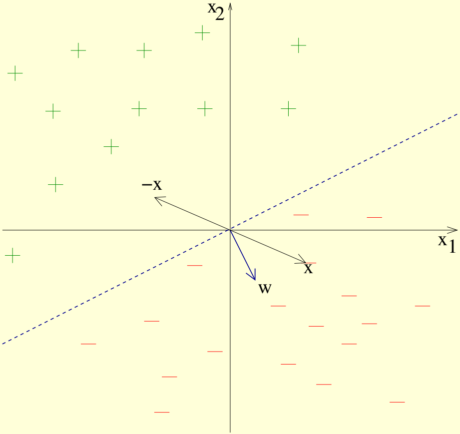
Step-through



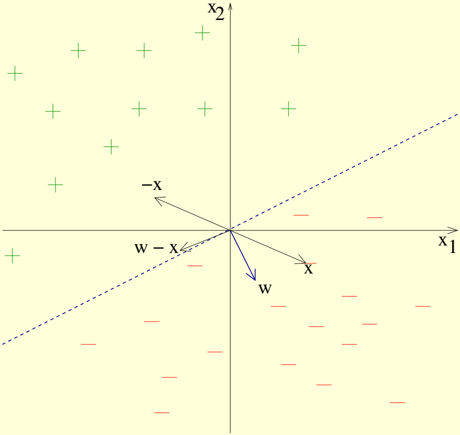
Step-through



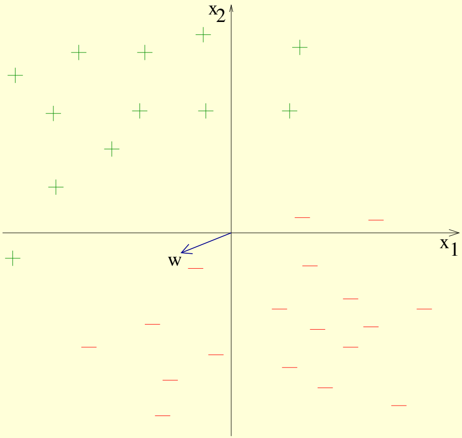
Step-through



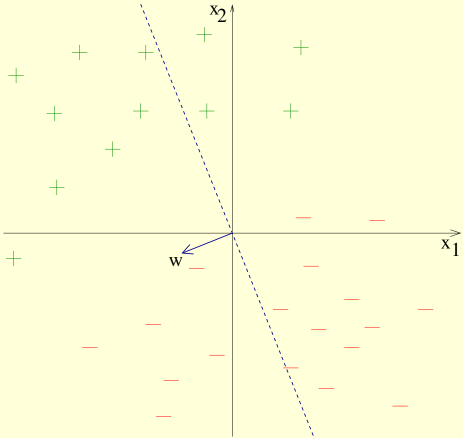
Step-through



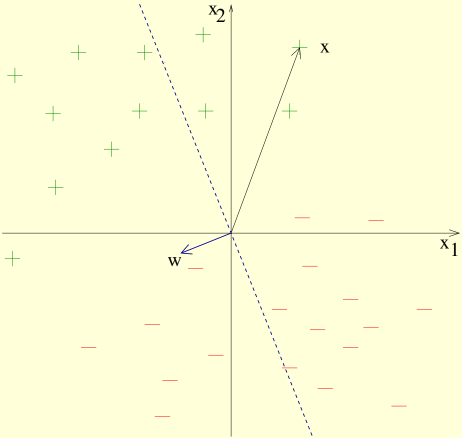
Step-through



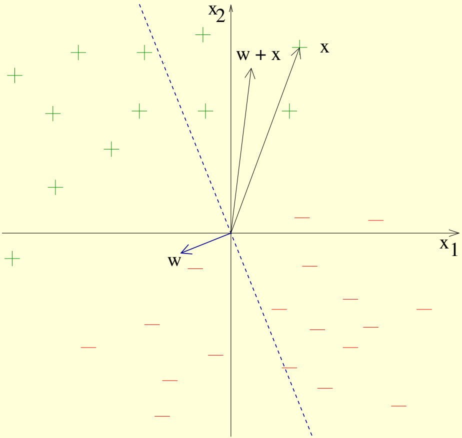
Step-through



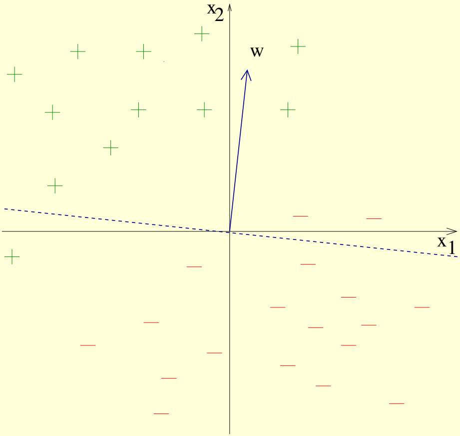
Step-through



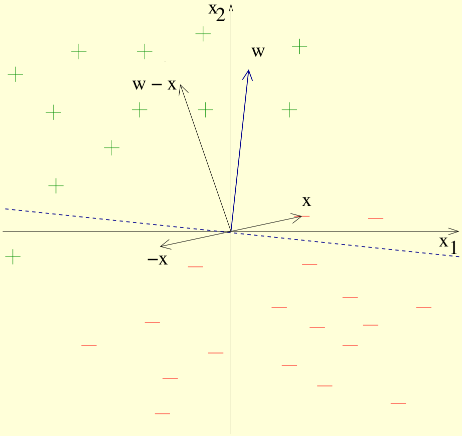
Step-through



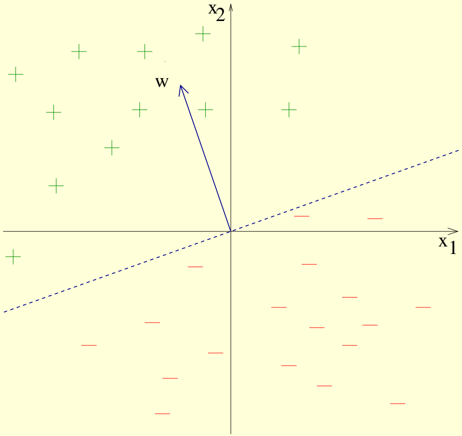
Step-through



Step-through



Step-through



Questions

- Will the algorithm necessarily terminate?

Questions

- Will the algorithm necessarily terminate?

  Yes, if indeed the data set is linearly separable by an origin-passing hyperplane.

Questions

- Will the algorithm necessarily terminate?
  Yes, if indeed the data set is linearly separable by an origin-passing hyperplane.

- What if the points are not linearly separable?

Questions

- Will the algorithm necessarily terminate?

  Yes, if indeed the data set is linearly separable by an origin-passing hyperplane.

- What if the points are not linearly separable?

  The algorithm will not terminate!

Questions

- Will the algorithm necessarily terminate?

  Yes, if indeed the data set is linearly separable by an origin-passing hyperplane.

- What if the points are not linearly separable?

  The algorithm will not terminate!

- If a point is classified correctly in iteration i, can it get *misclassified* in iteration $i + 1$?

Questions

- Will the algorithm necessarily terminate?

  Yes, if indeed the data set is linearly separable by an origin-passing hyperplane.

- What if the points are not linearly separable?

  The algorithm will not terminate!

- If a point is classified correctly in iteration i, can it get *misclassified* in iteration $i + 1$?

  Yes.

Questions

- Will the algorithm necessarily terminate?

  Yes, if indeed the data set is linearly separable by an origin-passing hyperplane.

- What if the points are not linearly separable?

  The algorithm will not terminate!

- If a point is classified correctly in iteration i, can it get *misclassified* in iteration $i + 1$?

  Yes.

- How to adapt the algorithm for more than two classes?

Questions

- Will the algorithm necessarily terminate?

  Yes, if indeed the data set is linearly separable by an origin-passing hyperplane.

- What if the points are not linearly separable?

  The algorithm will not terminate!

- If a point is classified correctly in iteration i, can it get *misclassified* in iteration $i + 1$?

  Yes.

- How to adapt the algorithm for more than two classes?

  Find out! (About “one-versus-rest” models.)

Questions

- Will the algorithm necessarily terminate?

  Yes, if indeed the data set is linearly separable by an origin-passing hyperplane.

- What if the points are not linearly separable?

  The algorithm will not terminate!

- If a point is classified correctly in iteration i, can it get *misclassified* in iteration $i + 1$?

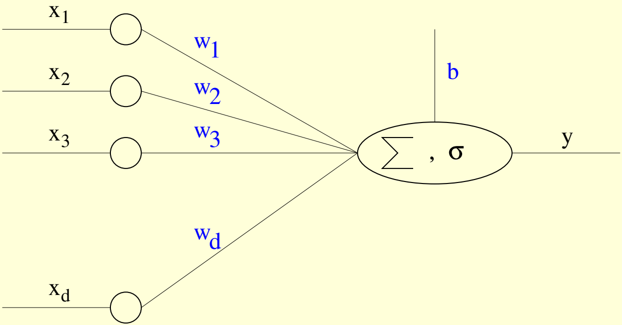
  Yes.

- How to adapt the algorithm for more than two classes?

  Find out! (About “one-versus-rest” models.)

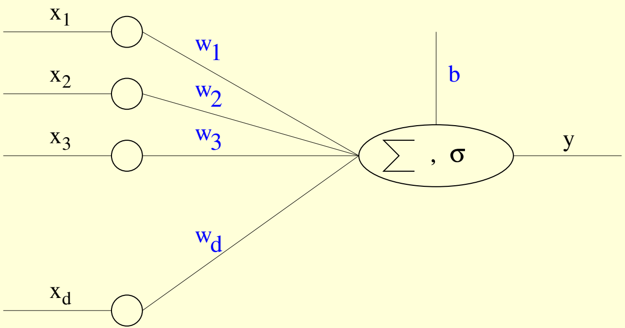
- Why is it called a *Perceptron*? (Is it related to electrons and neutrons?!)

Perceptron (Rosenblatt, 1957)



- An early version was built in hardware.

Perceptron (Rosenblatt, 1957)



- An early version was built in hardware.
- $y = \text{sign}(w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + b)$, where

$$\text{sign}(\alpha) = \begin{cases} 1 & \text{if } \alpha \geq 0 \\ -1 & \text{otherwise.} \end{cases}$$

b is the "bias", which we had assumed to be 0, but which is also easy to learn.

References

- Chapter 4, **A Course in Machine Learning**, Hal Daumé III. Available on-line at `http://ciml.info/`.