ONLINE MASTERS IN **DATA SCIENCE**

DSC 255 - MACHINE LEARNING FUNDAMENTALS

THE PERCEPTRON ALGORITHM

SANJOY DASGUPTA, PROFESSOR

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

Recall: Linear Separators for Binary Classification

Binary classification problem: data $x \in \mathbb{R}^d$ and labels $y \in \{-1, +1\}$

Linear Classifier:

- Parameters: $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$
- Decision boundary $w \cdot x + b = 0$
- On point x, predict label $sign(w \cdot x + b)$

If the true label on point x is y:

- Classifier correct if $y = sign(w \cdot x + b)$
- Equivalent: if $y(w \cdot x + b) > 0$

A Loss Function for Classification

What is the **loss** of our linear classifier (given by w, b) on a point (x, y)?

One idea for a loss function:

- If $y(w \cdot x + b) > 0$: correct, no loss
- If $y(w \cdot x + b) < 0$: loss $-y(w \cdot x + b)$

A Simple Learning Algorithm

Fit a linear classifier w, b to the training set using **stochastic gradient descent.**

- Update w, b based on just one data point (x, y) at a time
- If $y(w \cdot x + b) > 0$: zero loss, no update
- If $y(w \cdot x + b) \leq 0$: loss is $-y(w \cdot x + b)$

A Simple Learning Algorithm

Fit a linear classifier w, b to the training set using **stochastic gradient descent.**

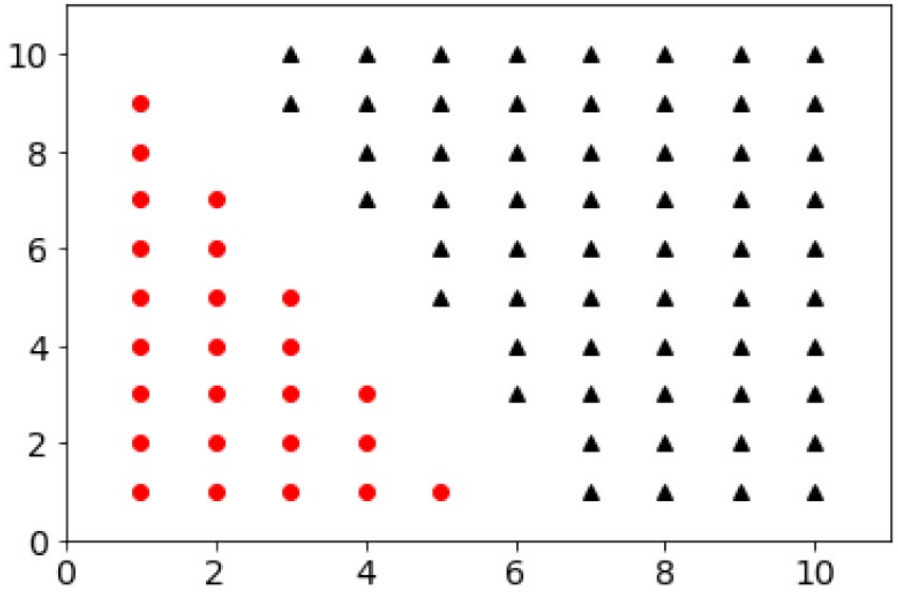
- Update w, b based on just one data point (x, y) at a time
- If $y(w \cdot x + b) > 0$: zero loss, no update
- If $y(w \cdot x + b) \leq 0$: loss is $-y(w \cdot x + b)$

The Perceptron Algorithm

- Initialize $w = 0$ and $b = 0$
- Keep cycling through the training data (x, y):
  - If $y(w \cdot x + b) \leq 0$ (i.e., point misclassified):
    - $w = w + yx$
    - $b = b + y$

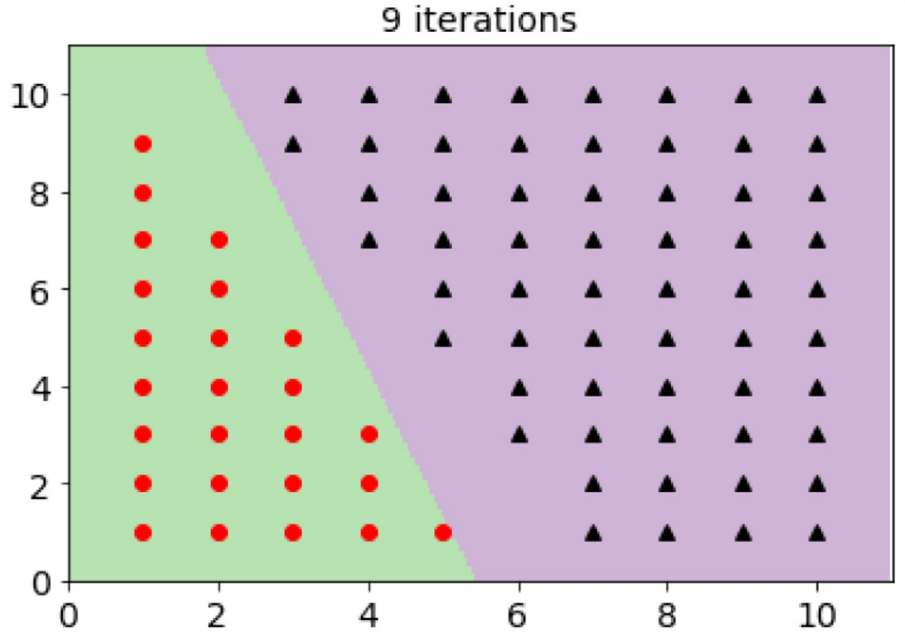
The Perceptron in Action

85 data points, linearly separable.



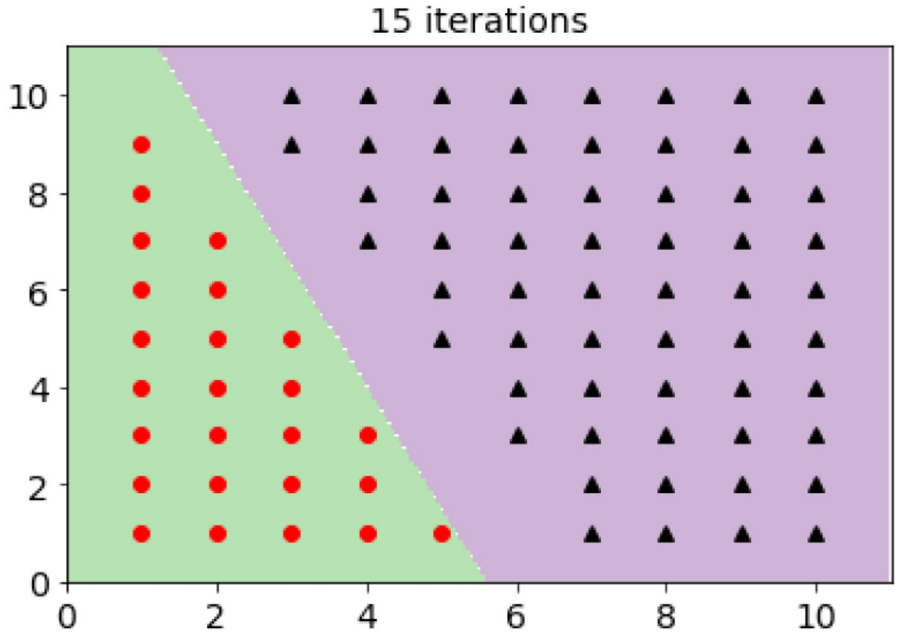
The Perceptron in Action

85 data points, linearly separable.



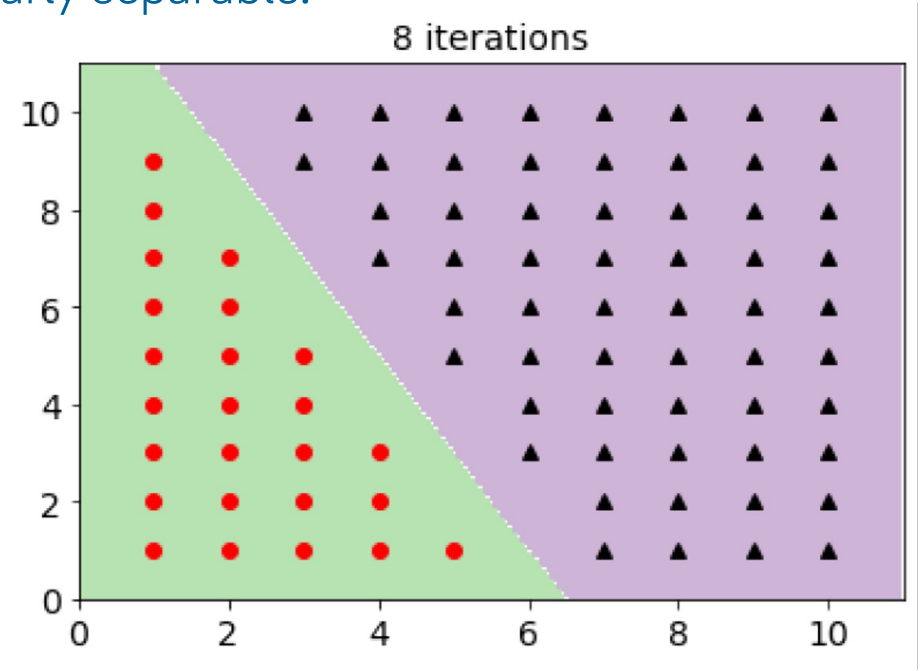
The Perceptron in Action

85 data points, linearly separable.



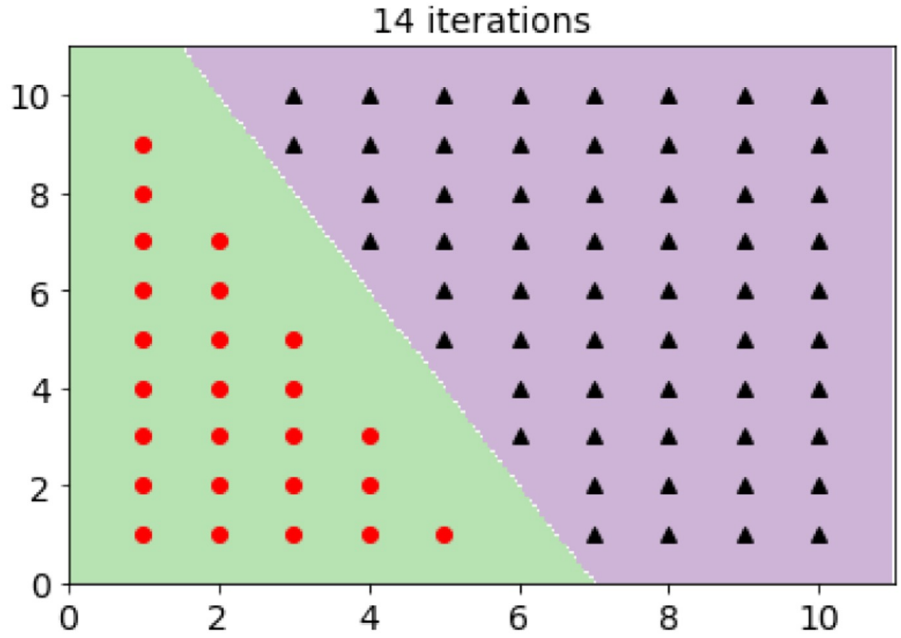
The Perceptron in Action

85 data points, linearly separable.



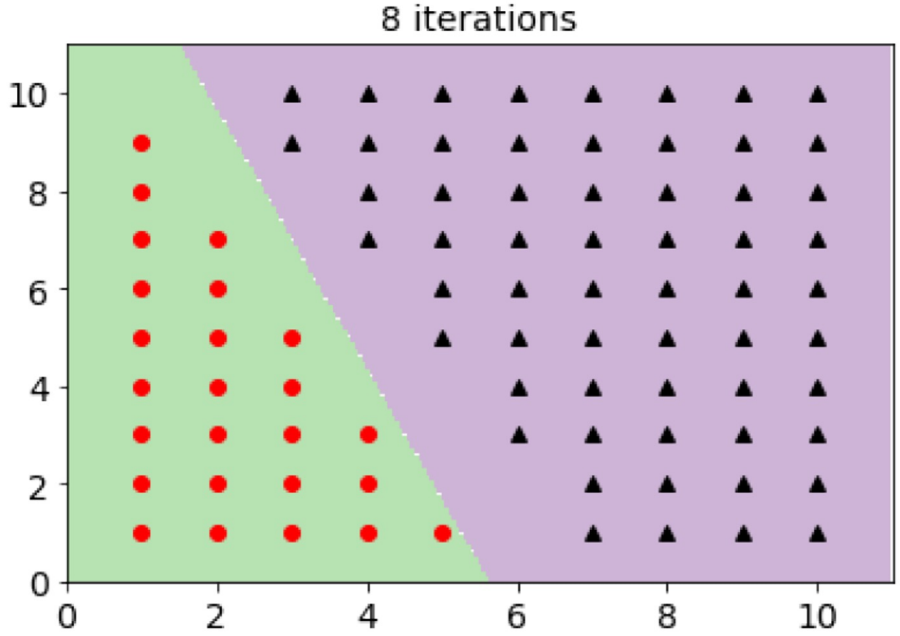
The Perceptron in Action

85 data points, linearly separable.



The Perceptron in Action

85 data points, linearly separable.



Perceptron: Convergence

If the training data is linearly separable:

- The Perceptron algorithm will and a linear classifier with zero training error.
- It will converge within a finite number of steps.
- The number of iterations can be bounded in terms of the *margin*: roughly, a measure of the space between the two classes.