CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019

Outline for March 18

- Recap common issues on Midterm 1
- Finish AdaBoost
  - Handout 8
  - Overview (Handout 9)
  - Decision trees with weighted examples
- Begin Support Vector Machines (SVMs)
  - Linearly separable datasets
  - Perceptron

- **Fill out partner form for Lab 5**
- **Office hours today 12:30-2pm**

Outline for March 18

- Recap common issues on Midterm 1
- Finish AdaBoost
  - Handout 8
  - Overview (Handout 9)
  - Decision trees with weighted examples
- Begin Support Vector Machines (SVMs)
  - Linearly separable datasets
  - Perceptron

Midterm 1 Curve

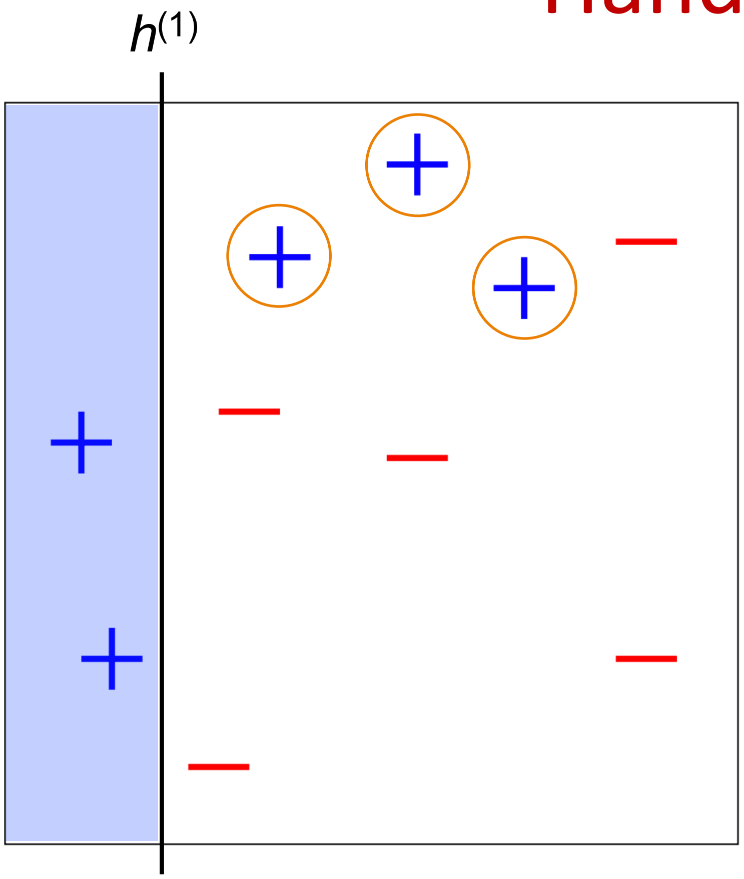
- A: 88-100
- B: 76-87
- C: 64-75
- D: 52-63
- Below 52: not passing (please meet with me)

Midterm 1 Solutions
(not posted in slides)

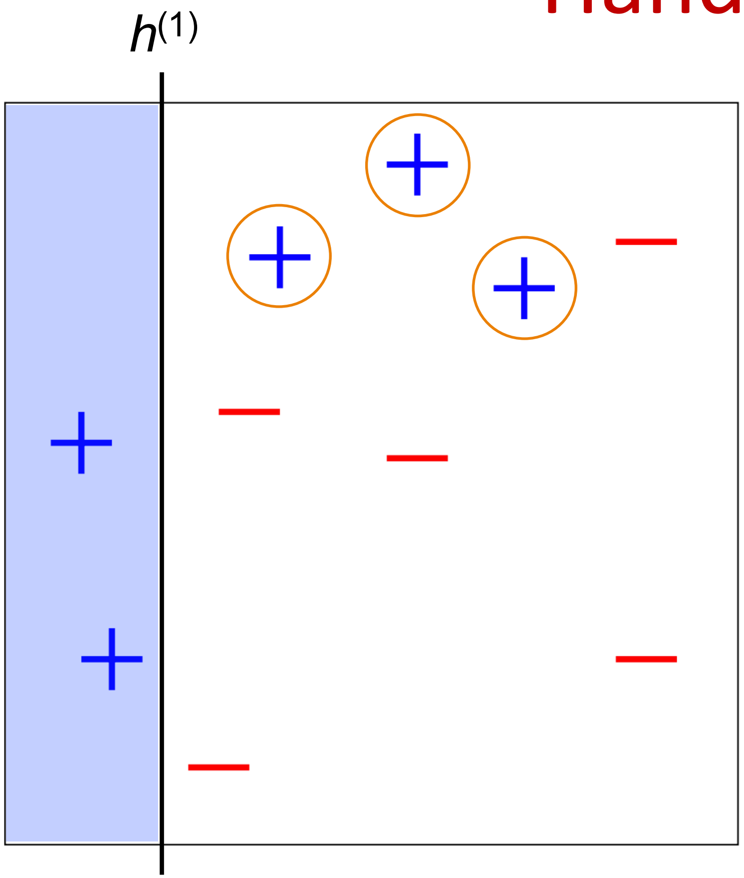
Outline for March 18

- Recap common issues on Midterm 1
- Finish AdaBoost
  - Handout 8
  - Overview (Handout 9)
  - Decision trees with weighted examples
- Begin Support Vector Machines (SVMs)
  - Linearly separable datasets
  - Perceptron

Handout 8: Round 1



Handout 8: Round 1



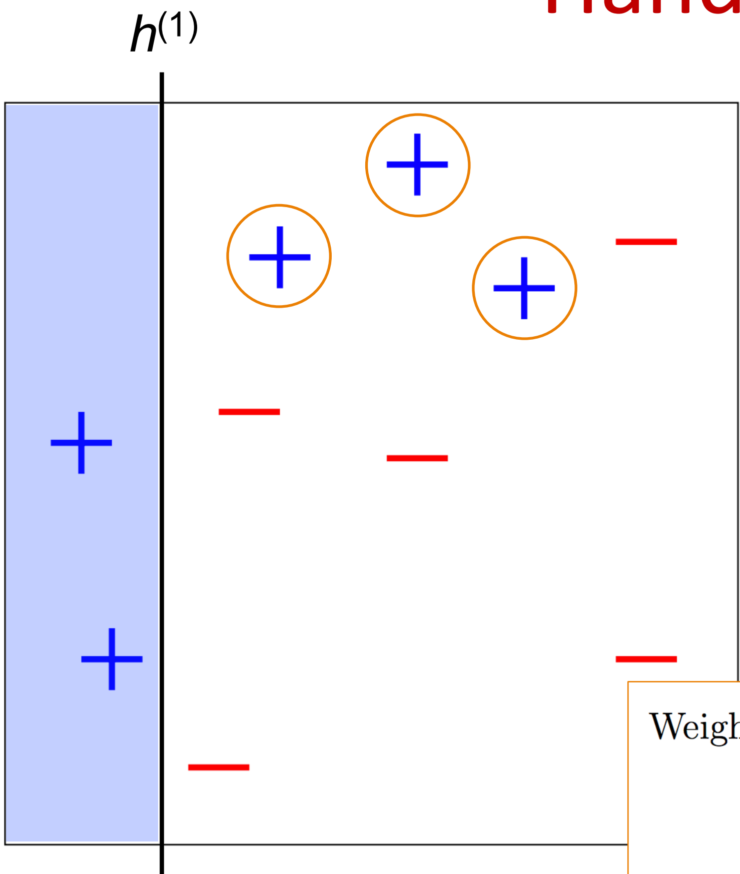
$w_i^{(1)} = \frac{1}{10}$ for all $i = 1, 2, \cdots, 10$.

$\epsilon_1 = \frac{3}{10}$ (three points incorrectly classified, all with weight $\frac{1}{10}$)

$\alpha_1 = \frac{1}{2} \ln \left(\frac{1 - \frac{3}{10}}{\frac{3}{10}} \right) = \ln \sqrt{\frac{7}{3}} \approx 0.42$

- correctly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left(- \ln \sqrt{\frac{7}{3}} \right)$
- incorrectly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left(\ln \sqrt{\frac{7}{3}} \right)$

Handout 8: Round 1



$w_i^{(1)} = \frac{1}{10}$ for all $i = 1, 2, \cdots, 10$.

$\epsilon_1 = \frac{3}{10}$ (three points incorrectly classified, all with weight $\frac{1}{10}$)

$\alpha_1 = \frac{1}{2} \ln \left(\frac{1-\frac{3}{10}}{\frac{3}{10}} \right) = \ln \sqrt{\frac{7}{3}} \approx 0.42$

- correctly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp\left(-\ln \sqrt{\frac{7}{3}}\right)$
- incorrectly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp\left(\ln \sqrt{\frac{7}{3}}\right)$

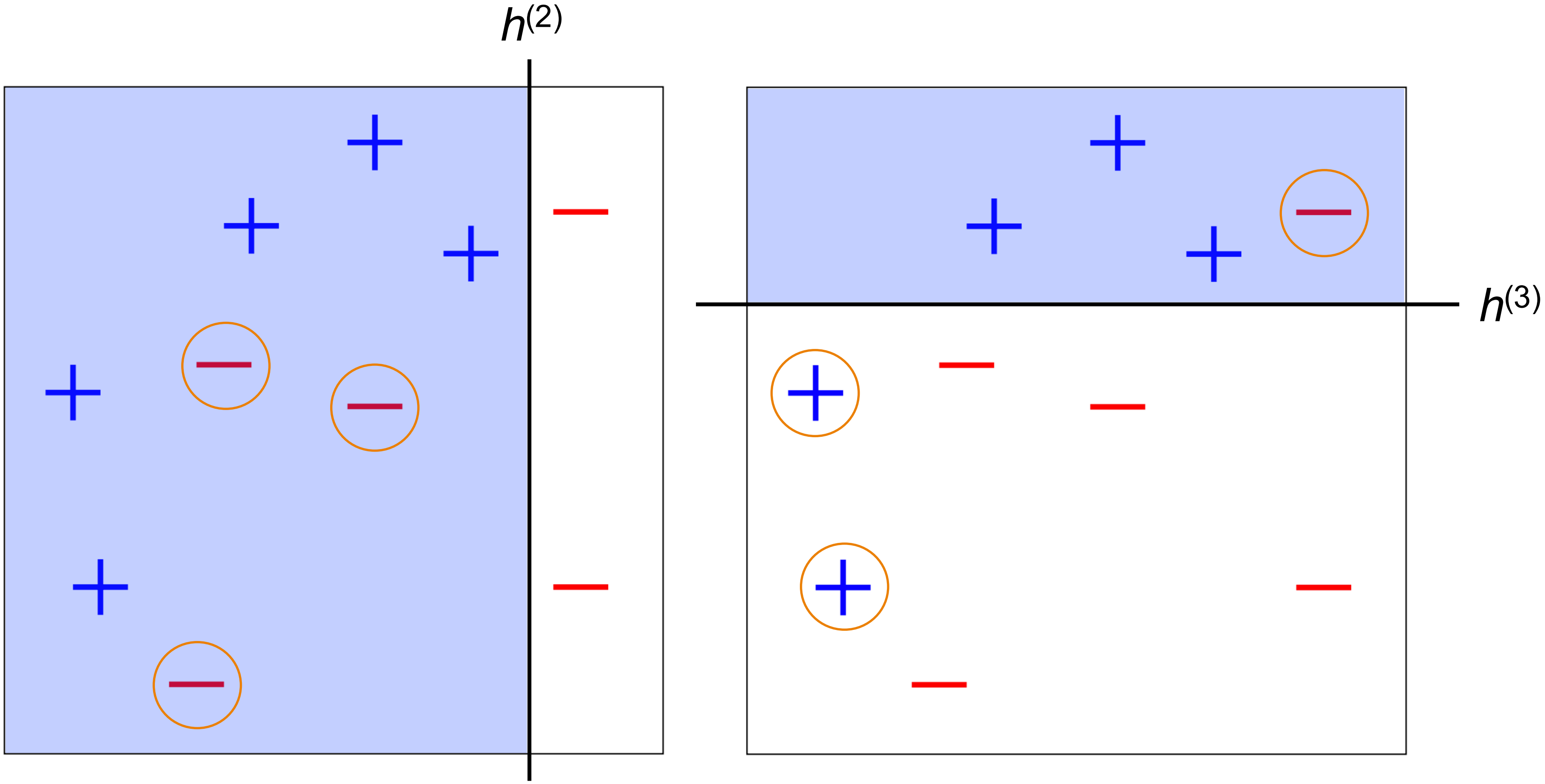
Weights must sum to 1, $\Rightarrow$

$$7 \cdot \frac{c_1}{10} \exp\left(-\ln \sqrt{\frac{7}{3}}\right) + 3 \cdot c_1 \cdot \frac{1}{10} \exp\left(\ln \sqrt{\frac{7}{3}}\right) = 1$$

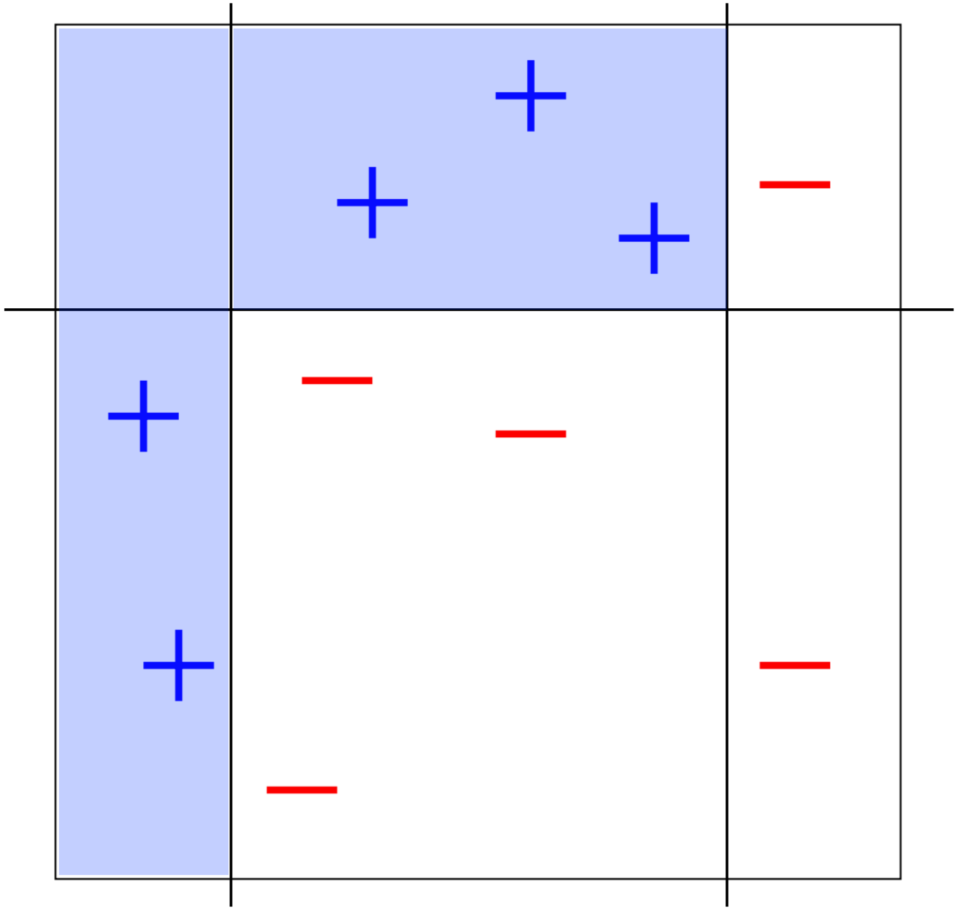
$$\Rightarrow \quad c_1 = \frac{5}{\sqrt{21}}$$

- correctly classified: $w_i^{(2)} = \frac{5}{\sqrt{21}} \cdot \frac{1}{10} \sqrt{\frac{3}{7}} = \frac{1}{14}$ decrease!
- incorrectly classified: $w_i^{(2)} = \frac{5}{\sqrt{21}} \cdot \frac{1}{10} \sqrt{\frac{7}{3}} = \frac{1}{6}$ increase!

Handout 8: Round 2 & 3 (exercise!)



Handout 8: final classifier



$$h(\boldsymbol{x}) = \operatorname{sign}\Big(0.42 \cdot h^{(1)}(\boldsymbol{x}) + 0.65 \cdot h^{(2)}(\boldsymbol{x}) + 0.92 \cdot h^{(3)}(\boldsymbol{x})\Big)$$

Outline for March 18

- Recap common issues on Midterm 1
- Finish AdaBoost
  - Handout 8
  - Overview (Handout 9)
  - Decision trees with weighted examples
- Begin Support Vector Machines (SVMs)
  - Linearly separable datasets
  - Perceptron

Decision Trees with Weighted Examples

Entropy

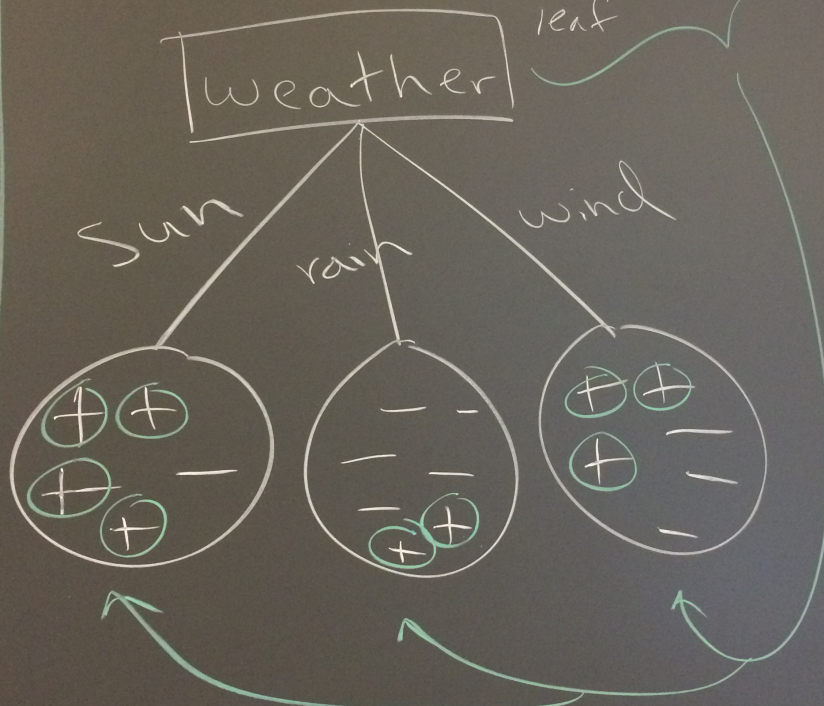
There should be a minus sign here!

$$H(Y|X_j=v) = -\sum_{c \in \text{vals}(Y)} \underbrace{P(Y=c|X_j=v)}_{\text{change!}} \log P(Y=c|X_j=v)$$

$$P(Y=c|X_j=v) = \frac{\sum_{i=1}^{n} w_i^{(t)} \mathbb{1}(y_i=c, x_{ij}=v)}{\sum_{i=1}^{n} w_i^{(t)} \mathbb{1}(x_{ij}=v)}$$

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

$$P(\text{positive}) = \frac{\sum_{\text{leaf}} w_i^{(t)} \mathbb{1}(y_i = 1)}{\sum_{\text{leaf}} w_i^{(t)}}$$



Outline for March 18

- Recap common issues on Midterm 1
- Finish AdaBoost
  - Handout 8
  - Overview (Handout 9)
  - Decision trees with weighted examples
- Begin Support Vector Machines (SVMs)
  - Linearly separable datasets
  - Perceptron

Hyperplane divides space into positive (+1) and negative (-1)

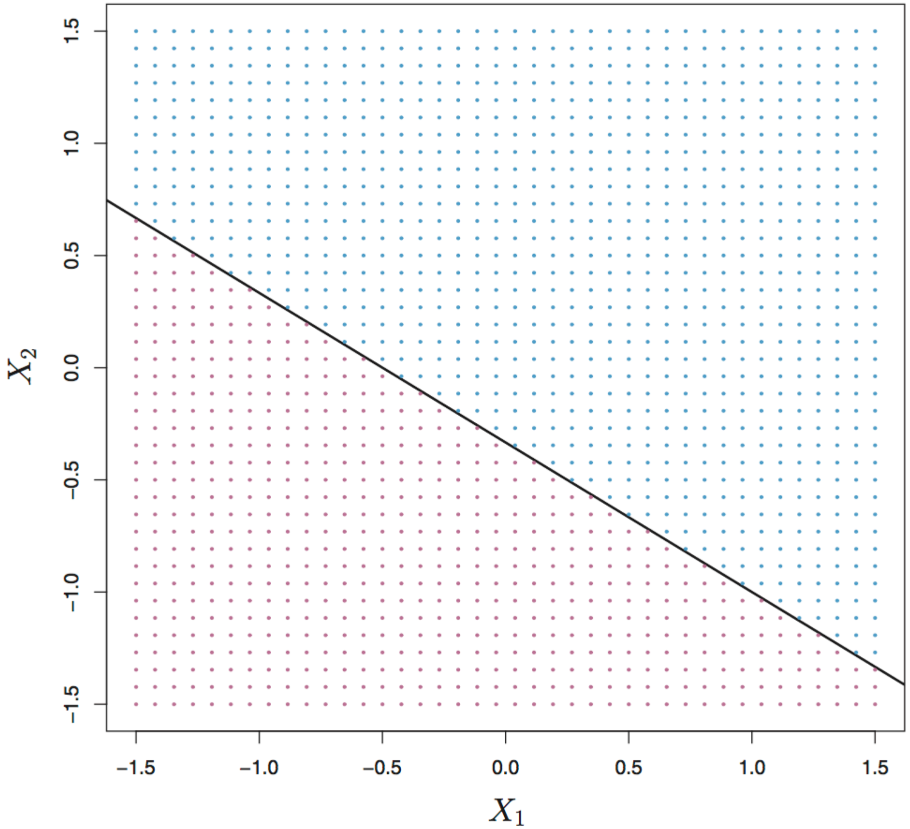


FIGURE 9.1. *The hyperplane* $1 + 2X_1 + 3X_2 = 0$ *is shown. The blue region is the set of points for which* $1 + 2X_1 + 3X_2 > 0$, *and the purple region is the set of points for which* $1 + 2X_1 + 3X_2 < 0$.

ISL: Figure 9.1

Goal: use training data to create a *separating* hyperplane

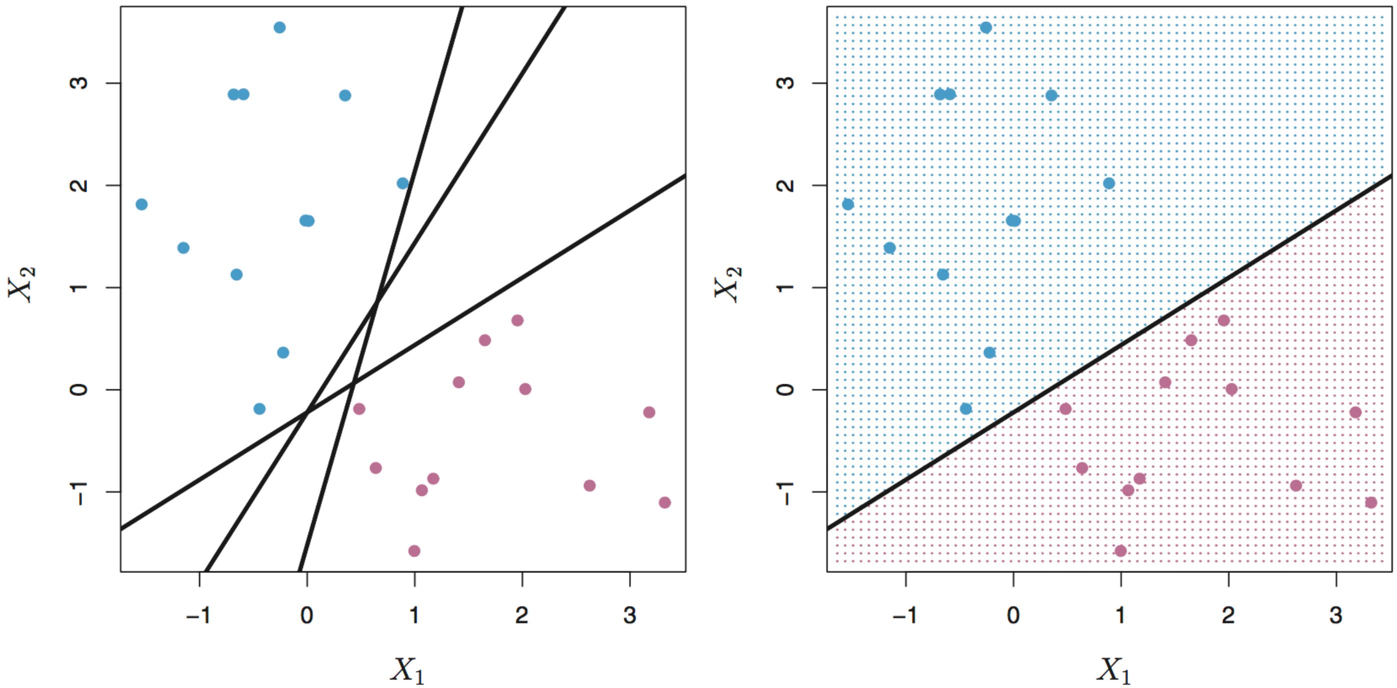


FIGURE 9.2. Left: *There are two classes of observations, shown in blue and in purple, each of which has measurements on two variables. Three separating hyperplanes, out of many possible, are shown in black.* Right: *A separating hyperplane is shown in black. The blue and purple grid indicates the decision rule made by a classifier based on this separating hyperplane: a test observation that falls in the blue portion of the grid will be assigned to the blue class, and a test observation that falls into the purple portion of the grid will be assigned to the purple class.*

ISL: Figure 9.2

These two hyperplanes would likely perform very differently on test data, but they both separate the training data

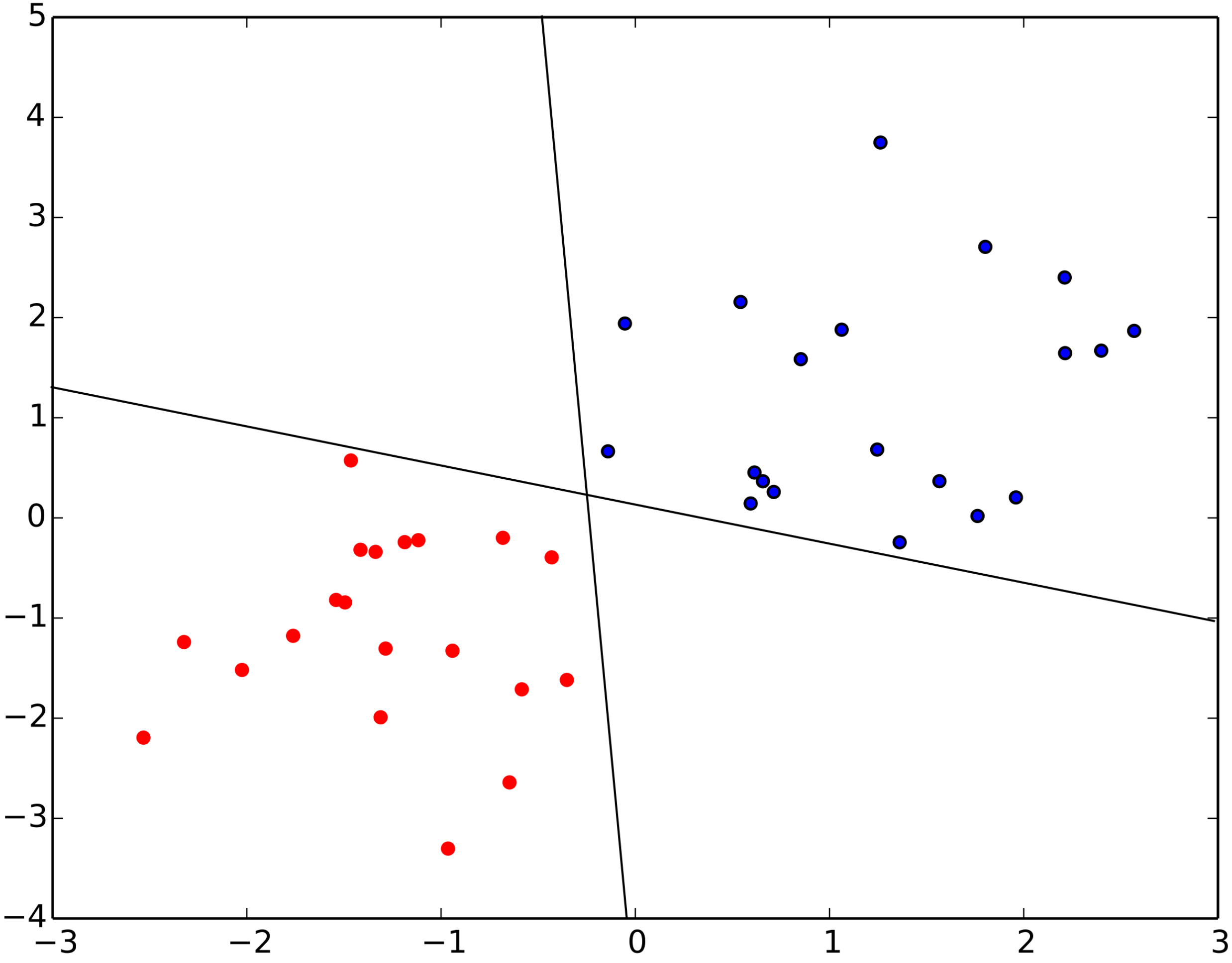


Image: Qwertyus, Wikipedia

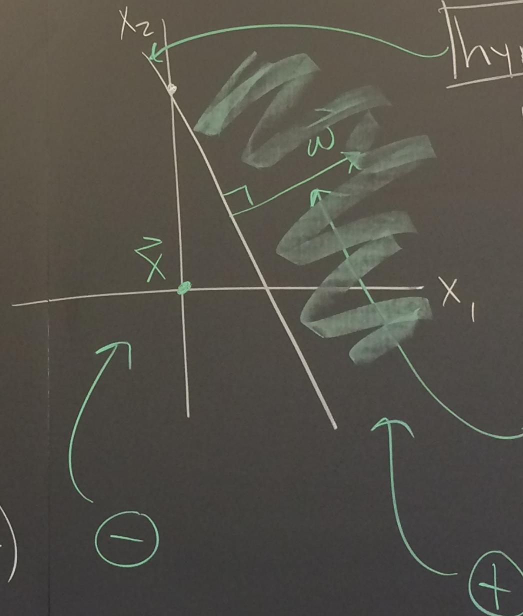
Perceptron

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

$h(\vec{x}) = \text{sign}(\vec{w}^T \vec{x})$, for some weight vector $\vec{w}$

$\boxed{w_j}$ is the weight on the jth feature

(similar to $\vec{b}$ from before)



$\boxed{\text{hyperplane}}$

$x_0 = 1$

$$\boxed{-5 + 2x_1 + x_2 = 0}$$

$$\underbrace{x_2 = -2x_1 + 5}_{\text{graphing}}$$

$$\vec{w} = \begin{bmatrix} -5 \\ 2 \\ 1 \end{bmatrix}$$

$$\boxed{\begin{array}{l} \text{if } \vec{w}^T \vec{x} > 0 \Rightarrow y = +1 \\ \text{if } \vec{w}^T \vec{x} \leq 0 \Rightarrow y = -1 \end{array}}$$