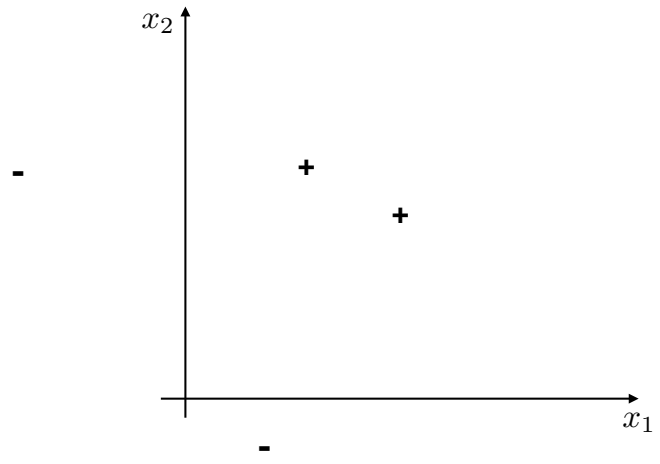
# Machine Learning Lecture 2

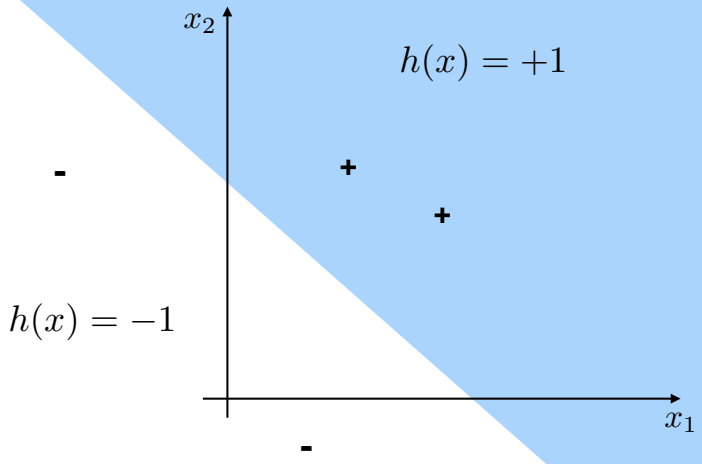
## Review of basic concepts

- Feature vectors, labels
- Training set
- Classifier
- Training error
- Test error
- Set of classifiers

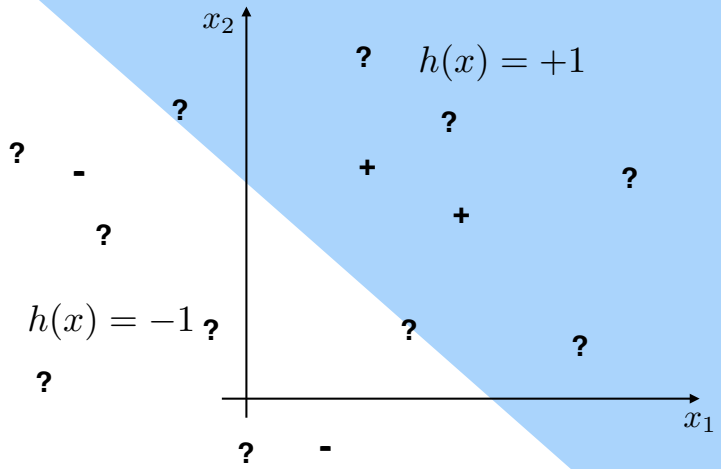
## Review: training set




## Review: a classifier



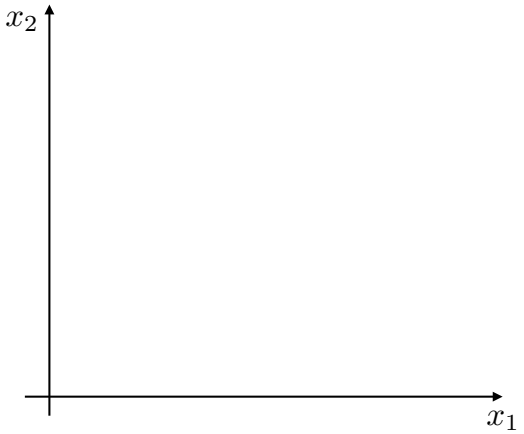
# Review: test set



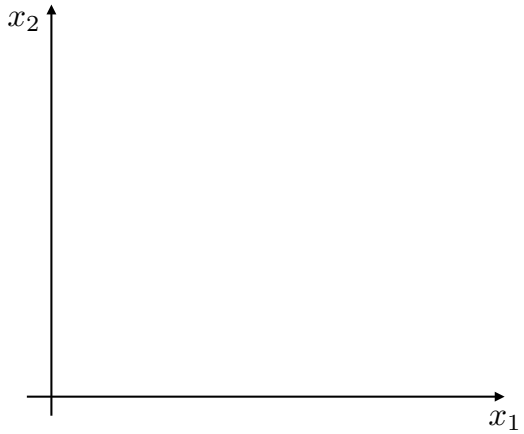

# This lecture

- The set of linear classifiers
- Linear separation
- Perceptron algorithm

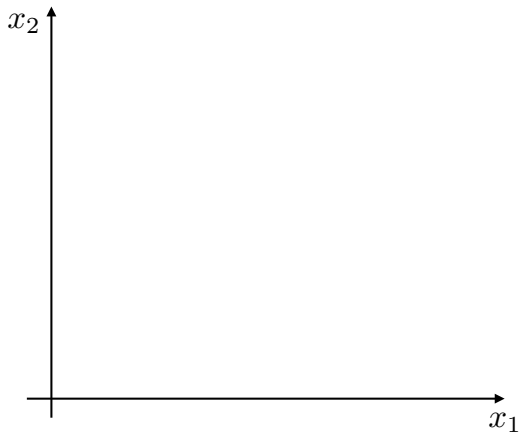
# Linear classifiers



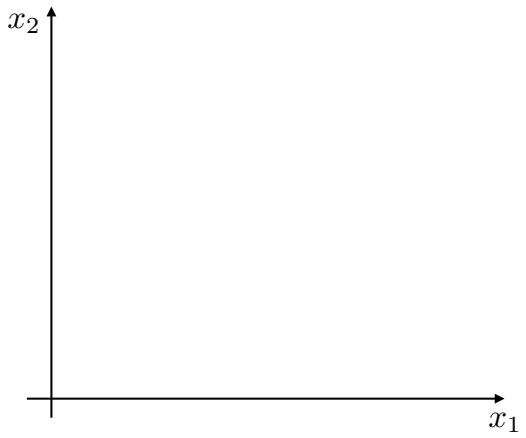

# Linear classifiers through origin



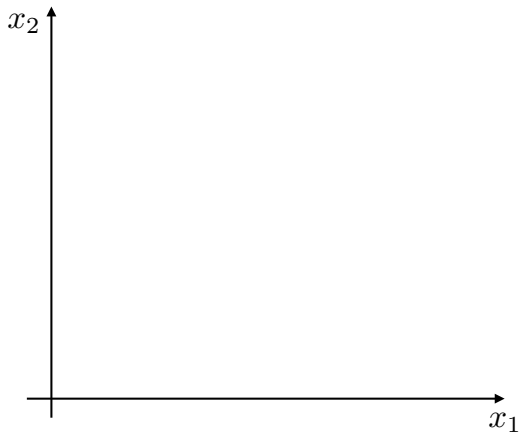
## Linear classifiers



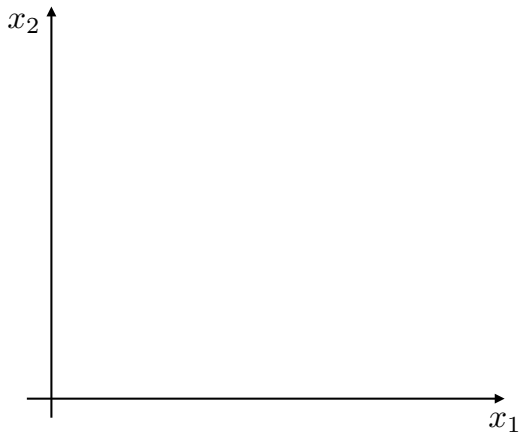

## Linear separation: ex




## Linear separation: ex




## Linear separation: ex



# Linear separation

**Definition:**
Training examples $S_n = \{(x^{(i)}, y^{(i)}\}), i = 1, \ldots, n\}$ are *linearly separable* if there exists a parameter vector $\hat{\theta}$ and offset parameter $\hat{\theta}_0$ such that $y^{(i)}(\hat{\theta} \cdot x^{(i)} + \hat{\theta}_0) > 0$ for all $i = 1, ..., n$.

# Learning linear classifiers

- Training error for a linear classifier (through origin)

# Learning linear classifiers

- Training error for a linear classifier

# Learning algorithm: perceptron

$\theta = 0$ (vector)

**if** $y^{(i)}(\theta \cdot x^{(i)}) \leq 0$ **then**
$\theta = \theta + y^{(i)} x^{(i)}$

## Learning algorithm: perceptron

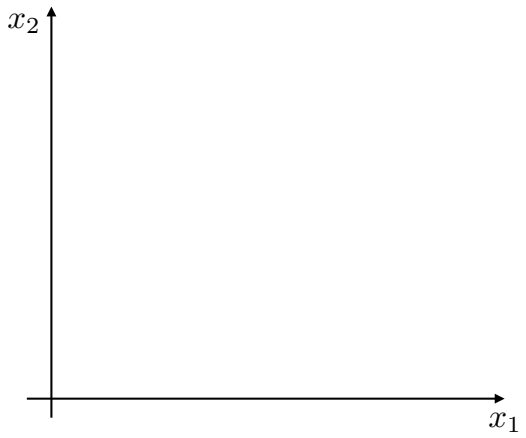
$\theta = 0$ (vector)

**for** $i = 1, \ldots, n$ **do**
  **if** $y^{(i)}(\theta \cdot x^{(i)}) \leq 0$ **then**
    $\theta = \theta + y^{(i)}x^{(i)}$

## Learning algorithm: perceptron

**procedure** PERCEPTRON($\{(x^{(i)}, y^{(i)}), i = 1, \ldots, n\}$, $T$)
  $\theta = 0$ (vector)
  **for** $t = 1, \ldots, T$ **do**
    **for** $i = 1, \ldots, n$ **do**
      **if** $y^{(i)}(\theta \cdot x^{(i)}) \leq 0$ **then**
        $\theta = \theta + y^{(i)}x^{(i)}$
  **return** $\theta$

## Perceptron algorithm: ex




## Perceptron (with offset)

1: **procedure** PERCEPTRON($\{(x^{(i)}, y^{(i)}), i = 1, \ldots, n\}$, $T$)
2:  $\theta = 0$ (vector), $\theta_0 = 0$ (scalar)
3:  **for** $t = 1, \ldots, T$ **do**
4:  **for** $i = 1, \ldots, n$ **do**
5:  **if** $y^{(i)}(\theta \cdot x^{(i)} + \theta_0) \leq 0$ **then**
6:  $\theta = \theta + y^{(i)}x^{(i)}$
7:  $\theta_0 = \theta_0 + y^{(i)}$
8:  **return** $\theta$, $\theta_0$

# Key things to understand

- Parametric families (sets) of classifiers
- The set of linear classifiers
- Linear separation
- Perceptron algorithm