TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

2.1 Linear Classifiers

Machine Learning 1: Foundations

Marius Kloft *(TUK)*

28 Apr – 5 May 2020

Recap

Machine learning

- computers learning from data
  how to make accurate predictions

Recap

Machine learning

- computers learning from data
  how to make accurate predictions

Formal problem setting and terminology

Recap

Machine learning

- computers learning from data
  how to make accurate predictions

Formal problem setting and terminology

- Given **training data** =
  - **inputs** $\mathbf{x}_1, \ldots, \mathbf{x}_n$
  - **labels** $y_1, \ldots, y_n$

Recap

Machine learning

- computers learning from data
  how to make accurate predictions

Formal problem setting and terminology

- Given **training data** =
  - **inputs** $\mathbf{x}_1, \ldots, \mathbf{x}_n$
  - **labels** $y_1, \ldots, y_n$
- Aim: to compute a function f (called **classifier** or **predictor**) predicting the unknown label y of a new input $\mathbf{x}$)

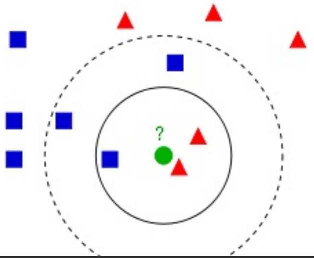
Recap

Machine learning

- computers learning from data
  how to make accurate predictions

Formal problem setting and terminology

- Given **training data** =
  - **inputs** $\mathbf{x}_1, \ldots, \mathbf{x}_n$
  - **labels** $y_1, \ldots, y_n$
- Aim: to compute a function f (called **classifier** or **predictor**) predicting the unknown label y of a new input $\mathbf{x}$)
- Example: k-nearest neighbor algorithm



Contents of this Video

Linear Classifiers

1. Math Notation & Recap
2. Linear Classifiers

1 Math Notation & Recap

2 Linear Classifiers

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.
- Matrices $M \in \mathbb{R}^{m \times n}$ have m rows and n columns and are denoted with normal letters.

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.
- Matrices $M \in \mathbb{R}^{m \times n}$ have m rows and n columns and are denoted with normal letters.
- Greek letters can refer to both scalars and vectors, but they are boldfaced if they denote vectors ($\boldsymbol{\lambda} \in \mathbb{R}^d$ vs. $\lambda \in \mathbb{R}$).

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.
- Matrices $M \in \mathbb{R}^{m \times n}$ have m rows and n columns and are denoted with normal letters.
- Greek letters can refer to both scalars and vectors, but they are boldfaced if they denote vectors ($\boldsymbol{\lambda} \in \mathbb{R}^d$ vs. $\lambda \in \mathbb{R}$).
- $\mathbf{0}$ and $\mathbf{1}$ are vectors in $\mathbb{R}^d$ with entries all zeros and ones, respectively.

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.
- Matrices $M \in \mathbb{R}^{m \times n}$ have m rows and n columns and are denoted with normal letters.
- Greek letters can refer to both scalars and vectors, but they are boldfaced if they denote vectors ($\boldsymbol{\lambda} \in \mathbb{R}^d$ vs. $\lambda \in \mathbb{R}$).
- $\mathbf{0}$ and $\mathbf{1}$ are vectors in $\mathbb{R}^d$ with entries all zeros and ones, respectively.
- Transposition of a vector or matrix:
  - if $\mathbf{v}$ is a column vector, then $v^\top$ is a row vector
  - if $M \in \mathbb{R}^{m \times n}$ then $M^\top \in \mathbb{R}^{n \times m}$.

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.
- Matrices $M \in \mathbb{R}^{m \times n}$ have m rows and n columns and are denoted with normal letters.
- Greek letters can refer to both scalars and vectors, but they are boldfaced if they denote vectors ($\boldsymbol{\lambda} \in \mathbb{R}^d$ vs. $\lambda \in \mathbb{R}$).
- $\mathbf{0}$ and $\mathbf{1}$ are vectors in $\mathbb{R}^d$ with entries all zeros and ones, respectively.
- Transposition of a vector or matrix:
  - if $\mathbf{v}$ is a column vector, then $v^\top$ is a row vector
  - if $M \in \mathbb{R}^{m \times n}$ then $M^\top \in \mathbb{R}^{n \times m}$.
- Scalar product of two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$: $\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v}^\top \mathbf{w}$

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.
- Matrices $M \in \mathbb{R}^{m \times n}$ have m rows and n columns and are denoted with normal letters.
- Greek letters can refer to both scalars and vectors, but they are boldfaced if they denote vectors ($\boldsymbol{\lambda} \in \mathbb{R}^d$ vs. $\lambda \in \mathbb{R}$).
- $\mathbf{0}$ and $\mathbf{1}$ are vectors in $\mathbb{R}^d$ with entries all zeros and ones, respectively.
- Transposition of a vector or matrix:
  - if $\mathbf{v}$ is a column vector, then $v^\top$ is a row vector
  - if $M \in \mathbb{R}^{m \times n}$ then $M^\top \in \mathbb{R}^{n \times m}$.
- Scalar product of two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$: $\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v}^\top \mathbf{w}$
- Norm of a vector: $\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v}^\top \mathbf{v}}$

Math Notation

- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.
- Matrices $M \in \mathbb{R}^{m \times n}$ have m rows and n columns and are denoted with normal letters.
- Greek letters can refer to both scalars and vectors, but they are boldfaced if they denote vectors ($\boldsymbol{\lambda} \in \mathbb{R}^d$ vs. $\lambda \in \mathbb{R}$).
- $\mathbf{0}$ and $\mathbf{1}$ are vectors in $\mathbb{R}^d$ with entries all zeros and ones, respectively.
- Transposition of a vector or matrix:
  - if $\mathbf{v}$ is a column vector, then $v^\top$ is a row vector
  - if $M \in \mathbb{R}^{m \times n}$ then $M^\top \in \mathbb{R}^{n \times m}$.
- Scalar product of two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$: $\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v}^\top \mathbf{w}$
- Norm of a vector: $\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v}^\top \mathbf{v}}$
- $\mathbf{v} \leq \mathbf{w}$ means $\forall i = 1, \ldots, d : v_i \leq w_i$

$$\begin{pmatrix} 1.7 \\ 2.3 \\ -0.3 \end{pmatrix} \leq \begin{pmatrix} 1.7 \\ 3.2 \\ 0.4 \end{pmatrix}$$

Math Notation

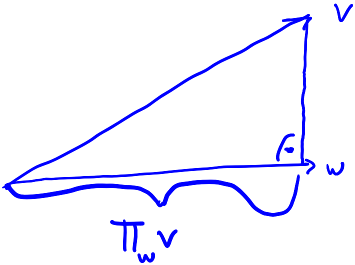
- Vectors $\mathbf{v} \in \mathbb{R}^d$ are thought of as column vectors and denoted with boldface letters.
- Scalars $s \in \mathbb{R}$ are denoted with normal letters.
- Matrices $M \in \mathbb{R}^{m \times n}$ have m rows and n columns and are denoted with normal letters.
- Greek letters can refer to both scalars and vectors, but they are boldfaced if they denote vectors ($\boldsymbol{\lambda} \in \mathbb{R}^d$ vs. $\lambda \in \mathbb{R}$).
- $\mathbf{0}$ and $\mathbf{1}$ are vectors in $\mathbb{R}^d$ with entries all zeros and ones, respectively.
- Transposition of a vector or matrix:
  - if $\mathbf{v}$ is a column vector, then $v^\top$ is a row vector
  - if $M \in \mathbb{R}^{m \times n}$ then $M^\top \in \mathbb{R}^{n \times m}$.
- Scalar product of two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^d$: $\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v}^\top \mathbf{w}$
- Norm of a vector: $\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v}^\top \mathbf{v}}$
- $\mathbf{v} \leq \mathbf{w}$ means $\forall i = 1, \ldots, d : v_i \leq w_i$
- The cardinality of a set S is denoted $|S|$

Math Recap: Projections

Recall from linear algebra:

Definition

The **scalar projection** of a vector $\mathbf{v} \in \mathbb{R}^d$ onto a vector $\mathbf{w} \in \mathbb{R}^d$ is $\Pi_{\mathbf{w}}\mathbf{v} = \mathbf{v}^\top \frac{\mathbf{w}}{\|\mathbf{w}\|}$.

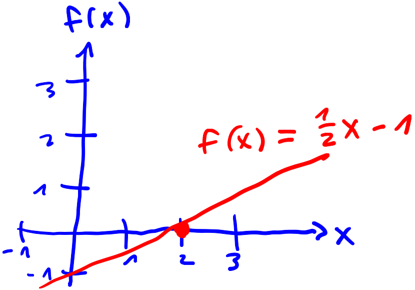


(with $\|w\|=1$)

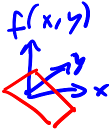
Math Recap: Hyperplanes

Definitions

- An (affine-)**linear function** is a function $f : \mathbb{R}^d \to \mathbb{R}$ of the form $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$, where $\mathbf{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$.
- A **hyperplane** is a subset $H \subset \mathbb{R}^d$ defined as $H := \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) = 0\}$.



$H = \{2\}$



Math Recap: Signed Distance

We define the sign here as positive if and only if $f(\mathbf{x}) \geq 0$.

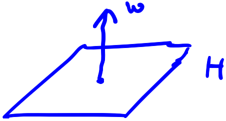
Math Recap: Signed Distance

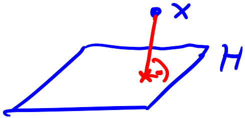
Proved on exercise sheet:

Proposition (properties of hyperplanes)

Let H be a hyperplane defined by the affine-linear function $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$. It holds:

1. The vector $\mathbf{w}$ is **orthogonal** to H, meaning that: for all $\mathbf{x}_1, \mathbf{x}_2 \in H$ it is $\mathbf{w}^\top(\mathbf{x}_1 - \mathbf{x}_2) = 0$.
2. The **signed distance** of a point $\mathbf{x}$ to H is given by
$$d(\mathbf{x}, H) \stackrel{\text{def.}}{:=} \pm \min_{\tilde{x} \in H} \|\mathbf{x} - \tilde{\mathbf{x}}\| \stackrel{\text{prop.}}{=} \frac{1}{\|\mathbf{w}\|} \left(\mathbf{w}^\top \mathbf{x} + b\right) .$$





We define the sign here as positive if and only if $f(\mathbf{x}) \geq 0$.

Linear Classifiers

Definition

A classifier of the form $f(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^\top \mathbf{x} + b)$ is called **linear classifier**.

Linear Classifiers

Definition

A classifier of the form $f(\mathbf{x}) = \operatorname{sign}(\mathbf{w}^\top \mathbf{x} + b)$ is called **linear classifier**.

What are advantages and disadvantages of linear classifiers?

Linear Classifiers

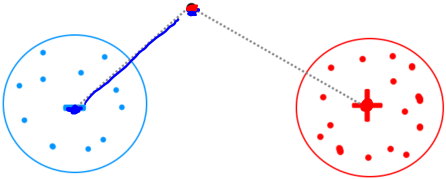
Definition

A classifier of the form $f(\mathbf{x}) = \text{sign}(\mathbf{w}^\top \mathbf{x} + b)$ is called **linear classifier**.

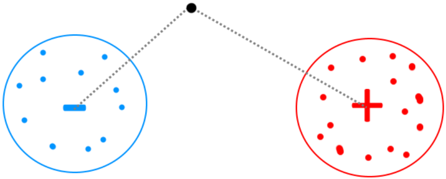
What are advantages and disadvantages of linear classifiers?

Please **pause** your video here and think about this question for a few minutes...

The **Nearest Centroid Classifier**



The Nearest Centroid Classifier

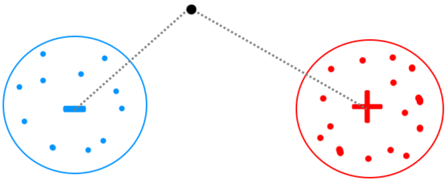


Let I_- and I_+ denote the indices of the data points labeled -1 and +1.

Training

- Compute $n_- = |I_-|$ and $n_+ = |I_+|$
- Compute $c_- = \frac{1}{n_-} \sum_{i \in I_-} x_i$ and $c_+ = \frac{1}{n_+} \sum_{i \in I_+} x_i$

The Nearest Centroid Classifier



Let I_- and I_+ denote the indices of the data points labeled -1 and +1.

Training

- Compute $n_- = |I_-|$ and $n_+ = |I_+|$
- Compute $c_- = \frac{1}{n_-} \sum_{i \in I_-} x_i$ and $c_+ = \frac{1}{n_+} \sum_{i \in I_+} x_i$

Prediction

- Given new x predict $\arg\min_{y \in \{-,+\}} \|x - c_y\|$

NCC is a Linear Classifier

Theorem

NCC is a linear classifier $f(\mathbf{x}) = \operatorname{sign}(\mathbf{w}^\top \mathbf{x} + b)$ with $\mathbf{w} := 2(c_+ - c_-)$ and $b = \|c_-\|^2 - \|c_+\|^2$.

Proof: Decision boundary: $H = \{x \in \mathbb{R}^d : \underbrace{\|x - c_-\| = \|x - c_+\|}_{(*)}\}$

$$(*) \iff \|x - c_-\|^2 = \|x - c_+\|^2$$

$$\iff \cancel{\|x\|^2} - 2c_-^\top x + \|c_-\|^2 = \cancel{\|x\|^2} - 2c_+^\top x + \|c_+\|^2$$

$$\iff \underbrace{2(c_+ - c_-)}_{=:w}{}^\top x + \underbrace{(\|c_-\|^2 - \|c_+\|^2)}_{=:b} = 0$$

$$\iff w^\top x + b = 0$$

Thus: $H = \{x \in \mathbb{R}^d : w^\top x + b = 0\}$ $\square$

Conclusion

Linear Classifier: $f(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^\top \mathbf{x} + b)$

- Fast and easy to understand

Example: NCC

- $f(\mathbf{x}) = \arg\min_{y \in \{-,+\}} \|x - c_y\|$