Machine Learning WS2012/13 (Unsupervised Learning)

Lecturer: Prof. Dr. Bethge

Examples of unsupervised learning applications

- *density estimation* (e.g. texture synthesis), image compression
- level set estimation
- clustering/mode finding e.g. spike sorting; algorithm Mixture of Gaussians
- metric learning e.g. 3D visualization; algorithm multidimensional scaling
- feature extraction -¿ representation learning, e.g. whitening; algorithm principle component analysis

1. Set algebras, measures, and probabilities

- **Overview for the mathematical construction of random variables**
  - Sample Space S (often also Ω). $s \in S$ "outcomes"
  - Boolean Algebra (or Sigma Algebra) $\mathcal{B}(S)$ and $A \in \mathcal{B}(S)$ are called "events". An event A is realized if an outcome s is in A.
    For example, $\mathcal{B}(S) = \mathcal{P}(S)$
  - Measure $m : \mathcal{B}(S) \to [0, \infty)$ and $m(A \dot\cup B) = m(A) + m(B)$
  - probability measure $p(S) = 1$, if m is a measure with $m(S) < \infty$ then $p(A) := m(A)/m(S)$ is a probability measure.
  - random variables: consider $M \in \mathcal{B}(S)$ with $M = \{s \in S : X(s) < \vartheta\} \subseteq S$
- **Power Set:** $\mathcal{P}(S) := \{A \subseteq S\}$, $|\mathcal{P}(S)| = 2^{|S|}$
- **Boolean Algebra:** $\mathcal{B}(S) \subseteq \mathcal{P}(S)$ is called a *Boolean Algebra* iff

  (i) $A \in \mathcal{B}(S) \Rightarrow \overline{A} \in \mathcal{B}(S)$
  (ii) $A, B \in \mathcal{B}(S) \Rightarrow A \cup B \in \mathcal{B}(S)$

  Examples:

  (i) $\{\emptyset, S\}$ and $\mathcal{P}(S)$ are Boolean Algebras.
  (ii) If $a \in S$ then $\{\emptyset, \{a\}, S - \{a\}, S\}$ is the Boolean Algebra *generated by* a.
  (iii) $I(R)$ is defined to be the smallest Boolean Algebra which contains all open intervals (a, b) for which $-\infty \leq a < b \leq \infty$.
- **Sigma-Algebra:** $\mathcal{B}(S) \subseteq \mathcal{P}(S)$ is called a Sigma-Algebra if

  (i) $A \in \mathcal{B}(S) \Rightarrow \overline{A} \in \mathcal{B}(S)$
  (ii) For all sequences $A_1, A_2, \ldots \in \mathcal{B}(S) \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{B}(S)$
- **Venn diagram**
- **Proposition 1 (Boolean laws):**

(B1) Idempotency: $A \cup A = A$; $\quad A \cap A = A$

(B2) Associativity: $A \cap (B \cap C) = (A \cap B) \cap C$

(B3) Commutativity: $A \cup B = B \cup A$; $\quad A \cap B = B \cap A$

(B4) Distributivity: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $\quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(B5) de Morgan's law: $\overline{A \cup B} = \overline{A} \cap \overline{B}$; $\quad \overline{A \cap B} = \overline{A} \cup \overline{B}$

(B6) Complements: $\overline{\overline{A}} = A$; $\quad A \cap \overline{A} = \emptyset$; $\quad A \cup \overline{A} = S$

(B7) Properties of S and $\emptyset$: $\quad A \cup S = S$; $\quad A \cup \emptyset = A$; $\quad A \cap S = A$; $\quad A \cap \emptyset = \emptyset$

- **(Finite) Measure:** A mapping $m : \mathcal{B}(S) \to [0, \infty), A \in \mathcal{B}(S) \to m(A) \in [0, \infty)$ is called a *measure* if for all $A, B \in \mathcal{B}(S)$ with $A \cap B = \emptyset$ holds: $m(A \cup B) = m(A) + m(B)$.
  (This property can be written compactly as $m(A \dot{\cup} B) = m(A) + m(B)$)

  Examples:

  (i) $m(A) := |A|$ is the *counting measure*.

  (ii) For $S = [a, b]$ and $\mathcal{B}(S) = \mathcal{I}([a, b])$ with $|a|, |b| < \infty$ the *Lebesgue Measure* is defined as $m([c, d]) := d - c$.

  (iii) If $f(x) \geq 0$ for all $a \leq x \leq b$ with $\int_a^b f(x)dx < \infty$ then $m([c, d]) := \int_c^d f(x)dx$ is a measure.

  (iv) For $S = R$ and $\mathcal{B}(S) = \mathcal{I}(R)$ the *Dirac Maß* is given by $\delta_a(J) = 1$ if $a \in J$ and $\delta_a(J) = 0$ otherwise.

- **Proposition 2:**

  (i) If $B \subseteq A$ then $m(A - B) = m(A) - m(B)$.

  (ii) If $B \subseteq A$ then $m(B) \leq m(A)$.

  (iii) $m(\emptyset) = 0$.

  (iv) $m(A \cup B) = m(A) + m(B) - m(A \cap B)$.

- **Proposition 3:**
  If $A, B \in \mathcal{B}(S)$ then:

  (i) $(A \cap B) \in \mathcal{B}(S)$ since $(A \cap B) = \overline{\overline{A} \cup \overline{B}}$.

  (ii) $\emptyset, S$ belong to every Boolean Algebra, since for $A \in \mathcal{B}(S)$ also $\overline{A} \in \mathcal{B}(S)$ and $A \cap \overline{A} = \emptyset$ and $A \cup \overline{A} = S$.

  (iii) $\mathcal{B}_B(S) := B \cap \mathcal{B}(S) = \{A \cap B : A \in \mathcal{B}(S)\}$ is a Boolean Algebra if $B \neq \emptyset$.

  (iv) $\mathcal{B}_B(S) \subseteq \mathcal{B}(S)$.

- **Probability measure:** For a sample space S and a Sigma Algebra $\mathcal{B}(S)$ a measure P is a *probability measure* if $P(S) = 1$.

- **Probability space:** A probability space consists of the triple $(S, \mathcal{B}(S), P)$ where S is the sample space, $\mathcal{B}(S)$ the event algebra (which is a Boolean or a sigma algebra) and P is a probability measure on $\mathcal{B}(S)$.

- **Proposition 4:**

  (i) $0 \leq P(A) \leq 1$.

  (ii) $P(\overline{A}) = 1 - P(A)$.

  (iii) If $\mathcal{E} = \{E_1, E_2, \ldots, E_n\}$ is a partition of S (that is $S = \bigcup_{k=1}^n E_k$ and $E_l \cap E_m = \emptyset$), then $\sum_{k=1}^n P(E_k) = 1$ and $P(A) = \sum_{k=1}^n P(A \cap E_k)$.

- **Specification of probabilities**
  (i) As a limiting value of relative frequencies: $P(A) = \lim_{n\to\infty} \frac{r_n(A)}{n}$
  (ii) Symmetrie assumptions (e.g. if we assume $P(s) = P(s')$ for all $s, s' \in S$ then it follows: $P(s) = 1/|S|$).
  (iii) Subjective probabilities (based on Cox axioms) can be determined via betting games.
- **Conditional probabilities:** If $(S, \mathcal{B}(S), P)$ is a probability space and $B \in \mathcal{B}(S)$ then the restriction of P to the reduced Sigma-Algebra $\mathcal{B}_B(S) := \{B \cap A : A \in \mathcal{B}(S)\}$ defines a measure $M_B : \mathcal{B}_B \to [0, \infty)$. If in addition $P(B) > 0$, then $P_B : \mathcal{B}_B(S) \to [0, 1], \quad A_B \mapsto P_B(A_B) := M_B(A_B)/P(B) = P(A_B)/P(B), \forall A_B \in \mathcal{B}_B(S)$ is a probability measure on $\mathcal{B}_B(S)$.
  The $A_B \in \mathcal{B}_B(S)$ usually originate from intersection of the elements of the original Sigma-Algebra $\mathcal{B}(S)$ with the event B.
- **Independence:**
  Two events A and B are *statistically independent* iff $P(A \cap B) = P(A)P(B)$.
- **Proposition 5:**
  (i) Two events A and B are (statistically) independent if $P(A) = 0$ or $P(B) = 0$.
  (ii) If two events A and B are independent and $P(A) > 0$, then: $P(B|A) = P_A(B) = P(B)$.
  (iii) If two events A and B are independent and $P(A) > 0$, then A and $\overline{B}$ are independent as well.

Random Variables

- **Cumulative distribution function (cdf):** Any univariate random variable can be uniquely defined by the cumulative distribution function:

$$F(x) = P(\{s \in S : X(s) \leq x\}), \quad \forall x \in S'$$

  It holds:
  (i) F is a nondecreasing function.
  (ii) $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$
  (iii) Any function that fulfills (i)+(ii) is a cdf.
  (iv) $P(\{s \in S : X(s) > x\}) = 1 - F(x)$
  (v) $P(\{s \in S : x_1 < X(s) \leq x_2\}) = F(x_2) - F(x_1)$
  (vi) discrete case: probability mass function / point probability $p(x) = F(x) - \lim_{\epsilon\to 0} F(x - \epsilon)$
  (vii) continuous case: probability density function (pdf) $\rho(x) = \frac{d}{dx}\mathcal{F}(x)$
- **Expectation (value):**

$$E[X] = \sum_{k=1}^{n} p(x_k)x_k, \quad E[f(X)] = \sum_{k=1}^{n} p(x_k)f(x_k)$$

- **Moments:**

$$E[X^m] = \sum_{k=1}^{n} p(x_k)x_k^m$$

- **Variance:**

$$Var[X] = E[X^2] - E[X]^2$$