Data Bases Data Mining
Functional dependencies exercise sheet

We recall the following inference rules, where reflexivity, augmentation and transitivity constitute the so called Armstrong's system.

$$\frac{Y \subseteq X}{X \rightarrow Y}\ \sigma_R \text{ (reflexivity)} \qquad \frac{X \rightarrow Y \quad X \rightarrow Z}{X \rightarrow YZ}\ \sigma_C \text{ (composition)}$$

$$\frac{X \rightarrow Y}{WX \rightarrow WY}\ \sigma_A \text{ (augmentation)} \qquad \frac{X \rightarrow YZ}{X \rightarrow Y}\ \sigma_D \text{ (decomposition)}$$

$$\frac{X \rightarrow Y \quad Y \rightarrow Z}{X \rightarrow Z}\ \sigma_T \text{ (transitivity)} \qquad \frac{X \rightarrow Y \quad WY \rightarrow Z}{WX \rightarrow Z}\ \sigma_P \text{ (pseudo-transitivity)}$$

A proof of $X \rightarrow Y$ from Σ written $\Sigma \vdash X \rightarrow Y$ is a *sequence* $\langle f_0, \dots, f_p \rangle$ of FDs s.t. $f_p = X \rightarrow Y$ and $\forall i \in [0..p]$ either $f_i \in \Sigma$, or f_i is the *conclusion* of a rule from the Armstrong's system of which all its *antecedents* $f_0 \dots f_p$ appear before f_i in the sequence.

Exercice 1 : Proof of Functional Dependencies (FD)

Let Σ be the following set of FDs:

$$BC \rightarrow A \qquad D \rightarrow BE$$
$$AC \rightarrow B \qquad B \rightarrow DE$$
$$AE \rightarrow C \qquad C \rightarrow E$$

1. Prove using Armstrong's system that the following dependencies are entailed by Σ
    1. $AD \rightarrow C$
    2. $AB \rightarrow C$
    3. $AE \rightarrow BD$
    4. $AC \rightarrow D$
    5. $CD \rightarrow A$

Exercice 2 : Inference rules for FDs

1. Is the following inference rule correct? If true, prove it, otherwise, exhibit a counter-example.

$$\frac{XW \rightarrow Y \quad XY \rightarrow Z}{X \rightarrow (Z \setminus W)}$$

2. Show that any proof $F \vdash X \rightarrow Y$ using the σ_P rule can be transformed into a proof that uses solely σ_A et σ_T.
3. Show that any proof $F \vdash X \rightarrow Y$ using the σ_R, σ_A and σ_T rules can be transformed into a proof that uses solely σ_R et σ_P.
4. Conclude that the set of rules $\{\sigma_R, \sigma_P\}$ is sound and complete for the inference problem of FDs.

Exercice 3 : The syntatic closure is a closure

The *syntactic closure* of X w.r.t. Σ is defined as $X^\star = \{A \mid \Sigma \vdash X \to A\}$. In the usual (algebraic) sense a *closure* is a mapping $\phi : \wp(E) \to \wp(E)$[1] that satisfies the three following properties:

Extensive $X \subseteq \phi(X)$

Increasing $X \subseteq Y \Rightarrow \phi(X) \subseteq \phi(Y)$

Idempotent $\phi(\phi(X)) = \phi(X)$

1. Show that $X^\star$ is a closure in the algebraic sense using the Armstrong's system. For idempotency, you may prove first that $\Sigma \vdash X \to X^\star$.
2. Show that for all set of attributes Y, if $X \subseteq Y \subseteq X^\star$ then $Y^\star = X^\star$
3. Use the previous property to compute the *set of closed sets* $Cl(\Sigma) = \{X^\star \mid X \subseteq R\}$ for Σ as defined in exercise 1.

Exercice 4 : Functional dependencies and propositional logic ([1])

The goal of this exercise is to relate functional dependencies and (classical) propositional logic, to highlight the point that the syntactic resemblance between a FD and logical implication is not a sheer one.

Let $\mathcal{P}$ be an infinite enumerable set of propositional variables and let Σ be a set of FD over $\mathbf{R}$. For each attribute $A \in \mathbf{R}$ we associate a corresponding *propositional variable* $\underline{A} \in \mathcal{P}$. This mapping is extended to FD: for each FD $A_1 \dots A_p \to B_1 \dots B_q \in \Sigma$ we associate the propositional formula $\underline{A_1} \wedge \dots \wedge \underline{A_p} \Rightarrow \underline{B_1} \wedge \dots \wedge \underline{B_q}$ with variables in $\mathcal{P}$. Finally, we write $\underline{\Sigma}$ for the set of propositional formulas associated to Σ. Let $\alpha = A_1 \dots A_p \to B_1 \dots B_q$ be a FD, we want to show that the following are equivalent:

$$\Sigma \models \alpha \quad (1)$$

$$\Sigma \models_2 \alpha \quad (2)$$

$$\underline{\Sigma} \models_P \underline{\alpha} \quad (3)$$

We write $\Sigma \models \alpha$ for the usual logical entailment between a set of FD Σ and a single FD α. We write $\Sigma \models_2 \alpha$ for the logical entailment restricted to the case where *relations have at most two tuples*. In other words, $\Sigma \models_2 \alpha$ is defined as $\forall r.(|r| = 2 \wedge r \models \Sigma) \Rightarrow r \models \alpha$. Finally, $\underline{\Sigma} \models_P \underline{\alpha}$ is the classical propositional logical entailment, i.e., for all assignment of propositional variables $\nu : \mathcal{P} \to \{0, 1\}$ such that $\nu \models \underline{\Sigma}$ it is the case that $\nu \models \underline{\alpha}$ as well.

Lemma 1. *Let $\nu : \mathcal{P} \to \{0, 1\}$ be an assignment of propositional variables, and let $r = \{t_1, t_2\}$ be the instance with two tuples t_1 and t_2 such that $t_1[A] = 1$ and $t_2[A] = \nu(\underline{A})$ for all $A \in \mathbf{R}$. Let $\alpha = A_1 \dots A_p \to B_1 \dots B_q$ be a FD. Then $\nu \models_P \underline{\alpha}$ if and only if $r \models \alpha$.*

1. Show that proposition (1) is equivalent to proposition (2). For the (2) $\Rightarrow$ (1) direction, use proof by contradiction (show that $\Sigma \models_2 \alpha$ and $\Sigma \not\models \alpha$ is inconsistent).
2. Show lemma 1 by constructing a two tuples instance “à la Armstrong” from ν and vice-versa.
3. Show that proposition (3) is equivlente to proposition (2) using lemma 1. Use proof by contradiction for each direction.
4. Conclude the main theorem.

References

[1] R. Fagin. Functional dependencies in a relational database and propositional logic. *IBM J. Res. Dev.*, 21(6):534–544, Nov. 1977.

[1] $\wp(E)$ is the set of all subsets of E, formally $\wp(E) = \{X \mid X \subseteq E\}$