CS717: Warm up problems

I am going to grade this paper. However, marks on this paper will not count toward the final marks. I would like you to apply yourself as much as possible in answering. This paper might help you and me evaluate each other to realize if this course is appropriate for you. Answer below and behind the page containing each question. Write your roll number on the top of the first page.

1. You should have come across the support vector machine formulation. Without getting into specific details, write in words, what is the advantage of the dual formulation of svm, using the kernel? The question is really what advantage you have in using the kernel.

Transaction ID	Items
1	{D,O,N,K,E,Y}
2	{M,O,N,K,E,Y}
3	{M,A,K,E}
4	{M,U,C,K,Y}
5	{C,O,O,K,I,E}

$$\Sigma = \{A, C, D, E, I, K, M, N, O, U, Y\}$$

Figure 1: Dataset $\mathcal{D}$

2. Let us say you are provided the data set $\mathcal{D}$ as in Figure 1. Propose an efficient algorithm to find all subsets of Σ that are subsets of more than m transactions (for some fixed m). (This is the idea behind the classic apriori algorithm).

3. The optimization problem in (1) is an *Integer Linear Program* (ILP).

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} \leq \mathbf{b} \\ & x_i \in \{0,1\} \end{array} \tag{1}$$

where $\mathbf{x} \in \Re^n$, $A \in \Re^{m\times n}$, $\mathbf{b} \in \Re^m$ and $\mathbf{c} \in \Re^n$. In a general method called relaxation, the constraint that x_i be zero or one is replaced with the linear inequalities $0 \leq x_i \leq 1$. The problem in (2) is called the *Relaxation of the Linear Program* (RLP).

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} \leq \mathbf{b} \\ & 0 \leq x_i \leq 1 \end{array} \tag{2}$$

It turns out that the RLP (2) is far easier to solve than the original ILP (1).

(a) What inequality relationship exists between the solution to the RLP (2) and the solution to the original ILP (1).
(1 Mark)

(b) What can you say about the original ILP (1) if the RLP (2) is infeasible?
(1 Mark)

(c) It sometimes happens that the RLP (2) has a solution with $x_i \in \{0,1\}$. What can you say in this case?
(1 Mark)

4. Tough problem. Not expected to solve by default. **Convergence of Cutting Plane algorithm:** We present a generalized cutting plane algorithm for the optimization problem

$$\begin{array}{ll} \text{maximize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{D} \end{array} \tag{3}$$

for some closed and convex set $\mathcal{D}$ and concave f. Let $\mathbf{g}(\mathbf{x})$ be a subgradient[1] at some point $\mathbf{x}$ for the function f. A version of the general cutting plane algorithm consists of solving the following problem in the k^{th} iteration to get $\mathbf{x}^k$.

$$\mathbf{x}^k = \begin{array}{ll} \text{maximize} & f^k(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{D} \end{array} \tag{4}$$

where the function f is replaced by a polyhedral approximation f^k constructed using the points $\mathbf{x}^i$ generated so far, along with their subgradients $\mathbf{g}(\mathbf{x}^i) \equiv \mathbf{g}^i$. More specifically,

$$f^k(\mathbf{x}) = \min\left\{f(\mathbf{x}^0) + (\mathbf{x} - \mathbf{x}^0)^T\mathbf{g}^0), \ldots, f(\mathbf{x}^{(k-1)}) + (\mathbf{x} - \mathbf{x}^{(k-1)})^T\mathbf{g}^{(k-1)}\right\}$$

Assume that the maximum of f^k is attained for all k. Prove that the cutting plane algorithm, with the updates presented as above, converges finitely for the dual of a linear program, with atleast one strategy for choosing the subgradient (in fact, it converges for any choice of the subgradients). Also state the choice of the subgradients.

You can assume that the dual function for a linear progran is of the form

$$\min_{i \in \mathcal{I}} \left\{\mathbf{a}_i^T\mathbf{x} + b_i\right\}$$

where $\mathcal{I}$ is a finite index set and $a_i \in \Re^n$ and b_i are given vectors respectively.

(6 Marks)

[1] $\mathbf{g}$ is a subgradient at $\mathbf{x}$ for a concave function f if and only if $-\mathbf{g}$ is a subgradient at $\mathbf{x}$ for the convex function $-f$.