15780: GRADUATE AI (SPRING 2017)

Midterm Exam

February 27, 2017

Name:

Andrew ID:

Heuristic Search	20	
Learning Theory	25	
Optimization and ML	30	
Linear Programming	25	
Total	100	

1 Heuristic Search (20 points)

Consider the following search problem. There is a set of *operations* $O = \{o_1, \ldots, o_n\}$, and a set of *conditions* $C = \{C_1, \ldots, C_m\}$. Each operation $o_i \in O$ has a set of *preconditions* $P_i \subseteq C$, and a set of *effects* $E_i \subseteq C$. A state is defined by a subset of conditions $S \subseteq C$. An operation $o_i \in O$ can be applied at state S if and only if $P_i \subseteq S$, and it leads to the state $S \cup E_i$. The goal state is C, i.e., the state that contains all conditions. The initial state is the empty set (so initially only operations o_i that have an empty P_i can be applied).

We define the following heuristic function h for this search problem. Given a state S, $h(S)$ computes the optimal path to the goal state, in the modified problem where every operation o_i is replaced with the operation o'_i, which has the same set of effects E_i, but an empty set of preconditions. Informally, any of the "old" operations can be applied at any state. (The perceptive student may have noticed that computing $h(S)$ is equivalent to solving the Minimum Set Cover problem, that is, computing $h(S)$ happens to be computationally hard, so this is a pretty bad heuristic.)

Prove that A* graph search with the heuristic h is optimal (it always finds the shortest sequence of operations that leads to the goal state). You may rely on any theorem stated in class.

2 Learning Theory (25 points)

Q1. (10 pt) For a finite function class F, show that VC-dim$(F) \leq \log_2(|F|)$.

Q2. (5 pt) Give an example of an input space X and a function class F such that VC-dim$(F) = \log_2(|F|)$.

Q3. (10 pt) Give an example of an input space X and two function classes F_1 and F_2 such that VC-dim$(F_i) = 0$ for $i = 1, 2$, but VC-dim$(F_1 \cup F_2) = 1$.

3 Optimization and ML (30 points)

Q1. (10 pt) Consider the regression problem of minimizing the sum of absolute losses using a linear hypothesis function, that is

$$\underset{\theta \in \mathbb{R}^n}{\text{minimize}} \;\; \sum_{i=1}^{m} \ell(h_\theta(x^{(i)}), y^{(i)}) \tag{1}$$

where $\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ is given by $\ell(\hat{y}, y) = |\hat{y} - y|$ and $h_\theta(x) = \theta^T x$. Show that this is a convex optimization problem in θ.

Q2. (10 pt) Prove that we can find the solution of the absolute loss linear regression problem by solving the following linear program

$$\begin{aligned} \underset{\theta \in \mathbb{R}^n, z \in \mathbb{R}^m}{\text{minimize}} \quad & \sum_{i=1}^{m} z_i \\ \text{subject to} \quad & -z_i \leq \theta^T x^{(i)} - y^{(i)} \leq z_i \end{aligned} \tag{2}$$

Q3. (10 pt) Consider the regression problem of minimizing the 0/1 loss using a linear hypothesis function, that is

$$\underset{\theta \in \mathbb{R}^n}{\text{minimize}} \sum_{i=1}^{m} \ell(h_\theta(x^{(i)}), y^{(i)}) \tag{3}$$

where $\ell : \mathbb{R} \times \mathbb{R} \to \{0, 1\}$ is given by $\ell(\hat{y}, y) = \mathbf{1}\{\hat{y} \cdot y \leq 0\}$ and $h_\theta(x) = \theta^T x$.

Prove that we can find the linear classifier that minimizes 0/1 loss using the following *binary integer* programming problem, for a large enough value of M.

$$\begin{aligned} \underset{\theta \in \mathbb{R}^n, z \in \{0,1\}^m}{\text{minimize}} \quad & \sum_{i=1}^{m} z_i \\ \text{subject to} \quad & y^{(i)} \theta^T x^{(i)} \geq 1 - z_i M \end{aligned} \tag{4}$$

4 Linear Programming (25 points)

4.1 Standard Form (10 points)

Recall that a linear program is in the *standard form* if it is expressed as follows:

$$\begin{aligned} \text{minimize } & c^T x \\ \text{subject to } & Ax = b \\ & x \geq 0 \end{aligned}$$

with optimization variable $x \in \mathbb{R}^n$, and problem data $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.
Convert the following problem to the standard form:

$$\begin{aligned} \text{maximize } & x_1 + 2x_2 \\ \text{subject to } & x_1 + 3x_2 \leq 12 \\ & -2x_1 - x_2 \geq -8 \\ & 1 \leq x_1 \\ & 0 \leq x_2 \leq 4. \end{aligned}$$

Specifically, what is c, A, and b in the converted problem?

4.2 Simplex Algorithm (15 points)

The following is **part** of the simplex algorithm for solving a linear program in the standard form:

Repeat:

1. Given index set $\mathcal{J}$ such that $x_{\mathcal{J}} = A_{\mathcal{J}}^{-1} b \geq 0$.
2. Find $j \notin \mathcal{J}$ for which $\bar{c}_j = c_j - c_{\mathcal{J}}^T A_{\mathcal{J}}^{-1} A_j < 0$.
3. Compute step direction $d_{\mathcal{J}} = -A_{\mathcal{J}}^{-1} A_j$ and determine index to remove

$$i^\star = ?$$

4. Update index set: $\mathcal{J} \leftarrow \mathcal{J} - \{i^\star\} \cup \{j\}$.

Choose **one** correct answer for each of the following statements:

Q1. (5 pt) In the second step of the algorithm, no $j \notin \mathcal{J}$ satisfies $c_j - c_{\mathcal{J}}^T A_{\mathcal{J}}^{-1} A_j < 0$.
This means [① a solution is found, ② the problem is infeasible, ③ the problem is unbounded].

Q2. (5 pt) In the third step of the algorithm, $i^\star$ should be set to

$$\left[① \operatorname*{arg\,min}_{i \in \mathcal{J}: d_i < 0} x_i/d_i,\ ② \operatorname*{arg\,max}_{i \in \mathcal{J}: d_i < 0} x_i/d_i,\ ③ \operatorname*{arg\,min}_{i \in \mathcal{J}: d_i \geq 0} x_i/d_i,\ ④ \operatorname*{arg\,max}_{i \in \mathcal{J}: d_i \geq 0} x_i/d_i \right].$$

Q3. (5 pt) In the third step of the algorithm, every $i \in \mathcal{J}$ satisfies $d_i \geq 0$.
This means [① a solution is found, ② the problem is infeasible, ③ the problem is unbounded].

Scrap Paper