Homework 8

CS 221 (Autumn 2012–2013)

Submission instructions: Write your answers in one PDF file named `hw8.pdf`. Remember to include your name and SUNet ID. Copy the PDF file onto `corn.stanford.edu`, ssh in to the machine, type `/usr/class/cs221/WWW/submit`, and follow the instructions.

1. Unsupervised learning: clustering (*5 points*)

Recall that K-means clusters points based on reconstruction loss defined by the squared distance:

$$\min_z \min_\mu \sum_{i=1}^n \|\mu_{z_i} - \phi(x_i)\|^2.$$

a. (*2 points*) Assume all our points are in one dimension, and we have Train $= \{1, 2, 8, 9\}$ with $\phi(x) = x$. Initialize K-means with two means $\mu_1 = 0$ and $\mu_2 = 3$ ($K = 2$). Run K-means until convergence, showing the assignments z and μ after each iteration. What is the final reconstruction loss?

b. (*3 points*) So far, we've focused on assigning points to clusters: $z_i \in \{1, \dots, K\}$ is the cluster to which point i is assigned, and $\mu_k \in \mathbb{R}^d$ is the center of the k-th cluster.

However, often data points exhibit hierarchical structure. For example, if Train $= \{1, 2, 8, 9, 31, 32, 38, 39\}$, and $K = 4$, then we would find clusters $\{1, 2\}$, $\{8, 9\}$, $\{31, 32\}$, and $\{38, 39\}$. In addition, we might want to further group the first two clusters and the last two clusters into *meta-clusters*.

Let $\nu_1, \dots, \nu_M$ be the centers of M meta-clusters and let $y_k \in \{1, \dots, M\}$ be the assignment of cluster k to a meta-cluster. In summary, we've introduced the following new variables: the meta-cluster assignments $y = (y_1, \dots, y_K)$ and the meta-cluster centers $\nu = (\nu_1, \dots, \nu_M)$.

We can now define the following optimization problem which will minimize the reconstruction error for both the points and the cluster centers:

$$\min_y \min_\nu \min_z \min_\mu \sum_{k=1}^K \|\nu_{y_k} - \mu_k\|^2 + \sum_{i=1}^n \|\mu_{z_i} - \phi(x_i)\|^2.$$

Given this objective function, it is natural to derive an alternating minimization algorithm by looping through the following four steps:

1. Fixing ν, z, μ, minimize the loss with respect to y.
2. Fixing y, z, μ, minimize the loss with respect to ν.
3. Fixing y, ν, μ, minimize the loss with respect to z.
4. Fixing y, ν, z, minimize the loss with respect to μ.

Give the updates for each of the four steps. Your updates should be similar to E- and M-steps for K-means. For example, the first update will be $y_k = \arg\min_m ???$ and the second update will be $\nu_m =$ average over ???, etc. Hint: when minimizing with respect to a particular variable, ignore any terms that don't interact with that variable, and then notice similarities with the K-means updates.

2. Q-learning (*4 points*)

In a simple world, there are 3 states A, B, C. The agent can move between adjacent states (actions are $\leftarrow$ and $\rightarrow$). Suppose the agent acts according to a purely random strategy π_{act}, yielding the following sequence of states, actions, and rewards:

$$A, \rightarrow, 3, B, \rightarrow, 2, C, \leftarrow, 0, B, \rightarrow, 4, C, \leftarrow, 1, B$$

a. (*1 point*)
Run the Q-learning algorithm starting with $\hat{Q}(s,a) = 0$ using a constant step size of $\eta_t = 0.5$, assuming the discount $\gamma = 1$. This run should make 5 updates. What are the final values of $\hat{Q}(s,a)$?

b. (*1 point*)
Given $\hat{Q}(s,a)$, what is the estimated optimal policy $\hat{\pi}(s)$ (for each state s, given the optimal action)?

c. (*2 points*)
Suppose that we believe that (s,a) pairs corresponding to moving towards the center B are similar (and thus should have the same Q value) and (s,a) pairs that move away are similar. Define two features $\phi_1(s,a) = [(s,a) = (A, \rightarrow) \text{ or } (s,a) = (C, \leftarrow)]$ and $\phi_2(s,a) = [(s,a) = (B, \rightarrow) \text{ or } (s,a) = (B, \leftarrow)]$ to capture this, and let $\hat{Q}(s,a) = w_1\phi_1(s,a) + w_2\phi_2(s,a)$. Rerun the Q-learning algorithm with this function approximation. What are the resulting weights w_1, w_2 and the induced optimal policy $\hat{\pi}$?

3. More logical connectives (*6 points*)

Propositional logic is built out of have propositional symbols A, B, C and logical connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$. Given a formula f which is built out of these pieces, we have already defined the interpretation of f as $\mathcal{I}_w(f)$. In this problem, we will extend propositional logic with new connectives. Think of these connectives as new language features which provide syntactic sugar (they don't increase the expressivity of propositional logic but make things notationally simpler).

a. (*2 points*)
We introduce a new mystery connective $\triangle$, whose interpretation is defined as follows:

$\mathcal{I}_w(f)$	$\mathcal{I}_w(g)$	$\mathcal{I}_w(f \triangle g)$
0	0	0
0	1	1
1	0	1
1	1	0

Construct a propositional formula h that depends only on f and g and has the same interpretation as $f \triangle g$ (that is, $\mathcal{I}_w(h) = \mathcal{I}_w(f \triangle g)$ for all models w).
Describe in one sentence what the connective is doing.

b. (*2 points*)
Consider a new ternary connective $\circ$, whose interpretation is as follows:

$\mathcal{I}_w(r)$	$\mathcal{I}_w(f)$	$\mathcal{I}_w(g)$	$\mathcal{I}_w(f \circ_r g)$
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

Construct a propositional formula h that depends only on f, g and r and has the same interpretation as $f \circ_r g$ (that is, $\mathcal{I}_w(h) = \mathcal{I}_w(f \circ_r g)$ for all models w). Describe in one sentence what the connective is doing.

c. (*2 points*)
Now let us consider first-order logic and extend it with a new quantifier $\exists_2 x\, f$ whose interpretation is that there exist at least two different objects x such that formula f (which depends on x) is true. For example, if we have a model $w = (A : \{\text{Alice}, \text{Bob}\}, B : \{\text{Bob}\})$, then $\mathcal{I}_w(\exists_2 x\, A(x)) = 1$ but $\mathcal{I}_w(\exists_2 x\, B(x)) = 0$.

Write an equivalent first-order logic formula h such that $\mathcal{I}_w(h) = \mathcal{I}_w(\exists_2 x\, \exists y\, R(x, y))$ for all w, where R is a general binary relation.

4. Validity, contingency, unsatisfiability (*3 points*)

For each of following formulae, say whether the formula is *valid* (true in all models), *contingent* (true in some model, false in some model), or *unsatisfiable* (false in all models).

If the formula is *contingent*, give an example of a concrete model w in which the formula is true and a model in which the formula is false. If the formula is *valid*, give a proof based on converting the negation of the formulae to CNF and applying resolution to derive a contradiction (false). If the formula is *unsatisfiable*, give a proof based on converting the formulae to CNF and applying resolution to derive a contradiction (false).

a. (*1 point*) $P \to (P \vee Q)$

b. (*1 point*) $((P \to Q) \wedge (R \to S) \wedge (P \vee R)) \to (Q \vee S)$.

c. (*1 point*) $(\forall x\, \exists y\, P(x, y)) \to (\exists x\, P(x, x))$

5. Family trees (*8 points*)

Suppose you're building a system that stores family relationships. The system will maintain a knowledge base of first-order logical formulae maintain a set of family relationships. For example, the formula $\forall x\, \forall y\, Parent(x, y) \leftrightarrow Child(y, x)$ says that if x is a parent of y, then y is a child of x.

In this domain, we have the following constant symbols: $Alice, Bob, Carol, John, Edward, Sarah$. We also have the following relation symbols: $Male$, $Female$, $Sister$, $Brother$, $Sibling$, $Father$, $Mother$, $Parent$, Son, $Daughter$, $Child$, $Ancestor$, $Descendent$, $Relative$.

In the following, you will construct first-order logic formulae from the above symbols to express various facts, using your prior knowledge about how family relationships work.

a. (*1 point*) Using unary relations $Male, Female$, write down a set of formulae encoding the fact that Alice, Carol, Sarah are female and that Bob, John, and Edward are male.

b. (*1 point*) Using the relation *Child*, define formulae encoding the fact that Alice and Bob had two children together, Carol and John, and that Carol has two children, Edward and Sarah.

c. (*1 point*) Define two formulae relating *Sibling* and *Parent* to *Child*. Hint: you should use the $\leftrightarrow$ connective and the $=$ connective (among others).

d. (*1 point*) Define six formulae relating $Brother, Sister, Son, Daughter, Father, Mother$ to $Female, Male, Sibling, Child, Parent$.

e. (*1 point*) Define one formula relating *Descendent* to *Child*, one formulae relating *Ancestor* to *Descendent*, and one formulae relating *Relative* to $Parent, Child$. Note that some some relations might appear on both sides of the $\leftrightarrow$ connective.

f. (*3 points*)

Give a proof based on forward chaining that Sarah is related to Alice ($Relative(Sarah, Alice)$ holds). To do this, first convert any relevant formulae into (possibly several) first-order Horn clauses.