Problem Set 3

December 3, 2009

Due date: Wed, Dec 9 2009 at 4pm; submit by email.

Exercise 1: (30 points) You are asked to evaluate the performance of two classification models, M_1 and M_2. The test set you have chosen contains 26 binary attributes, labeled as A through Z. Table 1 shows the posterior probabilities obtained by applying the two models on the test set. (Only the posterior probabilities for the positive class are shown.) As this is a two-class problem, $P(-) = 1 - P(+)$ and $P(- \mid A, \ldots, Z) = 1 - P(+ \mid A, \ldots, Z)$. Assume that we are mostly interested in detecting instances from the positive class.

1. Plot the ROC curve for both M_1 and M_2. You should plot both curves on the same graph. Explain which model you think it's better and why. (5 points)
2. For model M_1, suppose you choose the cutoff threshold to be $t = 0.5$. In other words, any test instances with posterior probability greater than t will be classified as a positive example. Compute the precision, recall and F-measure for such model at this threshold value. (5 points)
3. Repeat question 2 above for model M_2 using the same cutoff threshold. (5 points)
4. Compare the F-measure results for both models. Which model is better? Are the results consistent with what you expect from the ROC curve? (5 points)
5. Repeat question 2 above for model M_1 using threshold $t = 0.1$. Which threshold do you prefer, $t = 0.5$ or $t = 0.1$ and why? Are the results consistent with what you expect from the ROC curve? (5 points)

Instance	True Class	$P(+ \mid A, \ldots, Z, M_1)$	$P(+ \mid A, \ldots, Z, M_2)$
1	+	0.73	0.61
2	+	0.69	0.03
3	-	0.44	0.68
4	-	0.55	0.31
5	+	0.67	0.45
6	+	0.47	0.09
7	-	0.08	0.38
8	-	0.15	0.05
9	+	0.45	0.01
10	-	0.35	0.04

Table 1: Table of posterior probabilities obtained by applying the two classification models on the test set.

Exercise 2: (10 points) Consider an undirected, connected and non-bipartite graph $G(V, E)$; V refers to the set of nodes of G ($|V| = n$) and E refers to the set of edges of G ($|E| = m$). This undirected graph induces a Markov Chain M_G as follows: the states of M_G are the vertices of G, and for any two vertices $u, v \in V$, the entry of the corresponding transition matrix P is $P(u, v) = 1/d(u)$ if $(u, v) \in E$ and 0 otherwise. We use $d(u)$ to refer to the degree of node u in G. Let π be stationary-distribution vector of Markov Chain M_G. Show that for every $v \in V$, $\pi(v) = d(v)/2m$.

Exercise 3: (25 points) Assume the Web graph G and let it consist of n nodes, none of which is a sink node. Let P be the transition matrix of a simple random walk on this graph, and P' the transition matrix of the random walk that is enhanced with a uniform jump, from any node to any other node in G, with probability α. That is, $P' = \alpha P + (1 - \alpha)uv^T$, where u is a vector of all 1's and v is a uniform vector (all entries have value $1/n$).

1. Show that since P is a stochastic matrix, then P' is also a stochastic matrix. That is, show that the sum of the entries in every row of P' is equal to 1. (10 points).
2. In class we said that the stationary distribution q of the random walk described by P' can be computed by a vanilla power-method computation. More specifically, it can be done using the following iterative procedure:
   - $q^0 = v$ // v is the uniform vector.
   - $t = 1$
   - **repeat**
     - $q^t = (P')^T q^{t-1}$
     - $\delta = ||q^t - q^{t-1}||$
     - $t = t + 1$
   - **until** $\delta < \epsilon$

We also discussed that we can speedup the computation of $y = (P')^T x$ using the following procedure:

1. $y = \alpha P^T x$
2. $\beta = ||x||_1 - ||y||_1$
3. $y = y + \beta v$

Prove that this procedure indeed computes $y = (P')^T x$. Also discuss the computational gains of using this procedure instead of the vanilla power method algorithm. (15 points).

Exercise 4: (10 points) Let P be an $n \times n$ a stochastic matrix that corresponds to the transition matrix of a Markov chain. As for every stochastic matrix, we have that $\sum_{j=1}^{n} P(i, j) = 1$ for all $i \in \{1, \ldots, n\}$. Now consider the case where also $\sum_{i=1}^{n} P(i, j) = 1$ for all $j \in \{1, \ldots, n\}$. Such matrices, for which the sum of the values in every row and in every column is equal to 1 are called *doubly stochastic matrices*. Show that the stationary distribution of the Markov chain described by any doubly stochastic matrix is the uniform distribution.

Exercise 5: (25 points) Let D the domain (or the universe) of n distinct objects, and let P be the set of distinct pairs of objects in D. Also, let σ_1, σ_2 be two rankings (permutations) of the elements in D. The Kendall's tau distance between two permutations is defined as follows: For each distinct pair $\{i, j\} \in P$ if i and j are in the same order in σ_1 and σ_2, then $K_{ij}(\sigma_1, \sigma_2) = 0$; if i and j are in the opposite order (such as i being ahead of j in σ_1 and j being ahead of i in σ_2), then $K_{ij}(\sigma_1, \sigma_2) = 1$. The Kendall's tau distance between σ_1 and σ_2 is given by $K(\sigma_1, \sigma_2) = \sum_{\{i,j\} \in P}(\sigma_1, \sigma_2)$.

Very often, instead of observing the whole ranking of the n objects we see only the sorted lists of the first k elements of the ranking. We call such list a top-k list. Let τ_1 and τ_2 be the top-k lists of two rankings of the elements in D. Then, we define the p-Kendall tau distance between τ_1 and τ_2 as follows. For a pair of objects $i, j \in D$ we the following cases.

1. If i and j both appear in τ_1 and τ_2 and are in the same order (such as i being ahead of j in both top-k lists), then $K_{ij}^p(\tau_1, \tau_2) = 0$.

2. If i and j both appear in τ_1 and τ_2, but in opposite order (such as i being ahead of j in τ_1 and j ahead of i in τ_2) then, $K_{ij}^p(\tau_1, \tau_2) = 1$.

3. If i and j both appear in one top-k list (say τ_1) and exactly one of i or j, say i, appears in the other top-k list (say τ_2), then if i is ahead of j in τ_1, then $K_{ij}^p(\tau_1, \tau_2) = 0$. Otherwise, $K_{ij}^p(\tau_1, \tau_2) = 1$. Intuitively, we know that i is ahead of j as far as τ_2 is concerned, since i appears in τ_2, but j does not.

4. If i, but not j, appears in one of the top-k lists (say τ_1) and j but not i appears in the other top-k list (say τ_2), then $K_{ij}^p(\tau_1, \tau_2) = 1$. Intuitively, we know that i is ahead of j as far as τ_1 is concerned and j is ahead of i as far as τ_2 is concerned.

5. If i and j both appear in one top-k list (say τ_1), but neither i nor j appears in the other top-k list (say τ_2). We call such pairs special pairs and we define $K_{ij}^p(\tau_1, \tau_2) = p$ with $0 \leq p \leq 1$.

We define the p-Kendall tau distance between two top-k lists to be: $K^p(\tau_1, \tau_2) = \sum_{\{i,j\} \in P_{\tau_1 \cup \tau_2}} K_{ij}^p(\tau_1, \tau_2)$, where $P_{\tau_1 \cup \tau_2}$ is the set of distinct pairs $\{i, j\} \in D_{\tau_1} \cup D_{\tau_2}$, (note that D_{τ_1} (D_{τ_2}) is the subset of elements from D that appear in τ_1 (resp. τ_2). You are asked to prove the following:

1. Prove that the Kendall's tau distance between two permutations σ_1 and σ_2, denoted by $K(\sigma_1, \sigma_2)$ satisfies the triangle inequality. (10 points)

2. Find the values of p for which the p-Kendall tau distance, K^p, satisfies the triangle inequality. (15 points)