Supplementary Material for the Multiple Instance Learning Gaussian Process Probit Model

1 Details of Variational Inference Updates

As noted in the main text, we aim to find

$$\text{argmax}_{Q\in\mathcal{Q}} \,\text{ELBO}(Q), \text{ where} \tag{1}$$
$$\text{ELBO}(Q) := E_Q[\log P(\mathbf{u},\mathbf{f},\mathbf{m}|\{Y_b\};\mathbf{x}) - \log Q(\mathbf{u},\mathbf{f},\mathbf{m})] \text{ and} \tag{2}$$

$\mathcal{Q}$ denotes distributions of the form

$$Q(\mathbf{u},\mathbf{f},\mathbf{m}) = Q(\mathbf{u})Q(\mathbf{m})P(\mathbf{f}|\mathbf{u};\mathbf{x},\mathbf{z}), \tag{3}$$

where P the *augmented* MIL-GP-Probit model is as described in Section 3.2 of the main text. We will derive coordinate ascent updates for $Q(\mathbf{u})$ and $Q(\mathbf{m})$, which will also simultaneously give us the parametric form Q has if it satisfies Equation 1. To proceed, we first write down ELBO(Q), where $Q(\mathbf{u},\mathbf{f},\mathbf{m}) = Q(\mathbf{u})Q(\mathbf{m})P(\mathbf{f}|\mathbf{u})$ and notation has been cleaned to reduce clutter:

$$\text{ELBO}(Q) = E_{Q(\mathbf{u})Q(\mathbf{m})P(\mathbf{f}|\mathbf{u})}[\log P(\mathbf{u} + \log P(\mathbf{m}|\mathbf{f})) + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b) - \log Q(\mathbf{u}) - \log Q(\mathbf{m})].$$

Coordinate ascent update of $Q(\mathbf{u})$: To maximize ELBO(Q) with respect to $Q(\mathbf{u})$ such that $\int Q(\mathbf{u})d\mathbf{u} = 1$, we write down the Lagrangian, omitting additive terms not depending on $Q(\mathbf{u})$, and then its functional derivative with respect to $Q(\mathbf{u})$:

$$\mathcal{L}(Q) = E_{Q(\mathbf{u})}[E_{P(\mathbf{f}|\mathbf{u})Q(\mathbf{m})}[\log P(\mathbf{u}) + \log P(\mathbf{m}|\mathbf{f})] - \log Q(\mathbf{u})] - \lambda \textstyle\int Q(\mathbf{u})d\mathbf{u}$$
$$\frac{d\mathcal{L}}{Q(\mathbf{u})}(Q) = E_{P(\mathbf{f}|\mathbf{u})Q(\mathbf{m})}[\log P(\mathbf{u}) + \log P(\mathbf{m}|\mathbf{f})] - \log Q(\mathbf{u}) - 1 - \lambda$$

Setting the derivative equal to 0 gives

$$Q(\mathbf{u}) \propto \exp(E_{P(\mathbf{f}|\mathbf{u})Q(\mathbf{m})}[\log P(\mathbf{m}|\mathbf{f}) + \log P(\mathbf{u})]) \tag{4}$$

Recall that $P(\mathbf{f}|\mathbf{u}) = \mathcal{N}(\mathbf{f}; K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mathbf{u}, \text{diag}(K_{\mathbf{xx}} - K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}))$ so that

$$\begin{aligned}
Q(\mathbf{u}) &\propto \exp(E_{P(\mathbf{f}|\mathbf{u})Q(\mathbf{m})}[\text{tr}(\mathbf{f}\mathbf{f}^T) + \mathbf{f}^T\mathbf{m} - \tfrac{1}{2}\text{tr}(K_{\mathbf{zz}}^{-1}\mathbf{u}\mathbf{u}^T)]) &(5)\\
&= \exp(-\tfrac{1}{2}\text{tr}(K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mathbf{u}(K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mathbf{u})^T) + (K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mathbf{u})^T E_{Q(\mathbf{m})}[m] - \tfrac{1}{2}\text{tr}(K_{\mathbf{zz}}^{-1}\mathbf{u}\mathbf{u}^T))\\
&= \exp(-\tfrac{1}{2}\text{tr}(K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mathbf{u}\mathbf{u}^T K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}) + \mathbf{u}^T K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}E_{Q(\mathbf{m})}[\mathbf{m}] - \tfrac{1}{2}\text{tr}(K_{\mathbf{zz}}^{-1}\mathbf{u}\mathbf{u}^T))\\
&= \exp(-\tfrac{1}{2}\text{tr}(K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}(\mathbf{u}\mathbf{u}^T)) + (K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}E_{Q(\mathbf{m})}[\mathbf{m}])^T\mathbf{u} - \tfrac{1}{2}\text{tr}(K_{\mathbf{zz}}^{-1}(\mathbf{u}\mathbf{u}^T))) &(6)
\end{aligned}$$

We recognize this as the density of a $\mathcal{N}(\mathbf{u}; \mathbf{\Sigma}^u K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}E_{Q(\mathbf{m})}[\mathbf{m}], \Sigma^u)$ distribution, where $\Sigma^u = (K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1} + K_{\mathbf{zz}}^{-1})^{-1}$. So, the coordinate ascent update to $Q(\mathbf{u})$ gives a distribution

$$Q(\mathbf{u}) = \mathcal{N}(\mathbf{u}; \mu^u, \Sigma^u), \text{ where} \tag{7}$$
$$\Sigma^u = (K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1} + K_{\mathbf{zz}}^{-1})^{-1} \tag{8}$$
$$\mu^u = \Sigma^u K_{\mathbf{zz}}^{-1}K_{\mathbf{zx}}E_{Q(\mathbf{m})}[\mathbf{m}] \tag{9}$$

Coordinate ascent update of $Q(\mathbf{m})$: Similarly, to maximize ELBO(Q) with respect to $Q(\mathbf{m})$ such that $\int Q(\mathbf{m})d\mathbf{m} = 1$, we write down the Lagrangian, omitting additive terms not depending on $Q(\mathbf{m})$, and then its functional derivative with respect to $Q(\mathbf{m})$:

$$\mathcal{L}(Q) = E_{Q(\mathbf{m})}[E_{Q(\mathbf{u})P(\mathbf{f}|\mathbf{u})}[\log P(\mathbf{m}|\mathbf{f})] + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b) - \log Q(\mathbf{m})] - \lambda \int Q(\mathbf{m})d\mathbf{m}$$
$$\frac{d\mathcal{L}}{Q(\mathbf{m})}(Q) = E_{Q(\mathbf{u})P(\mathbf{f}|\mathbf{u})}[\log P(\mathbf{m}|\mathbf{f})] + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b) - \log Q(\mathbf{m}) - 1 - \lambda$$

Setting the derivative to 0 gives

$$\begin{aligned}
Q(\mathbf{m}) &\propto \exp(E_{Q(\mathbf{u})P(\mathbf{f}|\mathbf{u})}[\log P(\mathbf{m}|\mathbf{f})] + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b)) \\
&= \exp(E_{Q(\mathbf{u})P(\mathbf{f}|\mathbf{u})}[\log \mathcal{N}(\mathbf{m}; \mathbf{f}, \mathbf{I}_N)] + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b)) \\
&\propto \exp(E_{Q(\mathbf{u})P(\mathbf{f}|\mathbf{u})}[\log(-\tfrac{1}{2}\operatorname{tr}(\mathbf{I}_N^{-1}\mathbf{m}\mathbf{m}^T) + (\mathbf{I}_N^{-1}\mathbf{f})^T\mathbf{m})] + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b)) \\
&= \exp(\log(-\tfrac{1}{2}\operatorname{tr}(\mathbf{I}_N^{-1}\mathbf{m}\mathbf{m}^T) + (\mathbf{I}_N^{-1}E_{Q(\mathbf{u})P(\mathbf{f}|\mathbf{u})}[\mathbf{f}])^T\mathbf{m}) + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b)) \\
&= \exp(\log(-\tfrac{1}{2}\operatorname{tr}(\mathbf{I}_N^{-1}\mathbf{m}\mathbf{m}^T) + (\mathbf{I}_N^{-1}E_{Q(\mathbf{u})}[K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mathbf{u}])^T\mathbf{m}) + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b)) \\
&= \exp(\log(-\tfrac{1}{2}\operatorname{tr}(\mathbf{I}_N^{-1}\mathbf{m}\mathbf{m}^T) + (\mathbf{I}_N^{-1}(K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mu^u))^T\mathbf{m}) + \sum_{b\in\mathcal{B}} \log P(Y_b|M_b)) \\
&= \mathcal{N}(\mathbf{m}; K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mu^u, \mathbf{I}_N)\Pi_{b\in\mathcal{B}}P(Y_b|M_b)
\end{aligned}$$

Thus, the coordinate ascent update to $Q(\mathbf{m})$ gives a distribution $Q(\mathbf{m}) = \mathcal{N}(\mathbf{m}; \mu^m, \mathbf{I}_N)\Pi_{b\in\mathcal{B}}P(Y_b|M_b)$, where $\mu^m = K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mu^u$. An alternate way to write the update is

$$Q(\mathbf{m}) = \Pi_{b\in\mathcal{B}}Q(M_b), \text{ where for } b \in \mathcal{B}, \tag{10}$$
$$Q(M_b) \propto \Pi_{i\in b}\mathcal{N}(m_i; \mu_i^m, 1)P(Y_b|M_b), \text{ where} \tag{11}$$
$$\mu^m = K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mu^u. \tag{12}$$

2 Details of Gibbs Sampling Updates

To derive the full conditional distribution of $\mathbf{u}$ under $P(\mathbf{u}, \mathbf{m}|\{Y_b\}; \mathbf{x}) = P(\mathbf{u}|\mathbf{m}; \mathbf{x})$ (by conditional independence), we can leverage the close relationship between mean-field variational inference and Gibbs sampling. Equations 11-12 gave the form of the update for $Q(\mathbf{u})$ given $Q(\mathbf{m})$. Based on standard mean-field theory, this update also equals

$$Q(\mathbf{u}) \propto \exp E_{Q(\mathbf{m})}[\log P(\mathbf{u}, \mathbf{m})].$$

If we set $Q(\mathbf{m})$ to a point mass at a given $\mathbf{m}$, we obtain

$$Q(\mathbf{u}) = P(\mathbf{u}|\mathbf{m})$$

The full conditional distribution $P(\mathbf{u}|\mathbf{m})$ is derived simply by setting $Q(\mathbf{m})$ to point mass at a given $\mathbf{m}$, and appealing to Equations 7-9.

To derive the full conditional distribution of m_i, i.e. $P(m_i|\{m_{i'}\}_{i'\in b, i'\neq i}, \mathbf{u}, \{Y_b\}; \mathbf{x})$, we again leverage the close relationship between mean-field variational inference and Gibbs sampling. Equation 10 gave the form of the update for $Q(\mathbf{m})$ given $Q(\mathbf{u})$. Based on standard mean-field theory, this update also equals

$$Q(\mathbf{m}) \propto \exp E_{Q(\mathbf{u})}[\log P(\mathbf{u}, \mathbf{m}, \{Y_b\})].$$

If we set $Q(\mathbf{u})$ to a point mass at a given $\mathbf{u}$, we obtain

$$Q(\mathbf{m}) = P(\mathbf{m}|\{Y_b\}, \mathbf{u})$$

Appealing to Equations 11-12, $P(\mathbf{m}|\{Y_b\}, \mathbf{u}) = \Pi_{i\in\mathcal{B}} P(M_b|Y_b, \mathbf{u})$, where

$$\begin{aligned} P(M_b|Y_b, \mathbf{u}) &\propto \Pi_{i\in b}\mathcal{N}(m_i; \mu_i^m, 1)P(Y_b|M_b), \text{ where} \\ \mu^m &:= (\mu_1^m, \ldots, \mu_N^m) = K_{\mathbf{xz}}K_{\mathbf{zz}}^{-1}\mathbf{u}. \end{aligned}$$

If $Y_b = 0$, then considering Equation 11 of the main text,

$$\begin{aligned} P(M_b|Y_b, \mathbf{u}) &\propto \Pi_{i\in b}\mathcal{N}(m_i; \mu_i^m, 1)1[0 > m_i], \text{ so that} \\ P(m_i|\{m_{i'}\}_{i\in b, i'\neq i}, Y_b, \mathbf{u}) &\propto \mathcal{N}(m_i; \mu_i^m, 1)1[0 > m_i]. \end{aligned}$$

If $Y_b = 1$, then considering Equation 12 of the main text

$$P(M_b|Y_b, \mathbf{u}) \propto (1 - \Pi_{i\in b}1[0 > m_i])\Pi_{i\in b}\mathcal{N}(m_i; \mu_i^m, 1), \text{ so that} \tag{13}$$

$$P(m_i|\{m_{i'}\}_{i\in b, i'\neq i}, Y_b, \mathbf{u}) \propto (1 - 1[0 > m_i]\Pi_{i'\in b, i'\neq i}1[0 > m_{i'}])\Pi_{i'\in b}\mathcal{N}(m_{i'}; \mu_{i'}^m, 1) \tag{14}$$

There are two cases for Equation 14. If $m_{i'} < 0 \;\forall\; i' \in b, i' \neq i$, then

$$\begin{aligned} P(m_i|\{m_{i'}\}_{i\in b, i'\neq i}, Y_b, \mathbf{u}) &\propto (1 - 1[0 > m_i])\Pi_{i'\in b}\mathcal{N}(m_{i'}; \mu_{i'}^m, 1) \\ &= 1[0 \leq m_i]\mathcal{N}(m_i; \mu_{i'}^m, 1) \end{aligned}$$

Else,

$$\begin{aligned} P(m_i|\{m_{i'}\}_{i\in b, i'\neq i}, Y_b, \mathbf{u}) &\propto \Pi_{i'\in b}\mathcal{N}(m_{i'}; \mu_{i'}^m, 1) \\ &= \mathcal{N}(m_i; \mu_i^m, 1) \end{aligned}$$

3 Experimental results for all methods on individual datasets

In Tables 1 and 2, we show the performance of all methods for all datasets in the 20 Newsgroups dataset collection and the addition 59 non-synthetic datasets, respectively.

Criteria	AUC				Loglik				MAP			
Method dataset	MIL-GP-PROBIT-GIBBS	MIL-GP-PROBIT	MIL-GP-LOGISTIC	MIL-GP-LOGISTIC-LM	MIL-GP-PROBIT-GIBBS	MIL-GP-PROBIT	MIL-GP-LOGISTIC	MIL-GP-LOGISTIC-LM	MIL-GP-PROBIT-GIBBS	MIL-GP-PROBIT	MIL-GP-LOGISTIC	MIL-GP-LOGISTIC-LM
alt.atheism	0.877	0.969	0.974	0.967	-0.036	-0.036	-0.158	-0.343	0.642	0.714	0.700	0.684
comp.graphics	0.878	0.901	0.928	0.926	-0.051	-0.052	-0.164	-0.343	0.765	0.796	0.787	0.783
comp.os.ms-windows.misc	0.795	0.903	0.922	0.920	-0.035	-0.036	-0.159	-0.343	0.498	0.543	0.541	0.536
comp.sys.ibm.pc.hardware	0.867	0.909	0.955	0.942	-0.037	-0.038	-0.156	-0.343	0.661	0.708	0.700	0.683
comp.sys.mac.hardware	0.877	0.943	0.947	0.946	-0.041	-0.042	-0.159	-0.343	0.707	0.761	0.763	0.766
comp.windows.x	0.861	0.946	0.972	0.972	-0.055	-0.056	-0.168	-0.343	0.667	0.734	0.736	0.728
misc.forsale	0.753	0.908	0.945	0.935	-0.033	-0.034	-0.156	-0.343	0.465	0.521	0.526	0.524
rec.autos	0.865	0.944	0.935	0.933	-0.050	-0.051	-0.170	-0.344	0.691	0.746	0.741	0.734
rec.motorcycles	0.861	0.979	0.981	0.971	-0.039	-0.040	-0.169	-0.344	0.617	0.685	0.720	0.722
rec.sport.baseball	0.866	0.945	0.976	0.970	-0.050	-0.051	-0.174	-0.344	0.706	0.759	0.776	0.773
rec.sport.hockey	0.905	0.988	0.990	0.990	-0.074	-0.075	-0.181	-0.344	0.827	0.914	0.923	0.923
sci.crypt	0.850	0.988	0.995	0.984	-0.041	-0.042	-0.161	-0.343	0.618	0.703	0.773	0.777
sci.electronics	0.967	0.990	0.967	0.954	-0.047	-0.048	-0.154	-0.343	0.907	0.926	0.918	0.907
sci.med	0.874	0.956	0.951	0.943	-0.053	-0.054	-0.171	-0.344	0.686	0.760	0.742	0.733
sci.space	0.836	0.962	0.981	0.980	-0.049	-0.049	-0.175	-0.344	0.650	0.731	0.752	0.746
soc.religion.christian	0.833	0.960	0.971	0.969	-0.039	-0.040	-0.178	-0.344	0.655	0.747	0.750	0.739
talk.politics.guns	0.839	0.979	0.975	0.971	-0.047	-0.048	-0.163	-0.343	0.623	0.702	0.723	0.719
talk.politics.mideast	0.862	0.974	0.974	0.975	-0.049	-0.050	-0.160	-0.343	0.723	0.805	0.850	0.857
talk.politics.misc	0.826	0.966	0.969	0.968	-0.036	-0.037	-0.153	-0.343	0.574	0.637	0.646	0.641
talk.religion.misc	0.835	0.932	0.937	0.933	-0.037	-0.038	-0.171	-0.344	0.489	0.561	0.531	0.521

Table 1: Performance of all considered methods on datasets of the 20 newsgroups dataset collection. Our method (MIL-GP-PROBIT) has higher predictive log-likelihood than MIL-GP-LOGISTIC [1] on all 20 datasets, and is comparable in terms of AUC and MAP. Despite its variational approximation, MIL-GP-PROBIT has comparable predictive log-likelihood to our exact inference method MIL-GP-PROBIT-GIBBS, whose AUC and MAP suffer due to the difficult of distinguishing between instances with similar true probabilities via sampling.

Criteria	AUC				Loglik				MAP			
Method	MIL-GP-Probit-Gibbs	MIL-GP-Probit	MIL-GP-Logistic	MIL-GP-Logistic-LM	MIL-GP-Probit-Gibbs	MIL-GP-Probit	MIL-GP-Logistic	MIL-GP-Logistic-LM	MIL-GP-Probit-Gibbs	MIL-GP-Probit	MIL-GP-Logistic	MIL-GP-Logistic-LM
dataset												
50Salad_0	0.854	0.852	0.831	0.522	-0.281	-0.252	-0.258	-0.346	0.530	0.530	0.477	0.178
50Salad_1	0.834	0.842	0.838	0.629	-0.287	-0.258	-0.255	-0.346	0.463	0.483	0.487	0.231
50Salad_2	0.848	0.859	0.800	0.674	-0.179	-0.166	-0.191	-0.345	0.474	0.483	0.363	0.176
50Salad_3	0.856	0.874	0.878	0.728	-0.472	-0.412	-0.287	-0.346	0.657	0.695	0.771	0.420
50Salad_4	0.993	0.993	0.991	0.963	-0.103	-0.106	-0.143	-0.338	0.968	0.969	0.974	0.696
50Salad_5	0.817	0.825	0.835	0.713	-0.216	-0.198	-0.185	-0.317	0.351	0.372	0.445	0.200
Voc12_0	0.664	0.643	0.542	0.419	-0.024	-0.025	-0.025	-0.024	0.013	0.012	0.009	0.007
Voc12_1	0.676	0.662	0.534	0.326	-0.073	-0.073	-0.075	-0.081	0.061	0.056	0.036	0.024
Voc12_10	0.548	0.551	0.446	0.462	-0.082	-0.083	-0.091	-0.087	0.043	0.044	0.034	0.036
Voc12_11	0.763	0.751	0.677	0.445	-0.058	-0.061	-0.063	-0.067	0.088	0.078	0.043	0.023
Voc12_12	0.831	0.817	0.746	0.418	-0.050	-0.051	-0.053	-0.055	0.114	0.092	0.047	0.017
Voc12_13	0.815	0.812	0.739	0.280	-0.068	-0.068	-0.072	-0.078	0.153	0.139	0.086	0.026
Voc12_14	0.721	0.724	0.714	0.650	-0.304	-0.304	-0.346	-0.347	0.527	0.534	0.518	0.464
Voc12_15	0.741	0.743	0.694	0.507	-0.088	-0.086	-0.090	-0.101	0.128	0.123	0.090	0.047
Voc12_16	0.696	0.708	0.576	0.473	-0.034	-0.034	-0.034	-0.034	0.021	0.023	0.015	0.011
Voc12_17	0.622	0.627	0.627	0.610	-0.082	-0.083	-0.085	-0.091	0.061	0.062	0.056	0.052
Voc12_18	0.568	0.552	0.385	0.319	-0.026	-0.028	-0.027	-0.026	0.011	0.010	0.007	0.007
Voc12_19	0.863	0.861	0.816	0.681	-0.061	-0.063	-0.068	-0.073	0.173	0.159	0.100	0.057
Voc12_2	0.715	0.727	0.696	0.703	-0.022	-0.024	-0.023	-0.023	0.017	0.017	0.014	0.015
Voc12_3	0.722	0.714	0.686	0.611	-0.040	-0.041	-0.041	-0.041	0.037	0.035	0.031	0.025
Voc12_4	0.550	0.554	0.528	0.524	-0.117	-0.114	-0.114	-0.134	0.076	0.077	0.072	0.073
Voc12_5	0.858	0.855	0.805	0.512	-0.057	-0.058	-0.064	-0.067	0.109	0.106	0.072	0.031
Voc12_6	0.877	0.876	0.853	0.677	-0.116	-0.117	-0.129	-0.164	0.368	0.359	0.285	0.111
Voc12_7	0.709	0.693	0.549	0.441	-0.040	-0.042	-0.041	-0.041	0.030	0.027	0.017	0.013
Voc12_8	0.807	0.808	0.807	0.780	-0.170	-0.169	-0.174	-0.264	0.377	0.375	0.357	0.282
Voc12_9	0.652	0.652	0.605	0.515	-0.054	-0.051	-0.052	-0.054	0.031	0.030	0.026	0.021
hja_birdsong_0	0.742	0.756	0.734	0.609	-0.111	-0.108	-0.159	-0.245	0.124	0.134	0.117	0.075
hja_birdsong_1	0.603	0.645	0.694	0.651	-0.202	-0.155	-0.129	-0.164	0.259	0.271	0.292	0.243
hja_birdsong_10	0.856	0.837	0.638	0.393	-0.041	-0.044	-0.055	-0.048	0.342	0.177	0.040	0.016
hja_birdsong_11	0.923	0.936	0.932	0.927	-0.170	-0.158	-0.101	-0.111	0.864	0.874	0.875	0.874
hja_birdsong_12	0.481	0.763	0.746	0.417	-0.040	-0.032	-0.037	-0.033	0.011	0.025	0.023	0.010
hja_birdsong_2	0.713	0.843	0.817	0.262	-0.106	-0.095	-0.123	-0.302	0.094	0.203	0.279	0.030
hja_birdsong_3	0.950	0.934	0.927	0.920	-0.079	-0.088	-0.088	-0.073	0.560	0.490	0.502	0.402
hja_birdsong_4	0.612	0.568	0.635	0.482	-0.023	-0.023	-0.028	-0.024	0.009	0.009	0.011	0.008
hja_birdsong_5	0.793	0.759	0.788	0.779	-0.060	-0.060	-0.082	-0.081	0.070	0.058	0.070	0.065
hja_birdsong_6	0.691	0.597	0.700	0.548	-0.011	-0.012	-0.015	-0.013	0.008	0.005	0.010	0.004
hja_birdsong_7	0.898	0.882	0.786	0.420	-0.061	-0.066	-0.077	-0.078	0.466	0.399	0.239	0.027
hja_birdsong_8	0.968	0.977	0.944	0.528	-0.016	-0.028	-0.050	-0.036	0.830	0.715	0.473	0.022
hja_birdsong_9	0.590	0.619	0.661	0.394	-0.034	-0.031	-0.043	-0.031	0.015	0.015	0.018	0.009
msrcv2_0	0.840	0.843	0.768	0.380	-0.157	-0.160	-0.184	-0.347	0.406	0.415	0.246	0.081
msrcv2_1	0.862	0.873	0.870	0.574	-0.179	-0.181	-0.215	-0.346	0.674	0.694	0.614	0.164
msrcv2_10	0.926	0.928	0.896	0.742	-0.080	-0.083	-0.094	-0.106	0.323	0.327	0.210	0.093
msrcv2_11	0.781	0.792	0.575	0.342	-0.065	-0.067	-0.071	-0.071	0.076	0.079	0.040	0.028
msrcv2_12	0.661	0.632	0.327	0.207	-0.046	-0.050	-0.049	-0.049	0.032	0.029	0.018	0.016
msrcv2_13	0.847	0.852	0.821	0.802	-0.029	-0.030	-0.032	-0.031	0.302	0.312	0.207	0.196
msrcv2_14	0.758	0.772	0.633	0.572	-0.033	-0.036	-0.035	-0.034	0.035	0.037	0.025	0.023
msrcv2_15	0.562	0.564	0.405	0.355	-0.074	-0.074	-0.075	-0.081	0.043	0.043	0.035	0.033
msrcv2_17	0.565	0.581	0.528	0.524	-0.032	-0.034	-0.032	-0.031	0.033	0.033	0.035	0.035
msrcv2_18	0.830	0.833	0.754	0.521	-0.153	-0.155	-0.169	-0.346	0.409	0.403	0.225	0.095
msrcv2_2	0.849	0.827	0.726	0.266	-0.168	-0.169	-0.188	-0.346	0.354	0.309	0.224	0.077
msrcv2_20	0.700	0.687	0.519	0.430	-0.038	-0.041	-0.040	-0.039	0.026	0.024	0.015	0.013
msrcv2_21	0.758	0.762	0.726	0.618	-0.098	-0.099	-0.105	-0.115	0.139	0.142	0.137	0.113
msrcv2_22	0.752	0.754	0.691	0.396	-0.066	-0.066	-0.070	-0.073	0.077	0.077	0.062	0.039
msrcv2_3	0.835	0.828	0.753	0.493	-0.076	-0.077	-0.082	-0.091	0.118	0.113	0.075	0.038
msrcv2_5	0.824	0.815	0.654	0.249	-0.052	-0.055	-0.056	-0.058	0.067	0.063	0.032	0.015
msrcv2_6	0.884	0.881	0.842	0.839	-0.149	-0.154	-0.192	-0.346	0.526	0.522	0.385	0.294
msrcv2_7	0.808	0.812	0.787	0.674	-0.039	-0.043	-0.048	-0.043	0.052	0.053	0.044	0.027
msrcv2_8	0.818	0.826	0.753	0.243	-0.046	-0.049	-0.054	-0.051	0.054	0.056	0.039	0.013
msrcv2_9	0.888	0.889	0.860	0.678	-0.080	-0.084	-0.095	-0.102	0.318	0.305	0.187	0.069

Table 2: Performance of all considered methods on the 59 non-synthetic datasets. Our methods (MIL-GP-Probit, MIL-GP-Probit-Gibbs), do better than those of [1] (MIL-GP-Logistic, MIL-GP-Logistic-LM), in terms of AUC and MAP. Despite its variational approximation, MIL-GP-Probit has comparable performance to MIL-GP-Probit-Gibbs, which characterizes the exact posterior via sampling, but is slower. MIL-GP-Logistic-LM does poorly.

References

[1] Manuel Haußmann, Fred A Hamprecht, and Melih Kandemir. Variational bayesian multiple instance learning with gaussian processes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6570–6579, 2017.