A Gaussian Latent Variable Model for Large Margin Classification of Labeled and Unlabeled Data: Supplementary Material

Do-kyum Kim, Matthew Der and Lawrence K. Saul
Department of Computer Science and Engineering, University of California, San Diego
{dok027, mfder, saul}@cs.ucsd.edu

S1 Posterior means of latent variables

For the EM algorithm of our models, it is necessary to compute the posterior means of their Gaussian latent variables. This calculation is most involved for the model in section 3.2, where we impose both large-margin constraints and a class-balancing constraint on the unlabeled examples. In this section we briefly justify the result in eq. (25) for computing the posterior means under these constraints.

As usual, we compute the model's likelihood by integrating over the Gaussian latent variables z_i and $\tilde{z}_j$ of the labeled and unlabeled examples. We denote the region of integration by:

$$\mathbf{\Omega} \;=\; \Bigg\{(z_1, z_2, \ldots, z_n, \tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_m)\;\Bigg|\; y_i z_i \geq 1 \text{ for } i=1,\ldots,n,\; |\tilde{z}_j| \geq 1 \text{ for } j=1,\ldots,m,\; \frac{1}{m}\sum_{j=1}^{m}\operatorname{sign}(\tilde{z}_j) \in [\tilde{\mu}_{\min}, \tilde{\mu}_{\max}]\Bigg\}, \tag{S1}$$

where the last three lines express the model's large-margin and class-balancing constraints. The (unregularized) likelihood of the model is given by the integral:

$$P_\Omega(\Theta) = \int_\Omega \prod_{i=1}^{n} P(z_i|\mathbf{x}_i, \Theta)\, \mathrm{d}z_i \prod_{j=1}^{m} P(\tilde{z}_j|\tilde{\mathbf{x}}_j, \Theta)\, \mathrm{d}\tilde{z}_j \tag{S2}$$

More concretely, substituting the Gaussian priors for the latent variables into this expression, we obtain:

$$P_\Omega(\Theta) = \int_\Omega \prod_{i=1}^{n} \frac{e^{-\frac{1}{2}(z_i-\xi_i)^2}}{\sqrt{2\pi}} \mathrm{d}z_i \prod_{j=1}^{m} \frac{e^{-\frac{1}{2}(\tilde{z}_j-\tilde{\xi}_j)^2}}{\sqrt{2\pi}} \mathrm{d}\tilde{z}_j, \tag{S3}$$

where $\xi_i = \mathbf{w}\cdot\mathbf{x}_i + b$ and $\tilde{\xi}_j = \mathbf{w}\cdot\tilde{\mathbf{x}}_j + b$ denote, respectively, the linear scores of the labeled and unlabeled examples.

The regularized log-likelihood of the model, with both large-margin and class-balancing constraints, is given by:

$$\mathcal{L}_{\mathrm{ss}}^{\mathrm{bal}}(\Theta) \;=\; \log P_\Omega(\Theta) - \frac{\lambda}{2}\|\mathbf{w}\|^2. \tag{S4}$$

We use $\hat{z}_j = \mathrm{E}[\tilde{z}_j|\{\tilde{\mathbf{x}}_k, \tilde{y}_k \neq 0\}_{k=1}^m, \tilde{\mu} \in [\tilde{\mu}_{\min}, \tilde{\mu}_{\max}], \Theta]$ to denote the posterior means of the latent variables for the model's unlabeled examples. From eq. (S3), we see that we can obtain these means by the method of differentiating under the integral sign:

$$\hat{z}_j \;=\; \frac{\partial}{\partial\tilde{\xi}_j}\left[\mathcal{L}_{\mathrm{ss}}^{\mathrm{bal}}(\Theta)\right] + \tilde{\xi}_j. \tag{S5}$$

The particular formulation of this result in eq. (25) follows from the decomposition of the model's log-likelihood in eq. (18).

S2 Proof of the Lyapunov condition

In Section 3 of the main paper, we use a Lyapunov central limit theorem (Billingsley, 1995) to approximate the intractable posterior distribution in eq. (19). The theorem requires the Lyapunov condition, and here we provide a proof of the condition in our setting.

To begin, consider a sequence of independent (but non-identical) random variables $\{Y_1, Y_2, \ldots\}$, each with finite mean μ_i and variance σ_i^2. The Lyapunov condition requires that certain higher moments of these variables exist; it also bounds their rate of growth. In particular, for some $\delta > 0$, the condition requires that:

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}\left[|Y_i - \mu_i|^{2+\delta}\right] = 0, \tag{S6}$$

where $s_n = \left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}$. The Lyapunov condition is sufficient to prove a central limit theorem for independent but non-identical random variables. In particular, if the condition is met, it follows that:

$$\frac{1}{s_n}\sum_{i=1}^{n}(Y_i - \mu_i) \;\xrightarrow{d}\; \mathcal{N}(0, 1);$$

this is a generalization of the standard central limit theorem for sums of i.i.d. random variables.

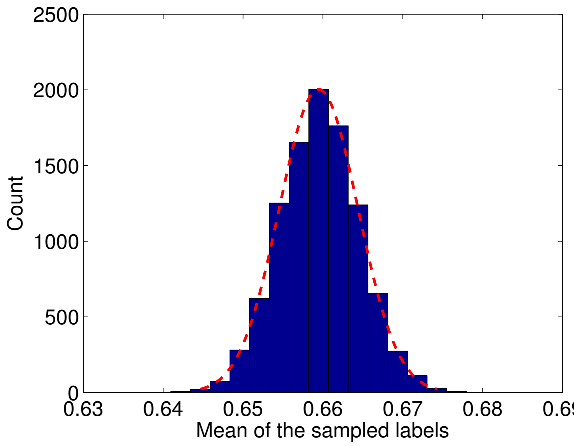

Figure S1: Histogram of the mean of the sampled labels. See text for details.

Next we verify the Lyapunov condition for the simple case of binary random variables $Y_i \in \{+1, -1\}$ with means μ_i and variances $\sigma_i^2 = 1-\mu_i^2$. In this case it is straightforward to show that:

$$\mathrm{E}\big[|Y_i - \mu_i|^4\big] \;=\; \sigma_i^2(4-3\sigma_i^2) \;\leq\; 4\sigma_i^2. \qquad \text{(S7)}$$

Now consider the ratio in eq. (S6) for the choice $\delta = 2$. From the upper bound in eq. (S7), it follows at once that:

$$0 \;\leq\; \frac{1}{s_n^4}\sum_{i=1}^{n}\mathrm{E}[|Y_i-\mu_i|^4] \;\leq\; \frac{4}{\sum_{i=1}^{n}\sigma_i^2}. \qquad \text{(S8)}$$

Note that as $n \to \infty$, the denominator on the right hand side increases without bound as long as some finite fraction of the variables Y_i have non-zero variance. Under this very weak condition, it follows at once that the ratio in eq. (S6) vanishes in the limit $n \to \infty$.

In Section 3 of the main paper, we apply the Lyapunov central limit theorem to the independent binary random variables $\tilde{y}_j \in \{-1, +1\}$. These variables store the missing labels of unlabeled examples; their posterior means and variances are given by eqs. (20–21). Note that all the variances are strictly greater than zero. Thus the Lyapunov condition also holds in this case.

S3 Empirical validation of the Lyapunov central limit theorem

Our experiments on $\mathrm{EMBLEM}_{\mathrm{ss}}^{\mathrm{bal}}$ rely on a Gaussian approximation from the Lyapunov central limit theorem. How accurate is this approximation? To investigate this question, we drew samples of the missing labels $\{\tilde{y}_j\}_{j=1}^m$ from the posterior distributions $p(\tilde{y}_j|\mathbf{x}_j, \tilde{y}_j \neq 0, \Theta)$ on the ccat data set. Fig. S1 shows a histogram of the means $\frac{1}{m}\sum_{j=1}^{m}\tilde{y}_j$ from 10000 repeated trials of this stochastic simulation; in the same plot, the dashed line shows the Gaussian approximation from the central limit theorem. The overall match is excellent.

References

P. Billingsley. *Probability and measure*. John Wiley & Sons, 3rd edition, 1995.