Summary Questions of the last lecture

What are the two options for out of sample extension in SSL?

→ One is to add the new sample to the graph and re-compute HF solution transductively and the other is to make the algorithms inductive.

Why do we have to make a classifier be smooth in inductive SSL for out of sample extension?

→ The smoothness deal with noisy samples by providing reasonable interpolation for new samples.

Summary Questions of the last lecture

What is the meaning of manifold regularization?

→ It enforces the classifier function to yield a reasonably interpolated value for any sample which is not given in the training data.

$$\lambda\Omega(\boldsymbol{f}) = \lambda_2 \boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f} + \lambda_1 \int_{\boldsymbol{x}\in\boldsymbol{X}} f(\boldsymbol{x})^2 d\boldsymbol{x}$$

For general kernels:

$$\min_{f\in\mathcal{H}_{\mathcal{K}}} \sum_{i}^{n_l} \Phi(\boldsymbol{x}_i, y_i, f(x_i))) + \lambda_1 \|f\|_{\mathcal{K}}^2 + \lambda_2 \boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f}$$

Smoothness for given samples ($\boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f}$)

Smoothness for unknown samples ($\|f\|_{\mathcal{K}}^2$)

Summary Questions of the last lecture

What is the key idea of Max-Margin Graph Cuts for SSL?

→ The self-training for SSL is done by using the confidently predicted labels obtained by regularized HF solution of SSL.

$$f^* = \underset{f \in \mathcal{H}_{\mathcal{K}}}{argmin} \sum_{i:|\ell_i^*| \geq \varepsilon} \Phi(\boldsymbol{x}_i, sgn(\ell_i^*), f(\boldsymbol{x}_i)) + \lambda \|f\|_{\mathcal{K}}^2$$

$$\text{s. t.} \quad \ell^* = \underset{\ell \in \mathbb{R}^N}{argmin}\, \ell^T (\boldsymbol{L} + \gamma_g \mathbf{I})\ \ell$$

$$\text{s. t.} \quad \ell_i = y_i, \quad \forall\, i = 1, \ldots, n_l$$

$$f^*(\boldsymbol{x}) = \sum_{i:|\ell_i^*| \geq \varepsilon} \alpha_i^* \mathcal{K}(\boldsymbol{x}, \boldsymbol{x}_i)$$

Summary Questions of the last lecture

What are the two options for online SSL?

→ One is to add the new sample to the graph and re-compute HF solution transductively and the other is to make the algorithms inductive. The key issue of the former option is to keep the computation cost and memory within a reasonable level even for continuous increment of unlabeled samples.

$$\boldsymbol{f}_u^q = (\boldsymbol{L}_{uu}^q + \gamma_g \boldsymbol{V})^{-1}(\boldsymbol{W}_{ul}^q \boldsymbol{f}_l) \text{ where } \boldsymbol{W}^q = \boldsymbol{V}\widetilde{\boldsymbol{W}}^q \boldsymbol{V}$$

Summary Questions of the last lecture

What is the key idea for keeping # of representative nodes?

→ Determine the minimum distance (R) between any two centroids and each new sample within a ball centered by a centroid and with radius R is added to the cluster. If there are samples that do not belong to any cluster, R is doubled.