Ensemble approaches to semi-supervised learning
Tanha, J.

Citation for published version (APA):
Tanha, J. (2013). Ensemble approaches to semi-supervised learning
Manifold

In graph based methods the inconsistency between $\{(x_i, y_i)\}_{i=1}^n$ and similarity matrix is:

$$F(y) = \sum_{i,j} S(x_i, x_j)(y_i - y_j)^2 = y^T Ly \quad (A.1)$$

where $L$ is the graph laplacian. We use the exponential function (5.18) in order to measure the inconsistency between the similarity and classifier predictions which can be reformulated as (using power series):

$$F(y, H) = \sum_{i,j} S(x_i, x_j)cosh((H(x_i) - H(x_j), \overrightarrow{1})) \simeq \sum_{i,j} S(x_i, x_j)$$
$$+ \sum_{i,j} S(x_i, x_j)((H(x_i) - H(x_j), \overrightarrow{1}))^2 \quad (A.2)$$

In (A.2), the first term is constant and the second term is equivalent to graph laplacian which is used as manifold regularization in many graph based methods. Then minimizing the inconstancy in (5.18) is equivalent to minimizing graph laplacian as in graph based methods [4] which are based on the manifold assumption.
Proof.1

Let the function \( \langle Y_i, e_k \rangle \) be equal to 1, if a data point with \( Y_i \) belongs to class \( k \). Then (5.22) can be rewritten as follows:

\[
F = \arg\min \sum_{i=1}^{n_1} \exp\left(-\frac{1}{K} \langle H^{t-1}(x_i) + \beta^t h^t(x_i), Y_i \rangle \right) + \sum_{i=1}^{n_2} \sum_{j=1}^{n_u} S^{lu}(x_i, x_j) e^{-\frac{1}{K} \sum_{k \in \ell(Y_i, e_k)} (H_{i}^{t-1} + \beta^t h^t, e_k)} + \sum_{i,j \in n_u} S^{uu}(x_i, x_j) e^{-\frac{1}{K} \sum_{k \in \ell} (H_{j}^{t-1} + \beta^t h^t, e_k)}
\]

(A.3)

Using the following inequality in (5.22):

\[
(c_1 c_2 ... c_n)^{\frac{1}{n}} \leq \frac{c_1 + c_2 + ... + c_n}{n}
\]

(A.4)

and then replacing the value of \( Y_i \) results in the (5.23), where \( \delta \) is defined in (5.24).

Proof.2

According to the formulation that we mentioned earlier, \( \frac{1}{K-1} \leq \beta^t h(x_i), e_k \leq 1 \), then multiplying it with \( \frac{1}{K-1} \) and using the exponential function gives the following inequality:

\[
e^{\frac{1}{K-1} \beta^t h(x_i)} \leq e^{\frac{1}{K-1} h(x_i), e_k} \leq e^{\frac{1}{K-1}}
\]

(A.5)

Using (B.4) for decomposing the third term of (5.23) leads to the following expression:

\[
e^{\frac{1}{K-1} (h(x_i) - h(x_j), e_k)} \leq e^{\frac{1}{K-1}} e^{-\frac{\beta^t}{K-1} (h(x_j), e_k)}
\]

(A.6)

Replacing the inequality (B.5) in the last term of (5.23) gives:

\[
F_1 \leq \sum_{i=1}^{n_1} \exp\left(-\frac{1}{K} \langle H^{t-1}(x_i), Y_i \rangle \right) \exp\left(-\frac{\beta^t}{K} (h^t, Y_i) \right) + \sum_{i=1}^{n_2} \sum_{j=1}^{n_1} \sum_{k \in \ell} S^{lu}(x_i, x_j) \exp\left(-\frac{1}{K-1} \langle H^{t-1}_j, e_k \rangle \right) \exp\left(-\frac{\beta^t}{K-1} (h^t_j, e_k) \right) \delta(Y_i, e_k) + \sum_{i,j \in n_u} \sum_{k \in \ell} S^{uu}(x_i, x_j) \exp\left(-\frac{1}{K-1} \langle H^{t-1}_j - H^{t-1}_i, e_k \rangle \right) \left( e^{\frac{1}{K-1}} e^{-\frac{\alpha}{K-1} h^t_j} \right)
\]

(A.7)
Factoring out the common term $e^{\frac{\beta t}{K}(h_i^t, e_k)}$ and re-arranging terms then results in:

$$F_1 \leq \sum_{i=1}^{n_i} \exp\left(-\frac{1}{K} \langle H_i^{t-1}, Y_i \rangle\right) \exp\left(-\frac{\beta t}{K} \langle h_i^t, Y_i \rangle\right) + \sum_{j=1}^{n_n} \sum_{k \in l} \exp\left(-\frac{\beta t}{K-1} \langle h_j^t, e_k \rangle\right) \left(\sum_{i=1}^{n_i} S^{uu}(x_i, x_j) \exp\left(-\frac{1}{K-1} \langle H_j^{t-1}, e_k \rangle\right)\right) \delta(Y_i, e_k)$$

As a result, it gives:

$$F_1 \leq \sum_{i=1}^{n_i} W_i \exp\left(-\frac{1}{K} \langle H_i^{t-1}, Y_i \rangle\right) + \sum_{i=1}^{n_i} \sum_{k \in l} P_{i,k} \exp\left(-\frac{\beta t}{K-1} \langle h_i^t, e_k \rangle\right)$$

where

$$W_i = \exp\left(-\frac{1}{K} \langle H_i^{t-1}, Y_i \rangle\right)$$

and

$$P_{i,k} = \sum_{j=1}^{n_n} S^{uu}(x_i, x_j) e^{\frac{\beta t}{K-1} \langle h_j^{t-1}, e_k \rangle} \delta(Y_i, e_k) + \sum_{j=1}^{n_n} S^{uu}(x_i, x_j) e^{\frac{\beta t}{K-1} \langle h_j^{t-1}, H_i^{t-1} - e_k \rangle} e^{\frac{\beta t}{K-1} \langle h_j^{t-1}, e_k \rangle}$$

**Proof.3**

We use the following inequality to decompose the elements of $F_2$:

$$\forall x \in [-1, 1] \quad \exp(\beta x) \leq \exp(\beta) + \exp(-\beta) + \beta x - 1,$$  \hspace{1cm} (A.12)

Replacing the inequality (B.10) in (5.25) gives:

$$F_2 \leq \sum_{i=1}^{n_i} W_i (e^{\frac{\beta t}{K}} + e^{\frac{\beta t}{K}} - 1) - \sum_{i=1}^{n_i} W_i \left(\frac{\beta t}{K} \langle h_i^t, Y_i \rangle\right) + \sum_{i=1}^{n_i} \sum_{k \in l} P_{i,k} (e^{\frac{\beta t}{K-1}} + e^{\frac{\beta t}{K-1}} - 1) - \sum_{i=1}^{n_i} \sum_{k \in l} P_{i,k} \left(\frac{\beta t}{K-1} \langle h_i^t, e_k \rangle\right)$$

Then, re-arranging the above inequality gives:

$$F_2 \leq \left(\sum_{i=1}^{n_i} W_i (e^{\frac{\beta t}{K}} + e^{\frac{\beta t}{K}} - 1) + \sum_{i=1}^{n_i} \sum_{k \in l} P_{i,k} (e^{\frac{\beta t}{K-1}} + e^{\frac{\beta t}{K-1}} - 1)\right)$$

$$- \left(\sum_{i=1}^{n_i} W_i \left(\frac{\beta t}{K} \langle h_i^t, Y_i \rangle\right) + \sum_{i=1}^{n_i} \sum_{k \in l} P_{i,k} \left(\frac{\beta t}{K-1} \langle h_i^t, e_k \rangle\right)\right)$$

(A.14)
The first term in inequality (A.14) is independent of $h_i$. Hence, to minimize $F_2$, finding the examples with the largest $P_{i,k}$ and $W_i$ is sufficient at each iteration of the boosting process. We assign the value $P_{i,k}$ of selected examples as weight for the newly-labeled examples, hence $w'_i = \max |P_{i,k}|$. To prove the second part of the proposition, we expand $F_2$ as follows:

\[
F_2 = \sum_{i=1}^{n_l} W_i \exp\left(-\frac{\beta t}{K} (h_i^t, Y_i)\right) + \sum_{i=1}^{n_n} \sum_{k \in l} P_{i,k} \exp\left(-\frac{\beta t}{K-1} (h_i^t, e_k)\right)
\]

\[
= \sum_{i=1}^{n_l} W_i \exp\left(-\frac{\beta t}{K-1}\right) + \sum_{i=1}^{n_n} W_i \exp\left(\frac{\beta t}{(K-1)^2}\right) + \sum_{i=1}^{n_n} \sum_{k \in l} P_{i,k} \exp\left(-\frac{\beta t}{K-1}\right) \delta'(h_i^t, e_k, P_i = k) + \sum_{i=1}^{n_n} \sum_{k \in l} P_{i,k} \exp\left(\frac{\beta t}{(K-1)^2}\right) \delta'(h_i^t, e_k, P_i \neq k)
\]

(A.15)

where $\delta'$ is defined as:

\[
\delta'(h_i^t, e_k, T) = \begin{cases} 
1 & \text{if } T \text{ is true} \\
0 & \text{otherwise} 
\end{cases}
\]

(A.16)

Differentiating the above equation w.r.t $\beta$ and equating it to zero, gives:

\[
\frac{\partial F_2}{\partial \beta} = \frac{-1}{K-1} \exp\left(-\frac{\beta}{K-1}\right) \sum_{i \in n_l} W_i \\
+ \frac{1}{(K-1)^2} \exp\left(-\frac{\beta}{(K-1)^2}\right) \sum_{i \in n_l} W_i \\
+ \frac{-1}{K-1} \exp\left(-\frac{\beta}{K-1}\right) \sum_{i=1}^{n_n} \sum_{k \in l} P_{i,k} \delta'(h_i^t, e_k, P_i = k) \\
+ \frac{1}{(K-1)^2} \exp\left(-\frac{\beta}{(K-1)^2}\right) \sum_{i=1}^{n_n} \sum_{k \in l} P_{i,k} \delta'(h_i^t, e_k, P_i \neq k) = 0
\]

(A.17)

Simplifying the above equation results in (5.29).