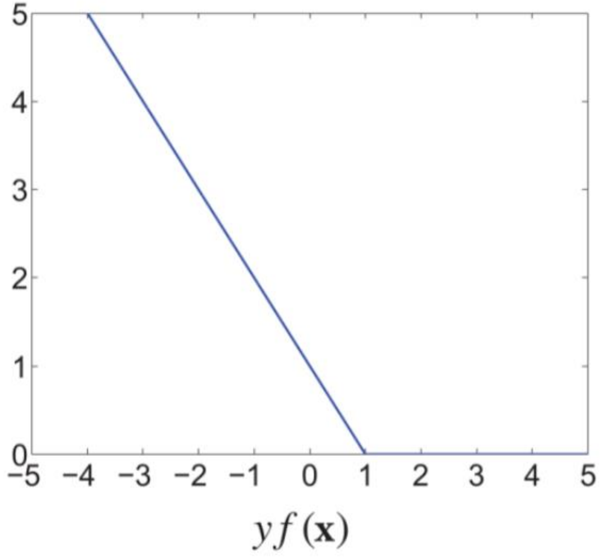
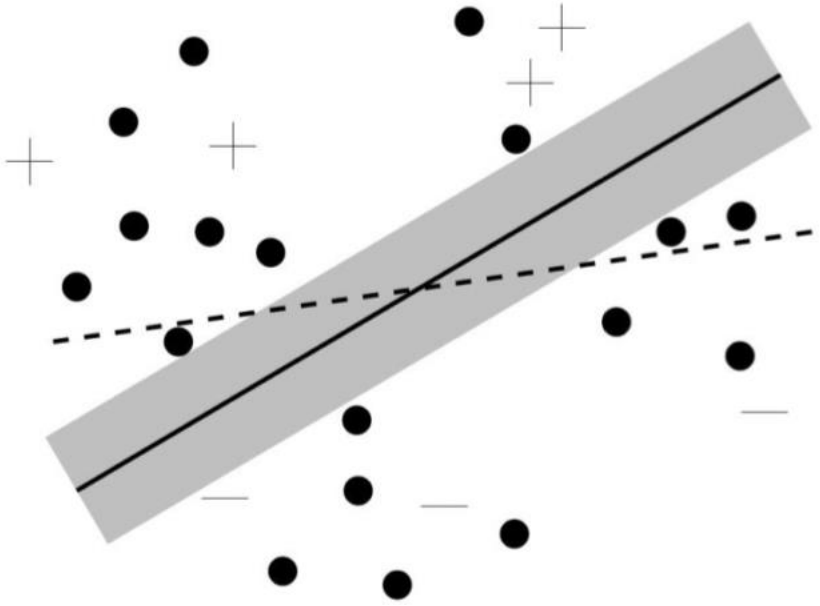
Questions of the last lecture

- What does hinge loss in SVM penalize?

→ Hinge loss penalizes the case that the classifier f(w, x) does not decide the correct class of labeled x with the score margin of y f(w, x)>=1.



hinge loss

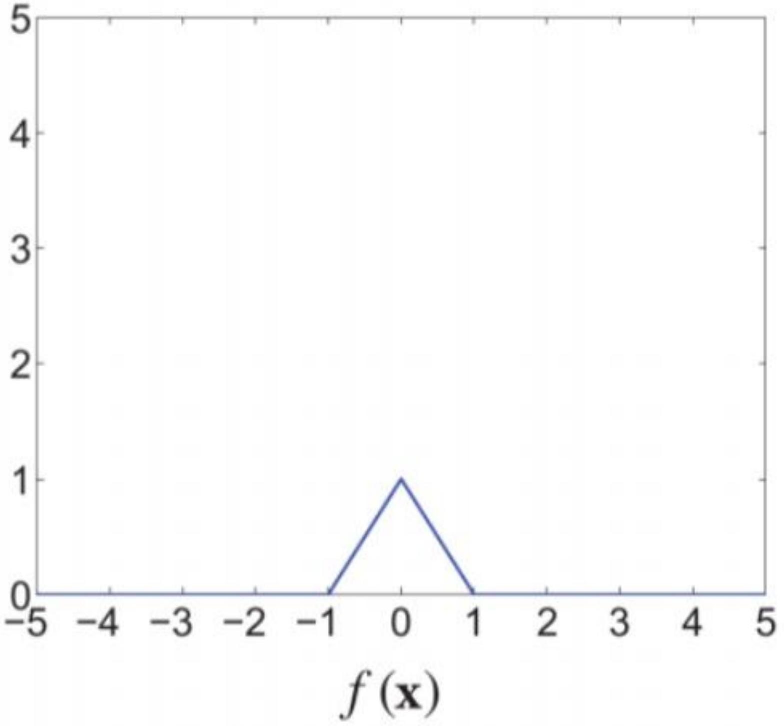


$$\Phi(\boldsymbol{x}_i, y_i, \boldsymbol{f}(\mathbf{w}, b; \boldsymbol{x}_i)) = \max(1 - y_i(\mathbf{w}^T \boldsymbol{x}_i + b), 0)$$

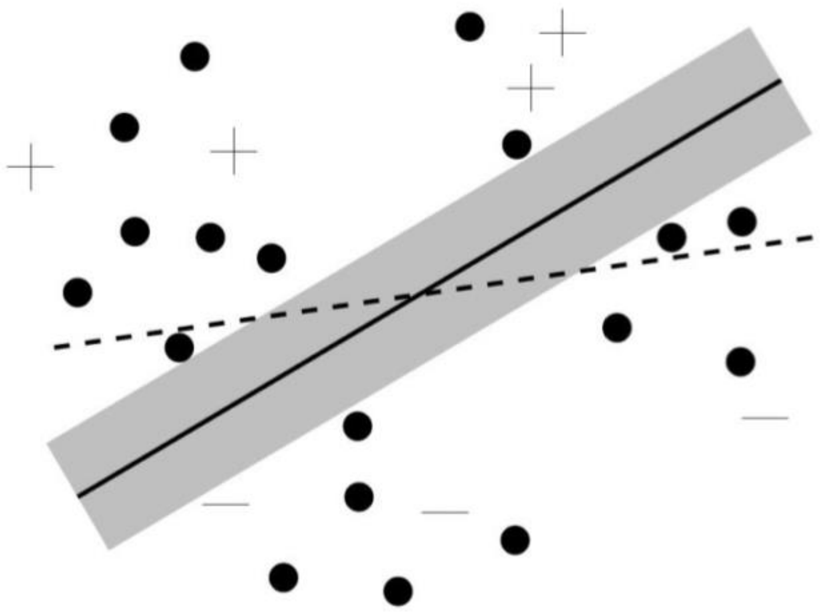
Questions of the last lecture

- What does hat loss in SVM for semi-supervised learning penalize?

→Hat loss penalizes the case that the classifier f(w, x) does not decide any class of unlabeled x with the score margin of |f(w, x)|>=1.



(b) the hat loss



Questions of the last lecture

- Why can't we use eigenvectors to solve MinCut-based SSL in graph?

→ It is because the eigenvector solution can not consider the labeled data.

Transductive SSL with graph: fixing $f(\boldsymbol{x}_i)$ for $i \in \mathcal{L}$

$$\min_{f \in \{\pm 1\}^{n_l+n_u}} \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j) \right)^2 + \infty \sum_{i}^{n_l} (f(\boldsymbol{x}_i) - y_i)^2$$

Solution:

An integer program: NP hard

Can we use eigenvectors? No. Why?

We need a better way to reflect the confidence.

Questions of the last lecture

- What is the meaning of harmonic function in SSL?

→ The harmonic function makes the label of a node to be harmonious(similar) with those of its neighboring nodes.

Relaxation: Transductive SSL with graph: fixing $f(\boldsymbol{x}_i)$ for $i \in \mathcal{L}$

$$\min_{f \in \boldsymbol{R}^{n_l+n_u}} \lambda \sum_{i,j=1}^{n_l+n_u} w_{ij} \left(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)\right)^2 + \infty \sum_{i}^{n_l} (f(\boldsymbol{x}_i) - y_i)^2$$

Naïve Solution

Right term solution: constrain f to match the supervised data

$$f(\boldsymbol{x}_i) = y_i \quad \forall\, i \in \{1, \dots, n_l\}$$

Left term solution: enforce the solution f to be harmonic (cf. aggregation, rw)

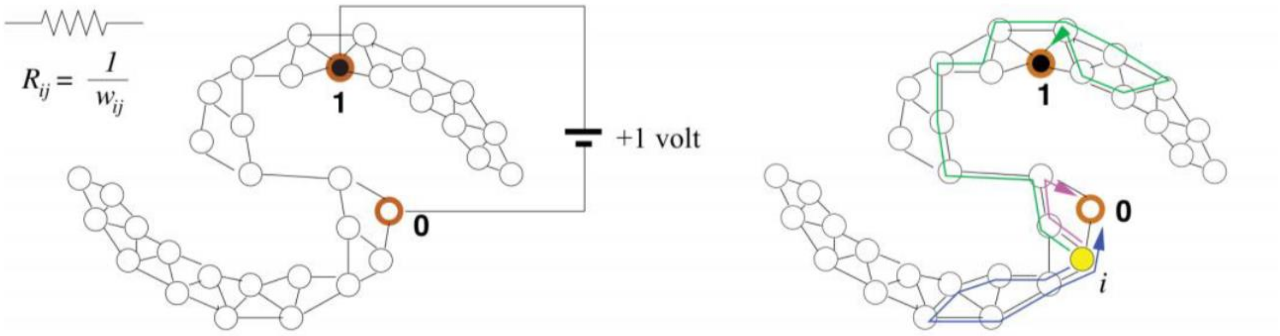
$$f(\boldsymbol{x}_i) = \frac{\sum_{ij} f(\boldsymbol{x}_j) w_{ij}}{\sum_{ij} w_{ij}} \quad \forall\, i \in \{n_l + 1, \dots, n_l + n_u\}$$

$$\boldsymbol{f}_u = \boldsymbol{L}_{uu}^{-1}(-\boldsymbol{L}_{ul}\boldsymbol{f}_l) = \boldsymbol{L}_{uu}^{-1}(\boldsymbol{W}_{ul}\boldsymbol{f}_l)$$

Questions of the last lecture

- What is random work interpretation of harmonic function-based SSL in graph?

→ The label of a node is assigned by the average of the harmonic labels of the vertices that are hit by random works.



(a) The electric network interpretation (b) The random walk interpretation

Random walk interpretation :

1) start from the vertex you want to label and randomly walk
2) $P(j|i) = \frac{w_{ij}}{\sum_k w_{ik}} \Leftrightarrow \boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{W}$
3) finish when a labeled vertex is hit

$$f(\boldsymbol{x}_i) = \frac{\sum_{ij} f(\boldsymbol{x}_j) w_{ij}}{\sum_{ij} w_{ij}}$$

4) $f(\boldsymbol{x}_i)$ is assigned by average of the labels of the hit vertices.

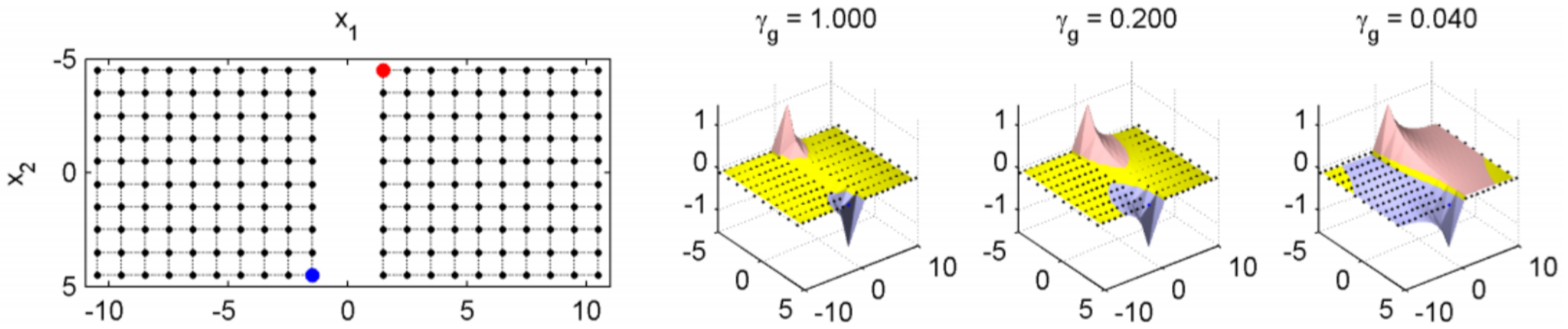
Questions of the last lecture

- What is a key point of regularized harmonic function-based SSL in graph?

→ A sink node is added to allow the random work to die at any nodes, which reduces the misleading by outliers.

$$\boldsymbol{f}_u = (\boldsymbol{L}_{uu} + \gamma_g \mathbf{I})^{-1}(\boldsymbol{W}_{ul}\boldsymbol{f}_l),$$

How does γ_g influence the solution?



Questions of the last lecture

- What is a key point of soft harmonic function-based SSL in graph?

→ The labeled data is not constrained strictly, where noisy labels may be smoothed by the soft harmonic function

Regularized HS objective with $\boldsymbol{Q} = \boldsymbol{L} + \gamma_g \mathbf{I}$:

Define $f_i \triangleq f(\boldsymbol{x}_i)$, $\boldsymbol{f} \triangleq [f_i, \dots, f_{n_l+n_u}]$

$$\min_{\boldsymbol{f} \in \mathbb{R}^{n_l+n_u}} \infty \sum_{i=1}^{n_l} (f_i - y_i)^2 + \lambda \boldsymbol{f}^T \boldsymbol{Q} \boldsymbol{f}$$

Soft constraints for $f(\boldsymbol{x}_i) = y_i, \forall i \in \mathcal{L}$: ∞ is replaced by finite values

$$\min_{\boldsymbol{f} \in \mathbb{R}^{n_l+n_u}} (\boldsymbol{f} - \boldsymbol{y})^T \boldsymbol{C} (\boldsymbol{f} - \boldsymbol{y}) + \boldsymbol{f}^T \boldsymbol{Q} \boldsymbol{f}$$