Graph Based Semi-supervised Learning

Aydın Gerek

Marmara University

May 26th, 2018

Learning Paradigms

- Supervised Learning

Learning Paradigms

- Supervised Learning
- Unsupervised Learning

Learning Paradigms

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning

Learning Paradigms

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
  - Inductive

Learning Paradigms

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
  - Inductive
  - Transductive

Graphs

$V =$ set of vertices

Graphs

$V =$ set of vertices

$E =$ set of edges

Graphs

$V =$ set of vertices $\qquad E =$ set of edges

$$A_{uv} = \begin{cases} 1 & (u, v) \in E \\ 0 & (u, v) \notin E \end{cases}$$

Graphs

$V =$ set of vertices $\qquad E =$ set of edges

$$A_{uv} = \begin{cases} w_{uv} & (u,v) \in E \\ 0 & (u,v) \notin E \end{cases}$$

Graphs

V = set of vertices $\qquad$ E = set of edges

$$A_{uv} = \begin{cases} w_{uv} & (u,v) \in E \\ 0 & (u,v) \notin E \end{cases} \qquad D_{uv} = \begin{cases} \sum_w A_{uw} & u = v \\ 0 & u \neq v \end{cases}$$

Graphs

V = set of vertices $\qquad$ E = set of edges

$$A_{uv} = \begin{cases} w_{uv} & (u,v) \in E \\ 0 & (u,v) \notin E \end{cases} \qquad D_{uv} = \begin{cases} \sum_w A_{uw} & u = v \\ 0 & u \neq v \end{cases}$$

$$\Delta = D - A$$

Graphs

V = set of vertices $\qquad E$ = set of edges

$$A_{uv} = \begin{cases} w_{uv} & (u,v) \in E \\ 0 & (u,v) \notin E \end{cases} \qquad D_{uv} = \begin{cases} \sum_w A_{uw} & u = v \\ 0 & u \neq v \end{cases}$$

$$\Delta = D - A \qquad L = D^{-1/2} \Delta D^{-1/2}$$

Gaussian Random Fields (GRF) (Zhu et al., 2003)

also known as Label Propagation

$$\arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \sum_{(u,v)\in E} A_{uv}(\hat{Y}_{u\ell} - \hat{Y}_{v\ell})^2$$

Gaussian Random Fields (GRF) (Zhu et al., 2003)

also known as Label Propagation

$$\arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \sum_{(u,v)\in E} A_{uv}(\hat{Y}_{u\ell} - \hat{Y}_{v\ell})^2$$

$$\text{s.t. } Y_{u\ell} = \hat{Y}_{u\ell}, \forall \ell \in \mathcal{Y}, \; \forall \text{ labeled } u$$

Gaussian Random Fields (GRF) (Zhu et al., 2003)

also known as Label Propagation

$$\underset{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}}{\arg\min}\sum_{\ell\in\mathcal{Y}}\sum_{(u,v)\in E}A_{uv}(\hat{Y}_{u\ell}-\hat{Y}_{v\ell})^2=\underset{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}}{\arg\min}\sum_{\ell\in\mathcal{Y}}{\hat{Y}_\ell}^T\Delta\hat{Y}_\ell$$

$$\text{s.t. } Y_{u\ell}=\hat{Y}_{u\ell},\forall\ell\in\mathcal{Y},\ \forall \text{ labeled } u$$

Gaussian Random Fields (GRF) (Zhu et al., 2003)

also known as Label Propagation

$$\arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \sum_{(u,v)\in E} A_{uv}(\hat{Y}_{u\ell} - \hat{Y}_{v\ell})^2 = \arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \hat{Y}_\ell{}^T \Delta \hat{Y}_\ell$$

$$\text{s.t. } Y_{u\ell} = \hat{Y}_{u\ell}, \forall \ell \in \mathcal{Y}, \ \forall \text{ labeled } u$$

$$\hat{Y}^{(t+1)}_{v\ell} \leftarrow \frac{\sum_{(u,v)\in E} A_{uv} \hat{Y}^{(t)}_{u\ell}}{\sum_{(u,v)\in E} A_{uv}}$$

Gaussian Random Fields (GRF) (Zhu et al., 2003)

also known as Label Propagation

$$\arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \sum_{(u,v)\in E} A_{uv}(\hat{Y}_{u\ell} - \hat{Y}_{v\ell})^2 = \arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \hat{Y}_\ell{}^T \Delta \hat{Y}_\ell$$

$$\text{s.t. } Y_{u\ell} = \hat{Y}_{u\ell}, \forall \ell \in \mathcal{Y}, \ \forall \text{ labeled } u$$

$$\hat{Y}_{v\ell}^{(t+1)} \leftarrow \frac{\sum_{(u,v)\in E} A_{uv}\hat{Y}_{u\ell}^{(t)}}{D_{vv}}$$

Gaussian Random Fields (GRF) (Zhu et al., 2003)

also known as Label Propagation

$$\arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \sum_{(u,v)\in E} A_{uv}(\hat{Y}_{u\ell} - \hat{Y}_{v\ell})^2 = \arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \hat{Y}_\ell{}^T \Delta \hat{Y}_\ell$$

$$\text{s.t. } Y_{u\ell} = \hat{Y}_{u\ell}, \forall \ell \in \mathcal{Y}, \ \forall \text{ labeled } u$$

$$\hat{Y}_{v\ell}^{(t+1)} \leftarrow \frac{\sum_{(u,v)\in E} A_{uv} \hat{Y}_{u\ell}^{(t)}}{D_{vv}}$$

$$\hat{Y}^{(t+1)} \leftarrow D^{-1} A \hat{Y}^{(t)}$$

Gaussian Random Fields (GRF) (Zhu et al., 2003)

also known as Label Propagation

$$\arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \sum_{(u,v)\in E} A_{uv}(\hat{Y}_{u\ell} - \hat{Y}_{v\ell})^2 = \arg\min_{\hat{Y}\in\mathbb{R}^{n\times|\mathcal{Y}|}} \sum_{\ell\in\mathcal{Y}} \hat{Y}_\ell^{\,T} \Delta \hat{Y}_\ell$$

$$\text{s.t. } Y_{u\ell} = \hat{Y}_{u\ell}, \forall \ell \in \mathcal{Y}, \ \forall \text{ labeled } u$$

$$\hat{Y}_{v\ell}^{(t+1)} \leftarrow \frac{\sum_{(u,v)\in E} A_{uv} \hat{Y}_{u\ell}^{(t)}}{D_{vv}}$$

$$\hat{Y}^{(t+1)} \leftarrow D^{-1} A \hat{Y}^{(t)}$$

Harmonic Property: node = average of neighbors

Manifold Regularization (Belkin et al., 2006)

Manifold Hypothesis: The data lives in a low dimensional manifold embedded in high-dimensional ambient space.

Manifold Regularization (Belkin et al., 2006)

Manifold Hypothesis: The data lives in a low dimensional manifold embedded in high-dimensional ambient space.

$$f^*(x) = \arg\min_{f \in \mathcal{H}_K} \frac{1}{n_l} \sum_{i=1}^{n_l} \mathcal{L}(x_i, y_i, f)$$

Manifold Regularization (Belkin et al., 2006)

Manifold Hypothesis: The data lives in a low dimensional manifold embedded in high-dimensional ambient space.

$$f^*(x) = \arg\min_{f \in \mathcal{H}_K} \frac{1}{n_l} \sum_{i=1}^{n_l} \mathcal{L}(x_i, y_i, f) + \lambda_A \|f\|_K^2$$

Manifold Hypothesis: The data lives in a low dimensional manifold embedded in high-dimensional ambient space.

$$f^*(x) = \arg\min_{f \in \mathcal{H}_K} \frac{1}{n_l} \sum_{i=1}^{n_l} \mathcal{L}(x_i, y_i, f) + \lambda_A \|f\|_K^2 + \lambda_I \|f\|_I^2$$

Manifold Regularization (Belkin et al., 2006)

Manifold Hypothesis: The data lives in a low dimensional manifold embedded in high-dimensional ambient space.

$$f^*(x) = \arg\min_{f \in \mathcal{H}_K} \frac{1}{n_l} \sum_{i=1}^{n_l} \mathcal{L}(x_i, y_i, f) + \lambda_A \|f\|_K^2 + \lambda_I \|f\|_I^2$$

$$\|f\|_I^2 = \int_{x \in \mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 d\mathcal{P}(x)$$

Manifold Regularization (Belkin et al., 2006)

Manifold Hypothesis: The data lives in a low dimensional manifold embedded in high-dimensional ambient space.

$$f^*(x) = \underset{f \in \mathcal{H}_K}{\arg\min} \frac{1}{n_l} \sum_{i=1}^{n_l} \mathcal{L}(x_i, y_i, f) + \lambda_A \|f\|_K^2 + \lambda_I \|f\|_I^2$$

$$\|f\|_I^2 = \int_{x \in \mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 d\mathcal{P}(x) \approx \frac{1}{n_l + n_u} \mathbf{f}^T \Delta \mathbf{f}$$

Manifold Regularization (Belkin et al., 2006)

Manifold Hypothesis: The data lives in a low dimensional manifold embedded in high-dimensional ambient space.

$$f^*(x) = \arg\min_{f \in \mathcal{H}_K} \frac{1}{n_l} \sum_{i=1}^{n_l} \mathcal{L}(x_i, y_i, f) + \lambda_A \|f\|_K^2 + \lambda_I \|f\|_I^2$$

$$\|f\|_I^2 = \int_{x \in \mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 d\mathcal{P}(x) \approx \frac{1}{n_l + n_u} \mathbf{f}^T \Delta \mathbf{f}$$

$$\mathbf{f}_i^T = f(x_i)$$

Manifold Regularization (Belkin et al., 2006)

Manifold Hypothesis: The data lives in a low dimensional manifold embedded in high-dimensional ambient space.

$$f^*(x) = \arg\min_{f \in \mathcal{H}_K} \frac{1}{n_l} \sum_{i=1}^{n_l} \mathcal{L}(x_i, y_i, f) + \lambda_A \|f\|_K^2 + \lambda_I \|f\|_I^2$$

$$\|f\|_I^2 = \int_{x \in \mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 d\mathcal{P}(x) \approx \frac{1}{n_l + n_u} \mathbf{f}^T L \mathbf{f}$$

$$\mathbf{f}_i^T = f(x_i)$$

Graph Convolutional Networks (Kipf and Welling 2017)

$X \in \mathbb{R}^{n \times d_i}$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n \times d_i}, \; Z \in \mathbb{R}^{n \times d_o}$$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n \times d_i}, \; Z \in \mathbb{R}^{n \times d_o}$$
$$H^{(l)} \in \mathbb{R}^{n \times d_l}$$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n \times d_i},\ Z \in \mathbb{R}^{n \times d_o}$$

$$H^{(l)} \in \mathbb{R}^{n \times d_l},\ H^{(0)} = X$$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n \times d_i},\; Z \in \mathbb{R}^{n \times d_o}$$
$$H^{(l)} \in \mathbb{R}^{n \times d_l},\; H^{(0)} = X, H^{(L)} = Z$$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n \times d_i},\ Z \in \mathbb{R}^{n \times d_o}$$
$$H^{(l)} \in \mathbb{R}^{n \times d_l},\ H^{(0)} = X, H^{(L)} = Z$$

$$H^{(l+1)} = f(H^{(l)}, A)$$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n\times d_i},\ Z \in \mathbb{R}^{n\times d_o}$$
$$H^{(l)} \in \mathbb{R}^{n\times d_l},\ H^{(0)} = X, H^{(L)} = Z$$

$$H^{(l+1)} = f(H^{(l)}, A)$$
$$f(H^{(l)}, A) = \sigma(AH^{(l)}W^{(l)})$$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n \times d_i}, \; Z \in \mathbb{R}^{n \times d_o}$$

$$H^{(l)} \in \mathbb{R}^{n \times d_l}, \; H^{(0)} = X, H^{(L)} = Z$$

$$H^{(l+1)} = f(H^{(l)}, A)$$

$$f(H^{(l)}, A) = \sigma(AH^{(l)} W^{(l)})$$

$$\widetilde{A} = A + I$$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n\times d_i},\ Z \in \mathbb{R}^{n\times d_o}$$
$$H^{(l)} \in \mathbb{R}^{n\times d_l},\ H^{(0)} = X, H^{(L)} = Z$$

$$H^{(l+1)} = f(H^{(l)}, A)$$
$$f(H^{(l)}, A) = \sigma(AH^{(l)}W^{(l)})$$
$$\widetilde{A} = A + I, \qquad D^{-1/2}AD^{-1/2}$$

Graph Convolutional Networks (Kipf and Welling 2017)

$$X \in \mathbb{R}^{n \times d_i}, \; Z \in \mathbb{R}^{n \times d_o}$$
$$H^{(l)} \in \mathbb{R}^{n \times d_l}, \; H^{(0)} = X, H^{(L)} = Z$$

$$H^{(l+1)} = f(H^{(l)}, A)$$
$$f(H^{(l)}, A) = \sigma(AH^{(l)}W^{(l)})$$
$$\widetilde{A} = A + I, \qquad D^{-1/2}AD^{-1/2}$$
$$f(H^{(l)}, A) = \sigma(\widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}H^{(l)}W^{(l)})$$