Semi-Supervised Learning with Graphs

Defn.

Undirected graph: $G = (V, E)$

- V = set of vertices
- E = set of edges $e = (u, v)$

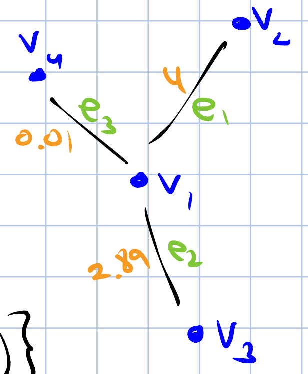
  $[u, v \in V,\ u \neq v]$

Defn

Weighted graph: (G, W)

- G = graph
- $W : E \to \mathbb{R}^+$ $(w(e) = 0 \iff \text{no edge})$

Ex.



$V = \{v_1, v_2, v_3, v_4\}$

$E = \{(v_1, v_2), (v_1, v_3), (v_1, v_4)\}$

$w(e_1) = 4, \quad w(e_2) = 2.89 \quad w(e_3) = 0.01$

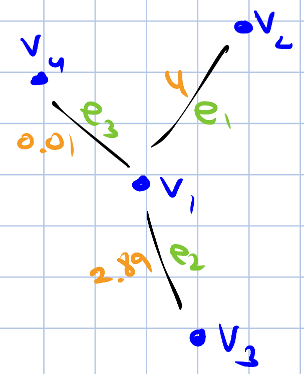
Incidence matrix:

- (G, W) - a weighted graph
- $V = \{v_1, \_\_, v_n\}$, $E = \{e_1, \_\_, e_m\}$
- $M \in \mathbb{R}^{m \times n}$ (m rows n colums)
- $M_{ij} = \begin{cases} \sqrt{w(e_i)} & e_i = (v_k, v_j) \\ -\sqrt{w(e_i)} & e_i = (v_j, v_k) \\ 0 & \text{else} \end{cases}$

  edge index

  vertex index

Ex.



$$M = \begin{array}{c|cccc|}
 & v_1 & v_2 & v_3 & v_4 \\
e_1 & -2 & 2 & 0 & 0 \\
e_2 & -1.7 & 0 & 1.7 & 0 \\
e_3 & -0.1 & 0 & 0 & 0.1
\end{array}$$

$E = \{(v_1, v_2), (v_1, v_3), (v_1, v_4)\}$

⊛ Graph Laplacian:

$$L = M^T M \in \mathbb{R}^{n \times n}$$

can show! $L_{i,j} = \begin{cases} -w((v_i, v_j)) & i \neq j \\ \deg(v_i) & v_i = v_j \end{cases}$

vertex index

$$\deg(v_i) = \sum_{j \neq i} w((v_i, v_j))$$

Ex.

$$L = \begin{array}{cccc|c}
v_1 & v_2 & v_3 & v_4 & \\
\hline
6.9 & -4 & -2.89 & -0.01 & v_1 \\
-4 & 4 & 0 & 0 & v_2 \\
-2.89 & 0 & 2.89 & 0 & v_3 \\
-0.01 & 0 & 0 & 0.01 & v_4
\end{array}$$

DETOUR - CALCULUS

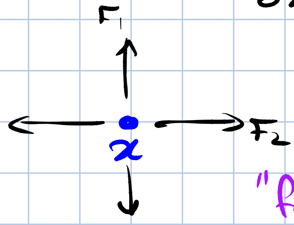
$f: \mathbb{R}^n \to \mathbb{R}$ $\qquad \nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$

(gradient)

$\vec{F} = \mathbb{R}^n \to \mathbb{R}^n$ $\qquad \nabla \cdot \vec{F} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \ldots + \frac{\partial F_n}{\partial x_n}$

$\vec{F} = (F_1, F_2, \ldots, F_n)$

(divergence)



"flux out of x"

Laplacian: $\nabla \cdot \nabla f = \frac{\partial^2 f}{\partial x_1^2} + \_\_\_ + \frac{\partial^2 f}{\partial x_n^2}$

Suppose $f: V \to \mathbb{R}$, $f = \begin{pmatrix} f(v_1) \\ f(v_2) \\ \vdots \\ f(v_n) \end{pmatrix}$

$M =$ incidence matrix.

$$\underbrace{(\underset{m \times n}{M} \cdot \underset{n \times 1}{f})}_{m \times 1}{}_i = \sum_{j=1}^{n} M_{i,j} \cdot f(v_j)$$

$$= \underbrace{\sqrt{w(e_i)}}_{??} \; \underbrace{(f(v_{j_2}) - f(v_{j_1}))}_{} \rightarrow \partial f$$

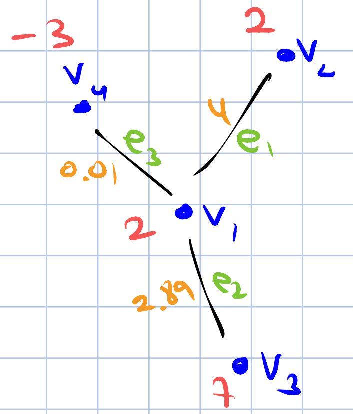
$e_i = (v_{j_1}, v_{j_2})$

$\frac{1}{\text{"distance"}} \rightarrow \frac{1}{\partial x_i}$

"$= \partial_{e_i} f$"

$$\text{"}M \cdot f = \nabla f\text{"}$$

Ex.



$$M = \begin{pmatrix} -2 & 2 & 0 & 0 \\ -1.7 & 0 & 1.7 & 0 \\ -0.1 & 0 & 0 & 0.1 \end{pmatrix}$$

$$f = \begin{pmatrix} 2 \\ 2 \\ 7 \\ -3 \end{pmatrix}$$

$$M \cdot f = \begin{pmatrix} 2 & \overbrace{(2-2)}^{\partial f} \\ 1.7 & (7-2) \\ 0.1 & (-3-2) \end{pmatrix} = \begin{pmatrix} 0 \\ 8.5 \\ -0.5 \end{pmatrix}$$

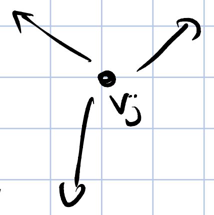
Suppose: $g: E \to \mathbb{R}$

$$\underbrace{(M^T g)_j}_{n\times 1} = \sum_{i=1}^{m} M_{ij}\, g(e_i)$$

$n \times m \quad m \times 1$

$$= \sum_{e=(v_j,\,?)} \pm\sqrt{w(e)}\; g(e)$$

sum of values connected to v_j



"divergence"

"$M^T g = \nabla \cdot g$"

Putting together:

$$\nabla \cdot \nabla f = M^T M f = L \cdot f$$

$M^T \cdot \underbrace{Mf}$

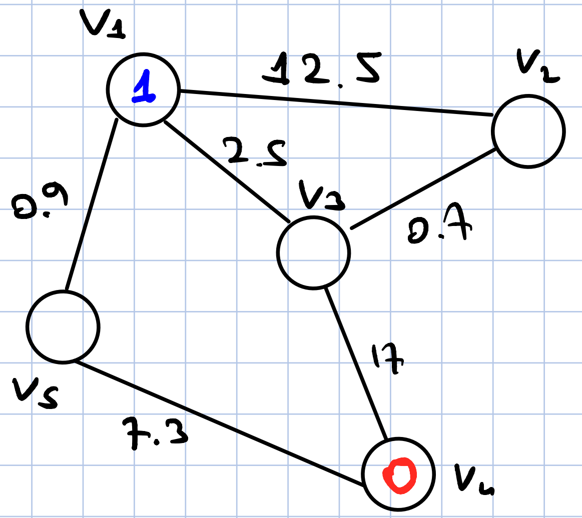
laplacian

CLASSIFICATION IN GRAPHS

INPUT: Weighted graph (G, W)

- some vertices are labeled as '0' or '1'

Ex.



GOAL: Decide on labels to <u>all</u> other vertices.

Key idea:

$e=(u,v) \rightarrow w(e)$ - high

$\downarrow$

u, v - "similar"

$\downarrow$

should have the same label.

FORMALLY:

V_L = labeled vertices $\subset V$

V_U = unlabeled vertices $\subset V$

$|V| = n$, $|V_L| = n_L$, $|V_U| = n_U$

Looking for: $f: V \rightarrow [0,1]$

Goal: Find f that minimises $\underbrace{\text{changes}}_{\|M \cdot f\|^2}$ along edges

OPTIMISATION PROBLEM:

known labels

$$f^* = \underset{f \in [0,1]^n}{\operatorname{argmin}} \{ \|M \cdot f\|^2 \text{ such that } f(v_i) = y_i, v_i \in V_L \}$$

Note: $f \in \mathbb{R}^n$ then:

$$f = f_L + f_U$$

$\neq 0$ on labeled

$\neq 0$ on unlabeled

Ex. $f = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \end{pmatrix} = \underbrace{\begin{pmatrix} f_1 \\ 0 \\ 0 \\ f_4 \\ 0 \end{pmatrix}}_{f_L} + \underbrace{\begin{pmatrix} 0 \\ f_2 \\ f_3 \\ 0 \\ f_5 \end{pmatrix}}_{f_U}$

We can write: $f_L = P_L^T P_L \cdot f$

$f_U = P_U^T P_U \cdot f$

where: P_L = projection on labeled

P_U = projection on unlabeled

Ex. $P_L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$ $P_U = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$

$P_L \cdot f = \begin{pmatrix} f_1 \\ f_4 \end{pmatrix}$ $P_U \cdot f = \begin{pmatrix} f_2 \\ f_3 \\ f_5 \end{pmatrix}$

Define: $y \in \{0,1\}^{n_L}$ - known labels

If f^* is a solution to ⊕

then: $P_L \cdot f^* = y$

$\Updownarrow$

$f^* = P_L^T \cdot y + P_U^T \cdot x$

$x = P_U \cdot f^*$, $x \in \mathbb{R}^{n_U}$

Equivalent opt. problem:

$$x^* = \underset{x \in [0,1]^{n_u}}{\arg\min} \left\{ \overbrace{\| \underbrace{M \cdot P_L^T y}_{\substack{b \\ \text{known vector}}} + \underbrace{M \cdot P_u^T}_{\substack{a \\ \text{known matrix}}} x \|^2}^{H(x) = \|ax+b\|^2} \right\}$$

Derivatives:

1-dimensional: $F(x) = (ax+b)^2$

$\downarrow$

$$F'(x) = 2 \cdot a \cdot (ax+b)$$

higher-dimensions: $F(x) = \| \underset{\text{matrix}}{a} \cdot \underset{\text{vector}}{x} + \underset{\text{vector}}{b} \|^2$

$\downarrow$

$$\nabla F(x) = 2 \cdot a^T (ax+b)$$

$$\Rightarrow \nabla H = \cancel{2} a^T (ax+b) = 0$$

$a = M \cdot P_u^T \qquad b = M P_L^T \cdot y$

$\Downarrow$

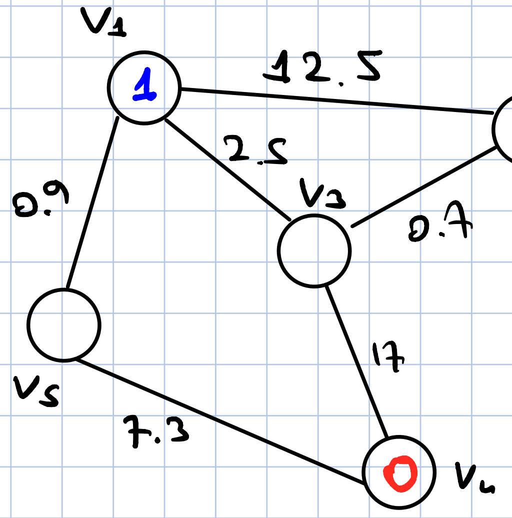
$$\underbrace{\underbrace{(M \cdot P_u^T)^T}_{a^T} \cdot \underbrace{(M \cdot P_u^T)}_{a}}_{A} \cdot x = -\underbrace{\underbrace{(M \cdot P_u^T)^T}_{a^T} \underbrace{(M P_L^T \cdot y)}_{b}}_{B}$$

$A = P_u (M^T M) P_u^T = P_u \cdot L \cdot P_u^T$ - matrix

$B = P_u \cdot L \cdot P_L^T \cdot y$ - vector

Solve: $Ax = -B$

Ex.



$$L = \begin{array}{c|ccccc|c} & L & U & U & L & U & \\ \hline & 15.9 & -12.5 & -2.5 & 0 & -0.9 & L \\ & -12.5 & 13.2 & -0.7 & 0 & 0 & U \\ & -2.5 & -0.7 & 20.2 & -17 & 0 & U \\ & 0 & 0 & -17 & 24.3 & -7.3 & L \\ & -0.9 & 0 & 0 & -7.3 & 8.2 & U \end{array}$$

$D_U L D_U^T$ = rows & columns of unlabeled vertices

$D_U L \cdot D_L^T$ = rows- unlabeled
columns- labeled

$$A = \begin{pmatrix} 13.2 & -0.7 & 0 \\ -0.7 & 20.2 & 0 \\ 0 & 0 & 8.2 \end{pmatrix}$$

$$B = \begin{pmatrix} -12.5 & 0 \\ -2.5 & -17 \\ -0.9 & -7.3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -12.5 \\ -2.5 \\ -0.9 \end{pmatrix}$$

Solve: $Ax = -B$

$$\Downarrow$$

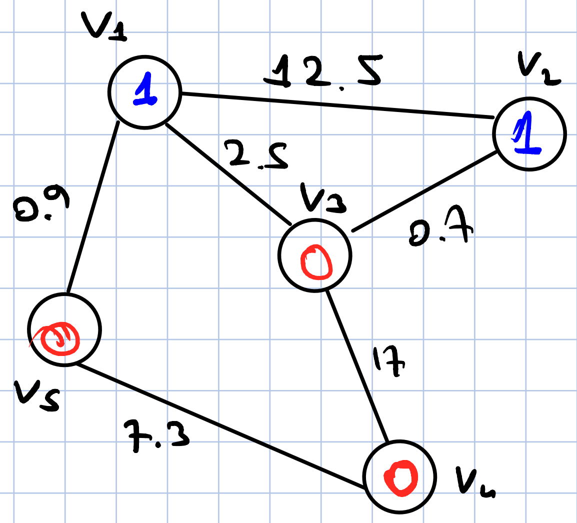
$$x^* = \begin{pmatrix} 0.955 \\ 0.156 \\ 0.109 \end{pmatrix}$$

$$\Downarrow$$

$$f^* = \begin{pmatrix} 1 \\ 0.955 \\ 0.156 \\ 0 \\ 0.109 \end{pmatrix}$$

$$\downarrow$$

labels: $\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$



SEMI-SUPERVISED BINARY CLASSIFICATION

$S = \{x_i\}_{i=1}^{n} \subseteq \mathbb{R}^d$ - data points

GOAL: classify as "0" or "1"

INPUT:
- The set S
- A small subset $R \subseteq S$ with labels

$R = \{x_{i_1}, x_{i_2}, \ldots, x_{i_r}\}$

$Y = \{y_1, y_2, \ldots, y_r\}$ - labels

0 OR 1

SOLUTION:

$G = (V, E)$

$V = S$.

$E = \{(x_i, x_j) : \|x_i - x_j\| \leq T\}$

chosen threshold

Weights: example

$w((x_i, x_j)) = e^{-\gamma \|x_i - x_j\|^2}$

$\Rightarrow$ used method shown earlier to classify all vertices in the graph.

SEMI-SUPERVISED → UNSUPERVISED

Semi-supervised:

- n points, $n_L = \#$ labels ($n_L << n$)
- Weighted graph (G, W)

Optimisation problem:

$$f^* = \underset{f \in [0,1]^n}{\operatorname{argmin}} \{ \|M \cdot f\|^2 \text{ such that } \cancel{f(v_i) = y_i}, v_i \in V_L \}$$

Unsupervised: no labels

Suggestion: try to solve the same way.

ISSUES:

- $f = 0$ is a (bad) solution
- $f = 1$ (all ones) is also a bad solution
- Given f suppose: $m = \|M \cdot f\|^2$

Take $f_\varepsilon = \varepsilon \cdot f$ then:

$$\|M \cdot f_\varepsilon\|^2 = \varepsilon^2 m$$

as small as we want

NEW OPTIMISATION PROBLEM:

$$\# \quad \underset{f \in \mathbb{R}^n}{\operatorname{argmin}} \frac{\|M \cdot f\|^2}{\|f\|^2} = \underset{f \perp \mathbb{1}}{\operatorname{argmin}} \boxed{\frac{f^T \cdot L \cdot f}{\|f\|^2}} \rightarrow \text{Rayleigh Quotient}$$

$0 \neq f \perp \underline{\mathbb{1}} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

NOTE:

(1) $f = \alpha \cdot \underline{1} \rightarrow$ not a solution

(2) $$\frac{\|M \cdot f_\varepsilon\|^2}{\|f_\varepsilon\|^2} = \frac{\cancel{\varepsilon^2}\|M \cdot f\|^2}{\cancel{\varepsilon^2}\|f\|^2} = \frac{\|Mf\|^2}{\|f\|^2}$$

(3) If f is an eigenvector of L, then:

$$(L \cdot f = \lambda \cdot f)$$

$$f^T \cdot L \cdot f = f^T \cdot (\lambda f) = \lambda \|f\|^2$$

$$\downarrow$$

$$\boxed{\frac{f^T L \cdot f}{\|f\|^2} = \lambda}$$

CLAIM: Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of L.

(1) $\lambda_1 = 0 \quad f_1 = \underline{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

(2) $\lambda_2 = \underset{f \perp \mathbb{1}}{\min} \frac{f^T L f}{\|f\|^2}$, f_2 = eignenvec

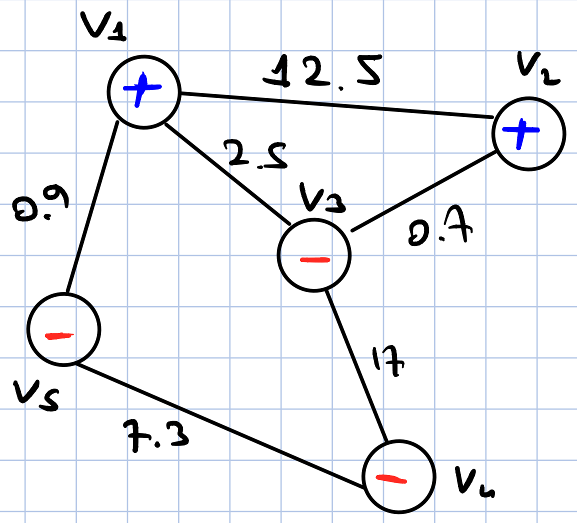
(3) $\lambda_2 > 0 \iff G$ is connected.

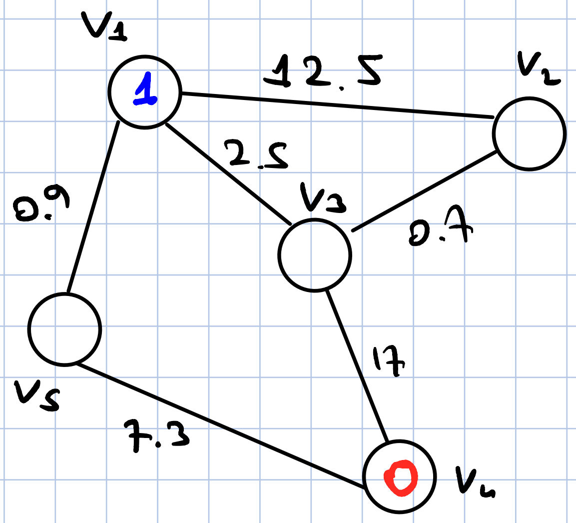
CONCLUSION:

We can solve (#) by eigenvector $\underline{\underline{f_2}}$ that corresponds to $\underline{\underline{\lambda_2}}$

Ex.

$$f_2 = \begin{pmatrix} 0.493 \\ 0.591 \\ -0.271 \\ -0.37 \\ -0.442 \end{pmatrix} \downarrow \begin{pmatrix} + \\ + \\ - \\ - \\ - \end{pmatrix}$$





$$L = \begin{pmatrix} 15.9 & -12.5 & -2.5 & 0 & -0.9 \\ -12.5 & 13.2 & -0.7 & 0 & 0 \\ -2.5 & -0.7 & 20.2 & -17 & 0 \\ 0 & 0 & -17 & 24.3 & -7.3 \\ -0.9 & 0 & 0 & -7.3 & 8.2 \end{pmatrix}$$