

# Aprendizado Semi-Supervisionado

1001513 – Aprendizado de Máquina 2  
Turma A – 2023/2  
Prof. Murilo Naldi



[naldi@ufscar.br](mailto:naldi@ufscar.br)



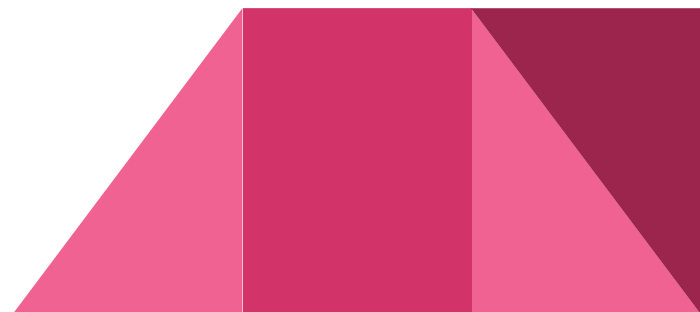
# Agradecimentos

- Pessoas que colaboraram com a produção deste material: Diego Silva, Ricardo Campello, Ricardo Cerri, Moacir Ponti
- Intel IA Academy

# Uma historinha pra começar a aula

Um professor e seu aluno quiseram fazer um *dataset*

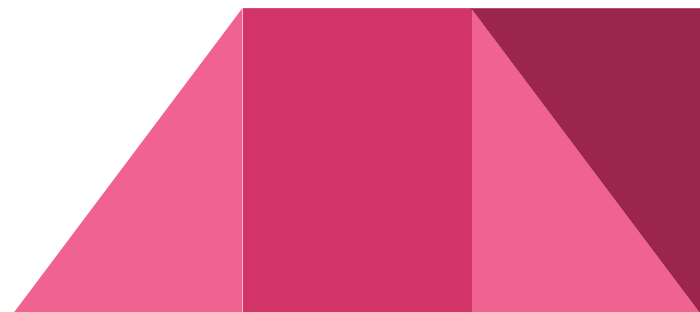
- ToLD-BR (<https://arxiv.org/abs/2010.04543>)
- Coletaram 10 milhões de *tweets*



# Uma historinha pra começar a aula

Um professor e seu aluno quiseram fazer um *dataset*

- ToLD-BR (<https://arxiv.org/abs/2010.04543>)
- Coletaram 10 milhões de *tweets*
- Como rotular isso?



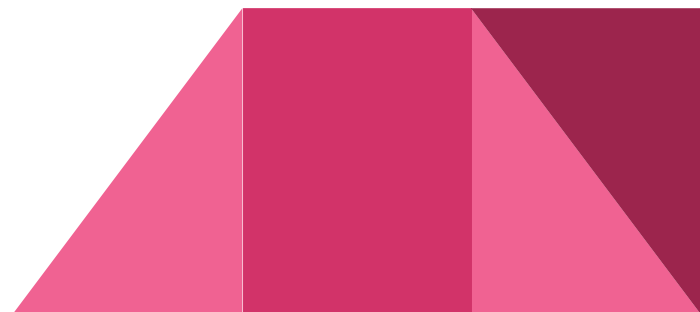
Uma historinha pra começar a aula



# Uma historinha pra começar a aula

Um professor e seu aluno quiseram fazer um *dataset*

- ToLD-BR (<https://arxiv.org/abs/2010.04543>)
- Coletaram 10 milhões de *tweets*
- Conseguiram 21 mil exemplos rotulados
  - Mas e se...



# Aprendizado semi-supervisionado

Grande volume de dados + poucos deles rotulados

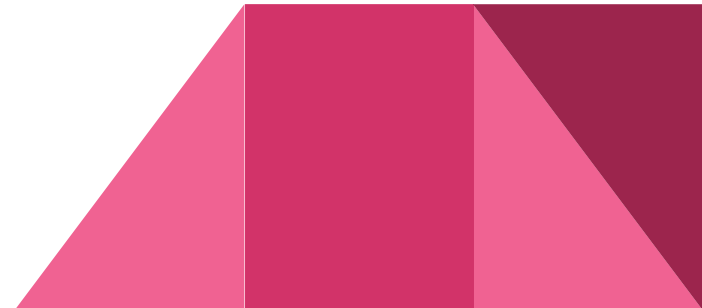
- Na verdade pode nem ser tão grande
- Pode ser classificação ou agrupamento
  - Ambos podem usar algumas informações sobre os rótulos
  - E o modelo resultante pode ser associado à criação de modelos de classificação



# Aprendizado semi-supervisionado

Aprendizado semi-supervisionado é a parte de aprendizado de máquina que combina inferência de rótulos (aprendizado supervisionado) a partir da forma em que os dados são estruturados (aprendizado não-supervisionado)

- Mistura de um pouco dos dois



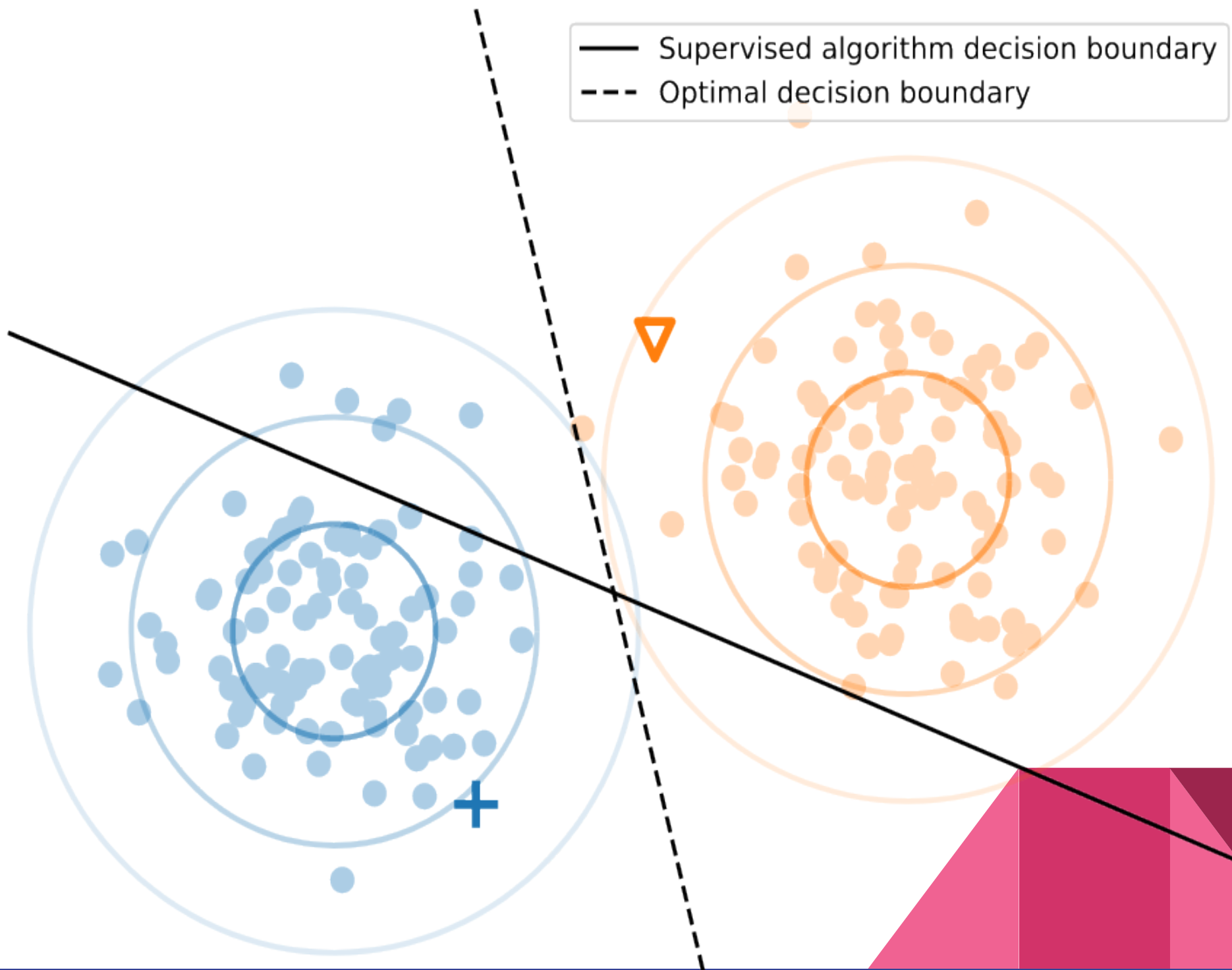


# Aprendizado semi-supervisionado

A maior parte dos trabalhos é focada em classificação semi-supervisionada

- Onde dados sem rótulos são usados para melhorar o resultado de um classificador
  - Melhoram a percepção do fronteira de decisão

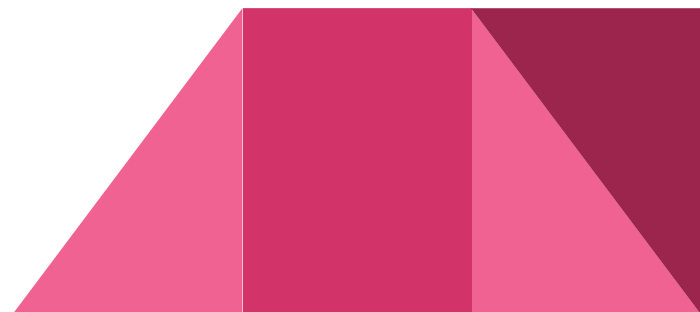




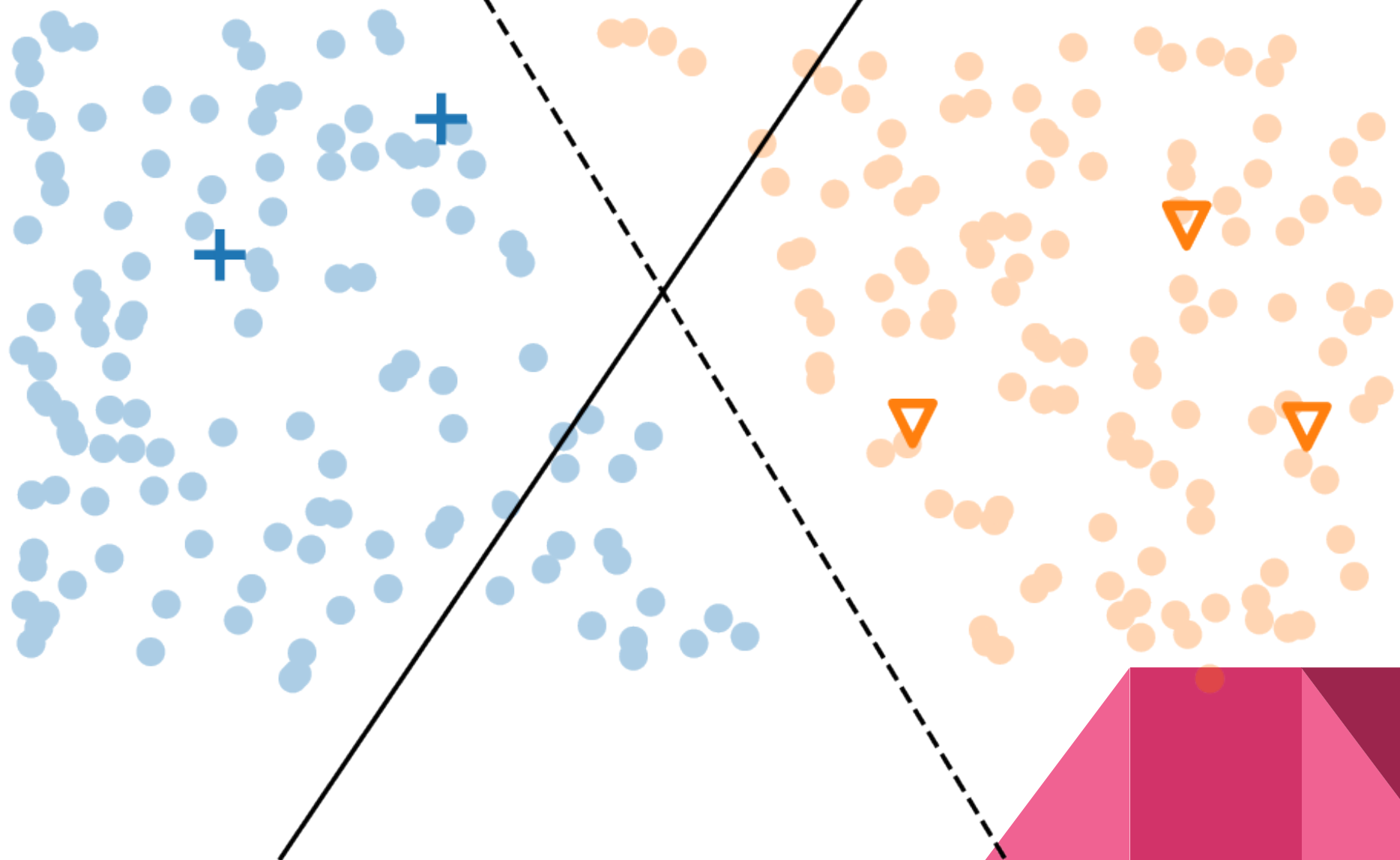
# Suposições

Algumas suposições são importantes :

- Suposição de suavidade: um objeto próximo de um objeto rotulado tende a possuir o mesmo rótulo
- Suposição de baixa densidade: a fronteira de decisão deve passar por uma região de baixa densidade de dados



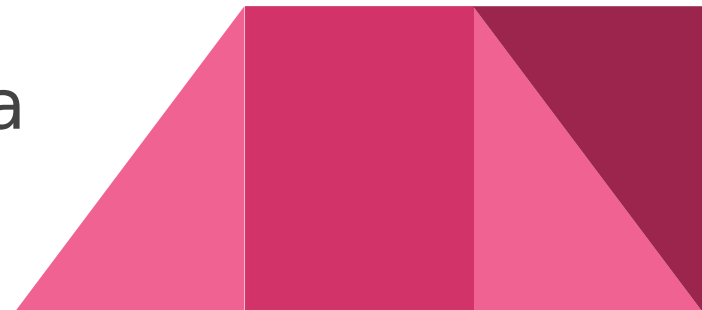
— Supervised algorithm decision boundary  
- - - Optimal decision boundary

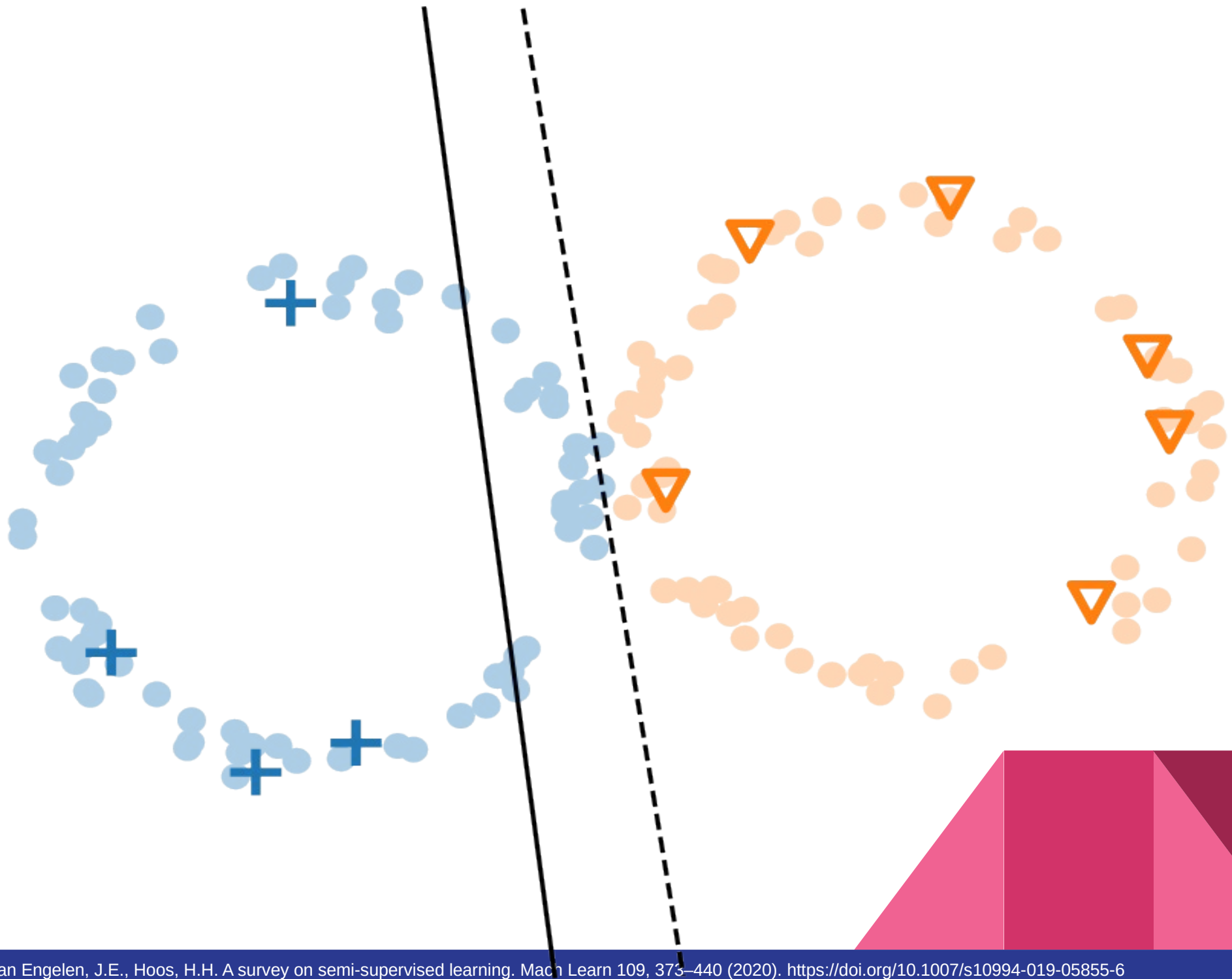


# Suposições

Algumas suposições são importantes :

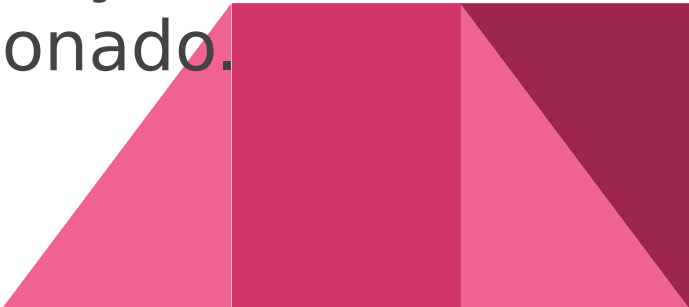
- Suposição de multiplicidade (*manifold*): as classes estão estruturadas em espaços topológico menores (*manifolds*) que se parece localmente com um espaço euclidiano nas vizinhanças de cada ponto
  - Exemplo: duas esferas podem ser divididas em uma variedade de pequenos círculos
  - Objetos que estão no mesmo *manifold* devem ser da mesma classe

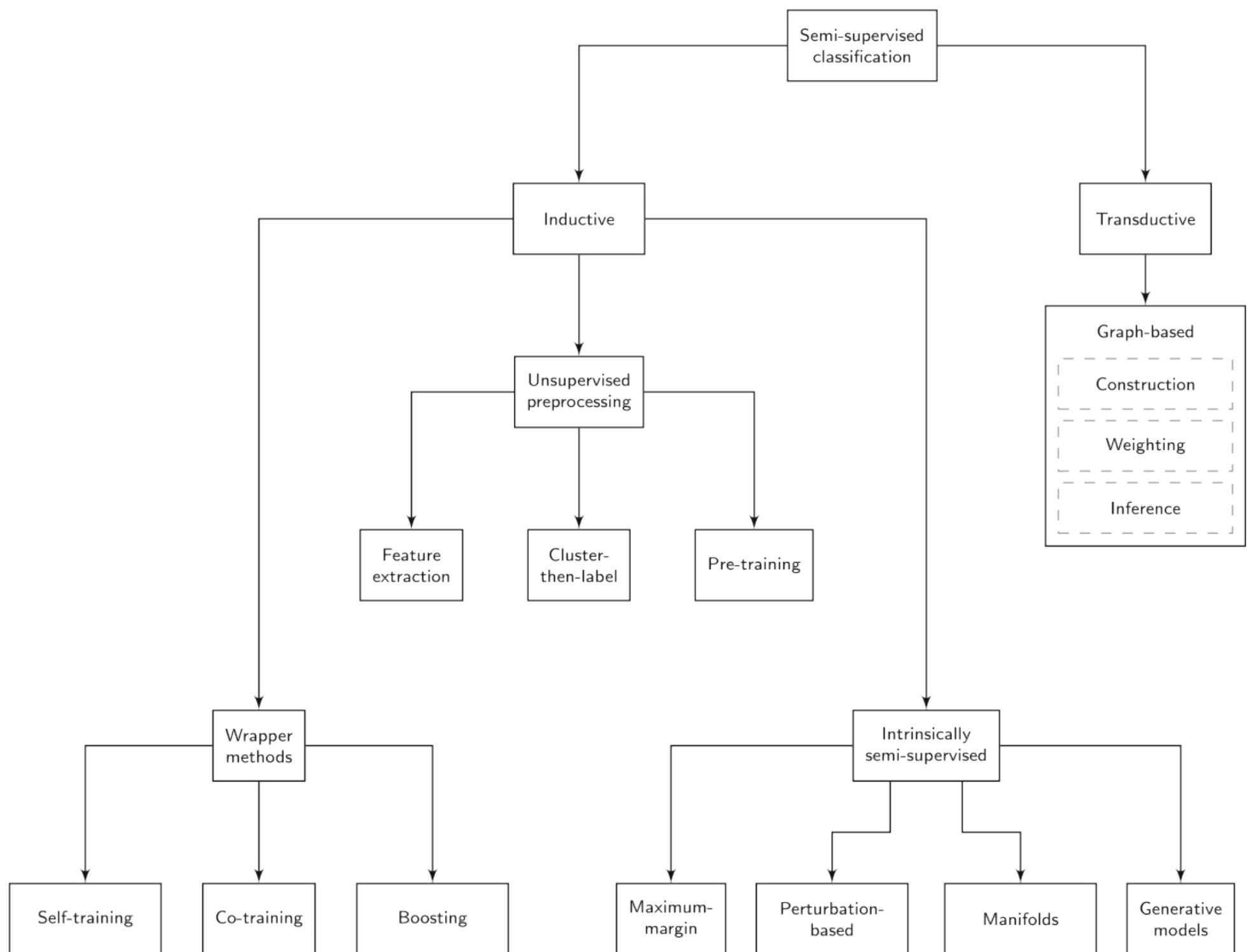




# Conexão com agrupamento

As suposições anteriores podem ser generalizadas como a “*suposição de agrupamento*”, ou seja, que os dados e suas classes se organizam como grupos:

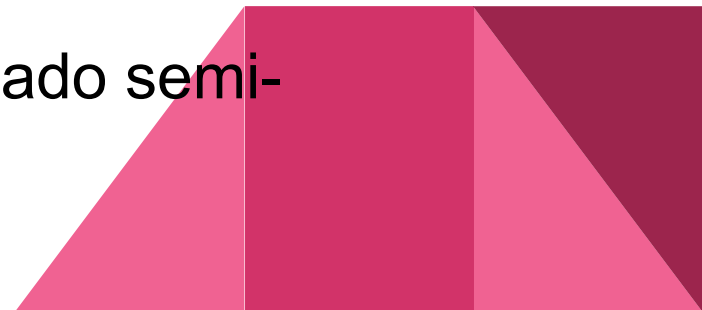
- Conceito de grupo por similaridade
  - Se os dados (não rotulados e rotulados) não puderem ser agrupados, não é possível que um método de aprendizado semi-supervisionado possa melhorar o resultado em relação a um método de aprendizado supervisionado.
- 





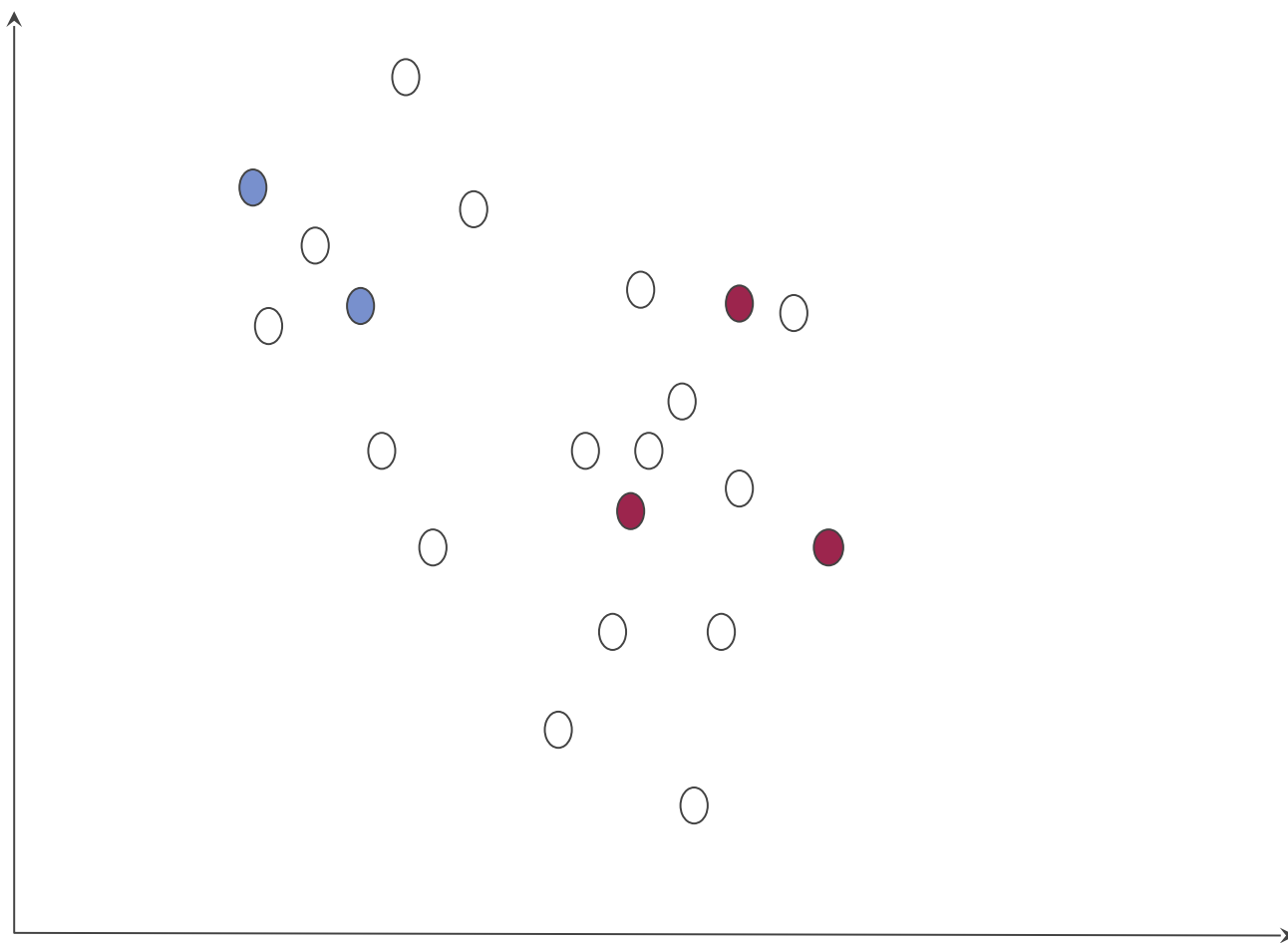
# Aprendizado semi-supervisionado indutivo

- Os métodos indutivos visam construir um classificador que possa gerar previsões para qualquer objeto
- Dados não rotulados podem ser usados ao treinar este classificador
  - Contudo, previsões para exemplos novos são independentes
- Alguns exemplos de como o modelo pode ser treinado são:
  - *Wrapper*
  - Aprendizado não supervisionado
  - Funções objetivo intrínsecas ao aprendizado semi-supervisionado



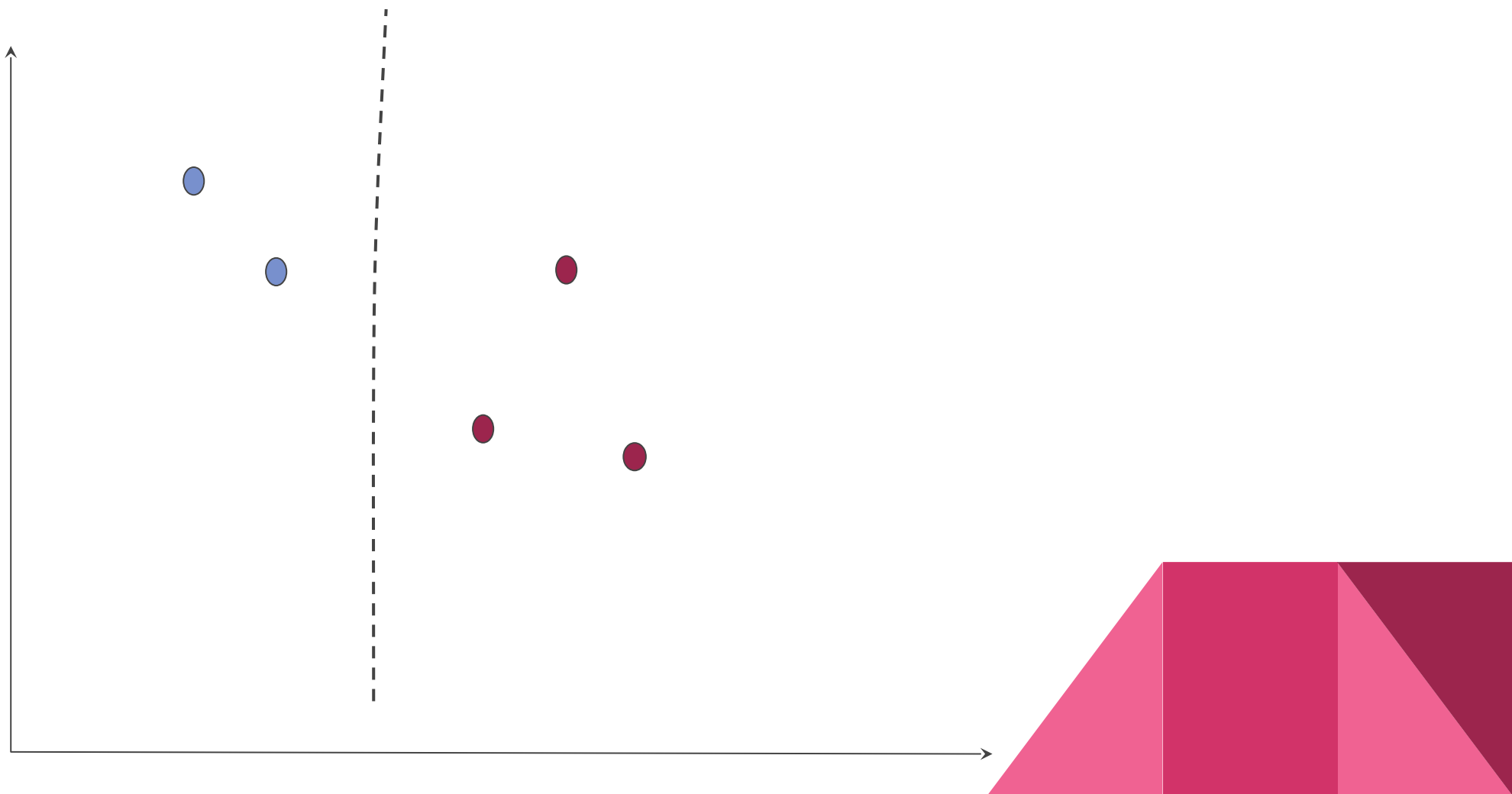
# Aprendizado semi-supervisionado

Aprendizado (semi-supervisionado) indutivo



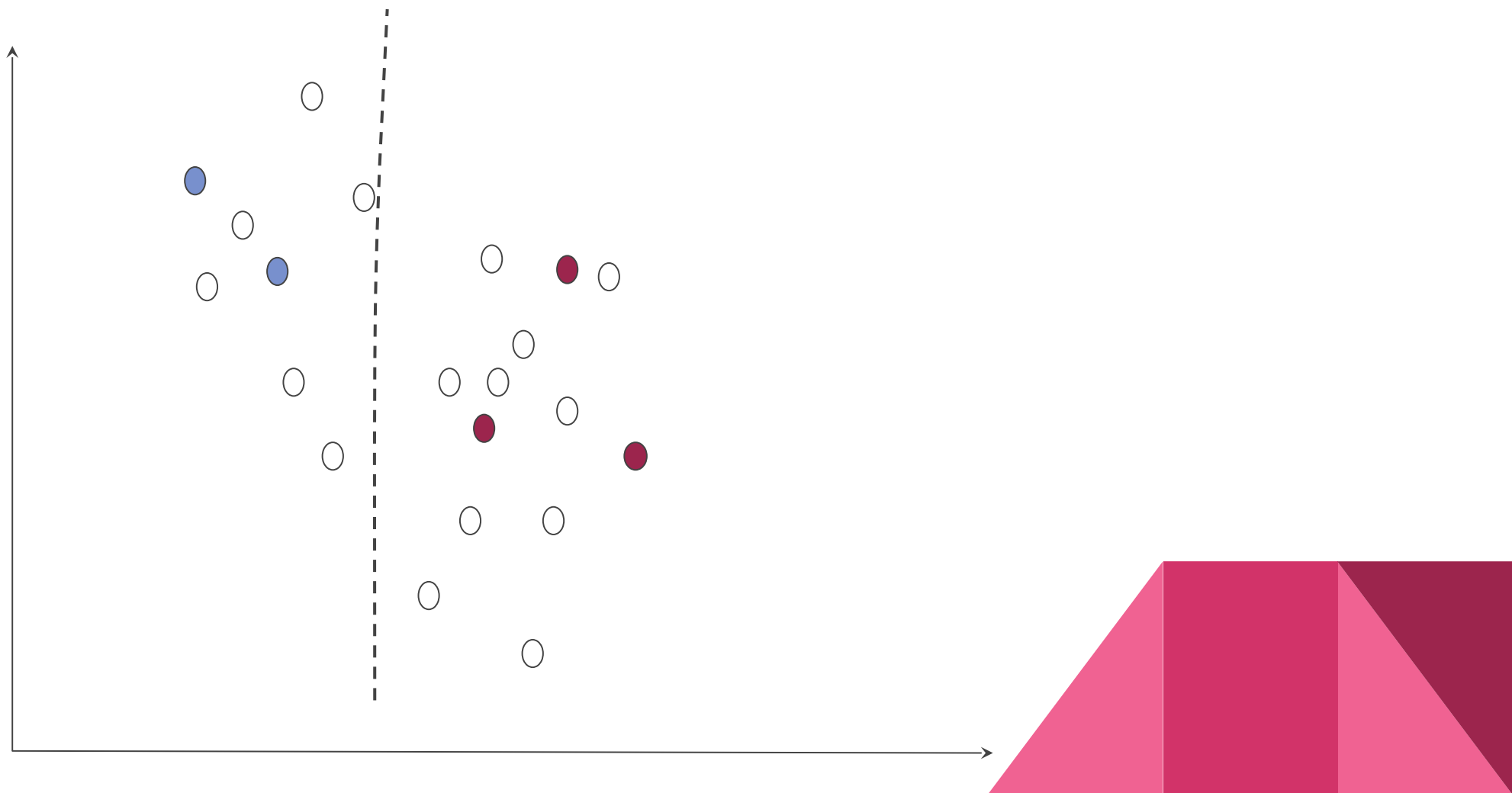
# Aprendizado semi-supervisionado

Aprendizado (semi-supervisionado) indutivo



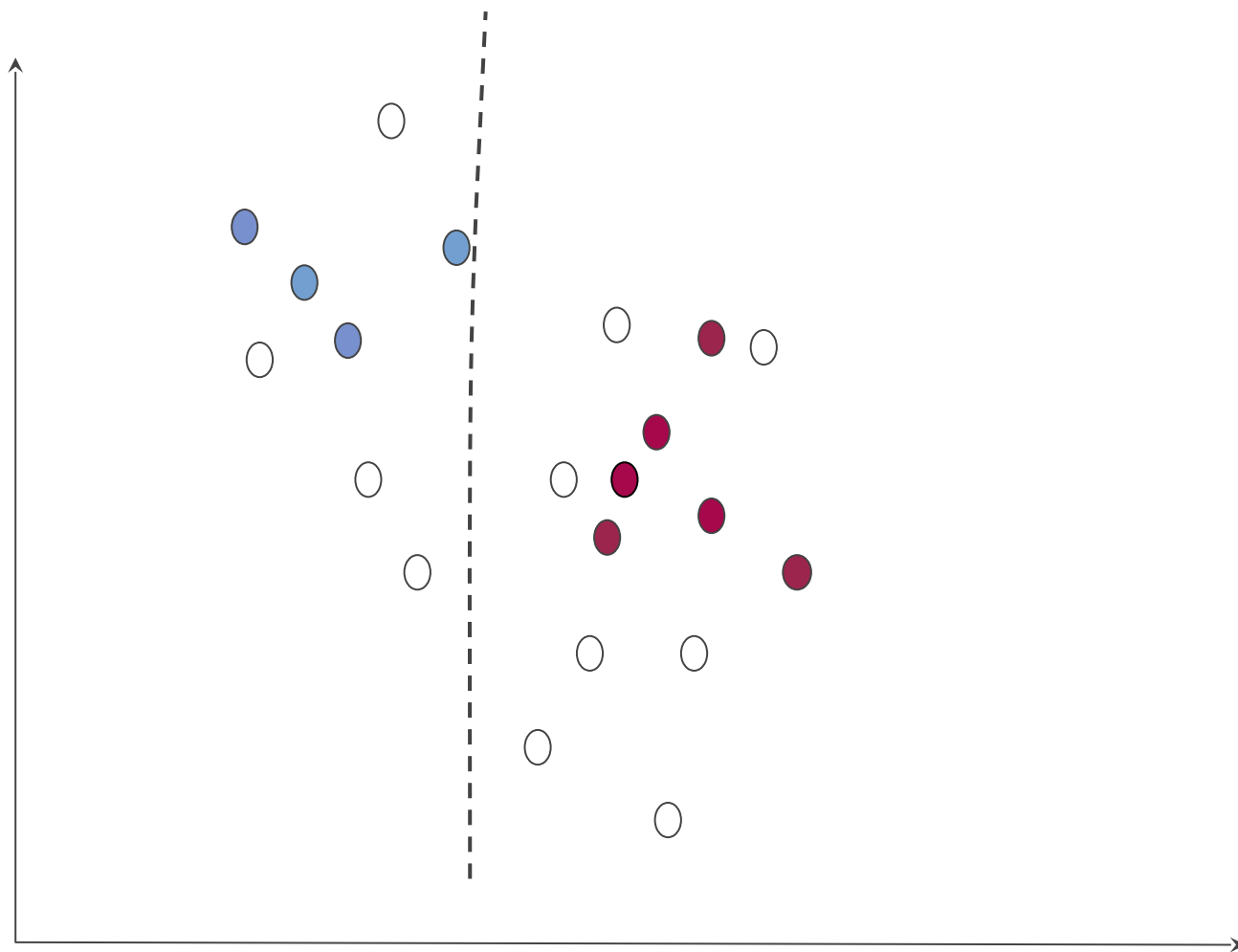
# Aprendizado semi-supervisionado

## Aprendizado (semi-supervisionado) indutivo



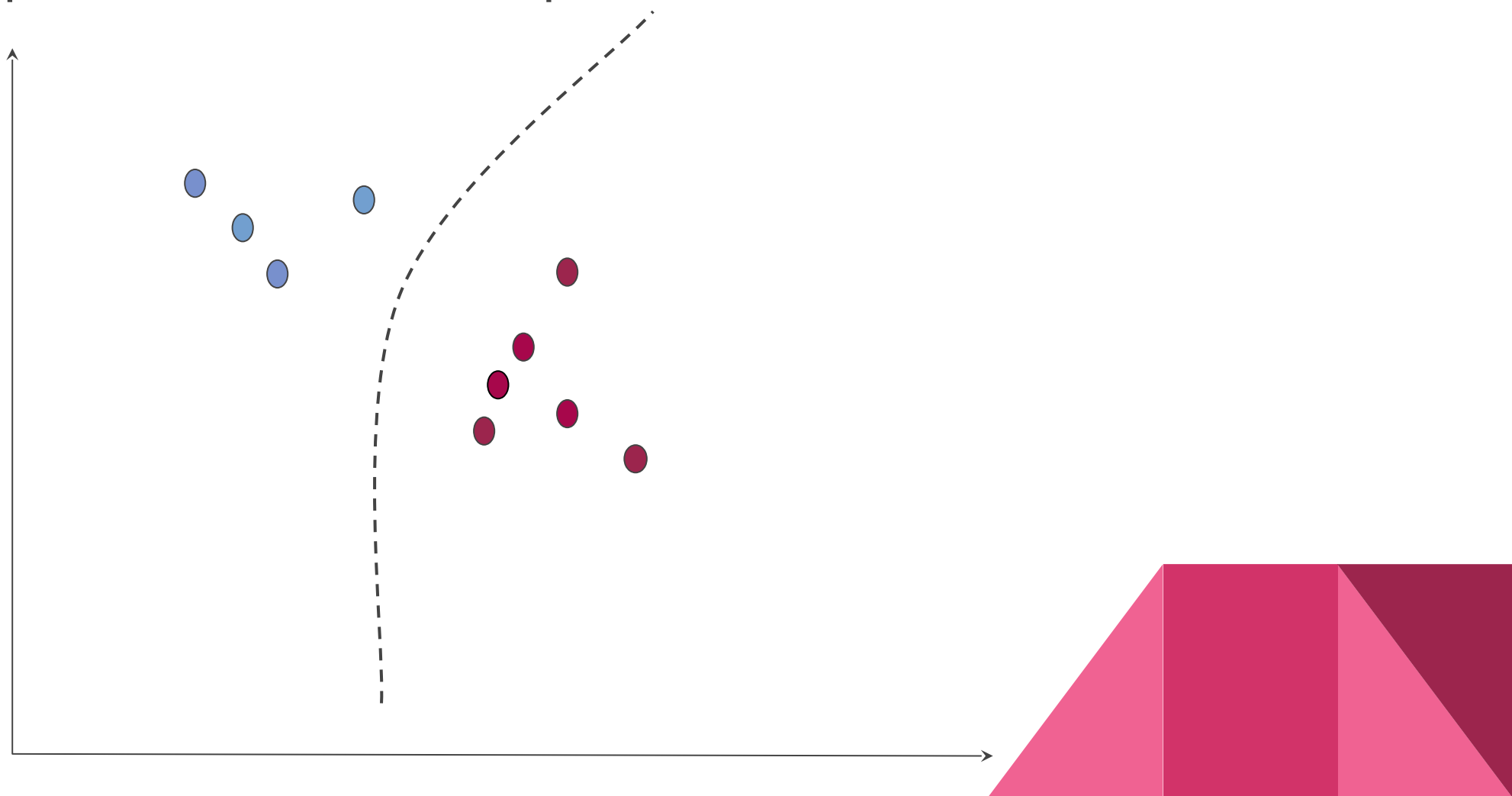
# Aprendizado semi-supervisionado

Aprendizado (semi-supervisionado) indutivo



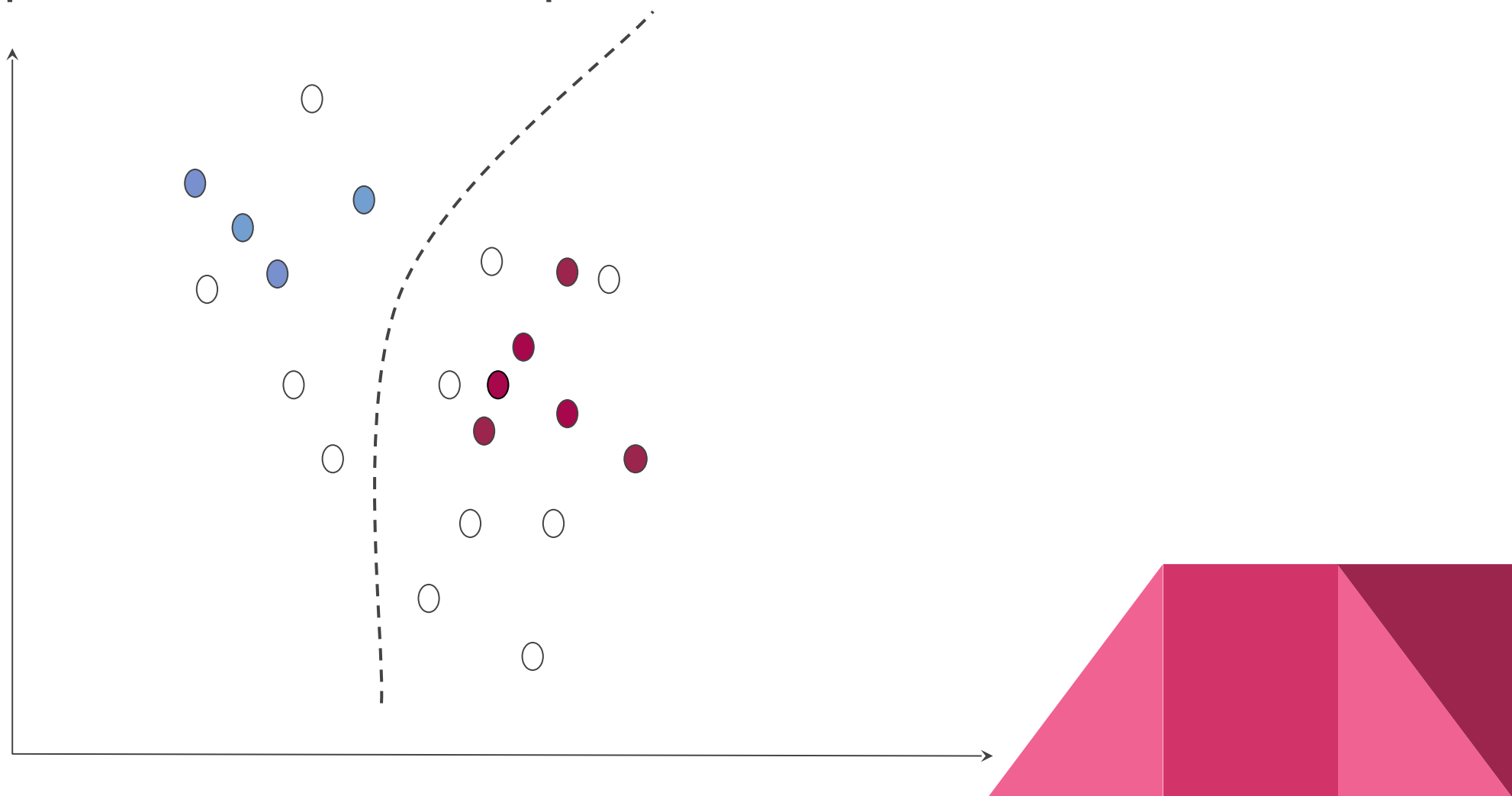
# Aprendizado semi-supervisionado

Aprendizado (semi-supervisionado) indutivo



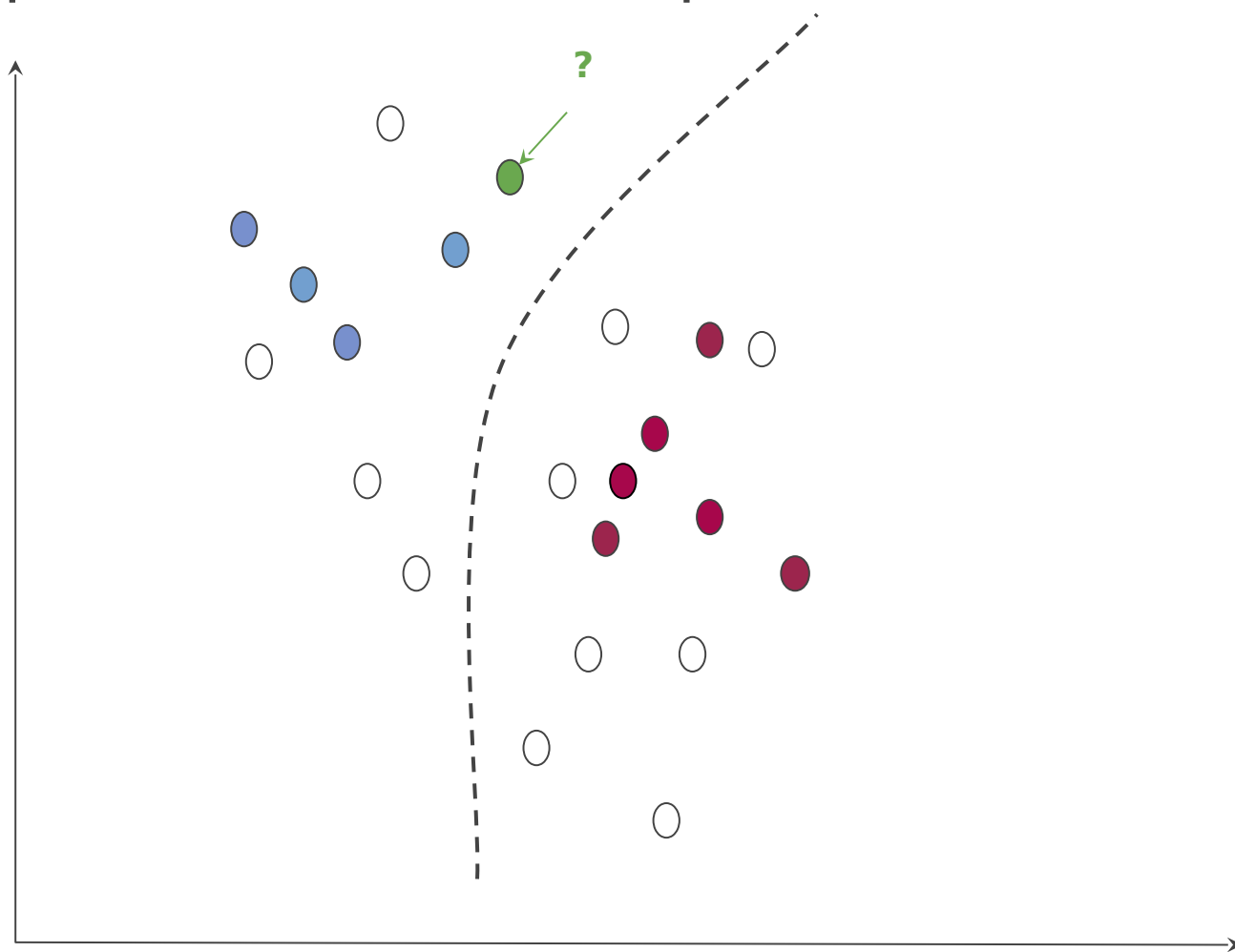
# Aprendizado semi-supervisionado

Aprendizado (semi-supervisionado) indutivo



# Aprendizado semi-supervisionado

Aprendizado (semi-supervisionado) indutivo





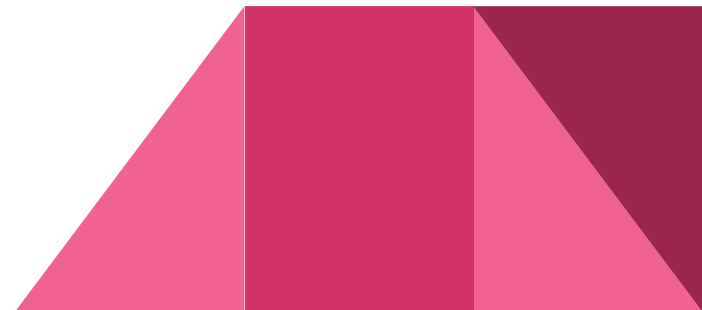
# Aprendizado semi-supervisionado indutivo

- *Wrapper*
  - Um dos métodos mais clássicos para aprendizado semi-supervisionado
  - Utilizam um ou mais classificadores básicos supervisionados e os treina iterativamente com os dados rotulados originais
    - Somados como com dados não rotulados adicionados por meio de previsões anteriores
  - Comumente chamados de dados pseudo-rotulados
  - Duas etapas:
    - Treinamento
    - Pseudo-rotulagem
  - Somente pontos com alta confiança são rotulados!



# Aprendizado semi-supervisionado indutivo

- Aprendizado não supervisionado
  - Diferentemente dos métodos *wrapper*, os dados não rotulados e rotulados são usados em dois estágios separados
  - Normalmente, o estágio não supervisionado compreende a extração ou a transformação de características da amostra não rotulada
    - Para enriquecer o espaço de características
  - Ou agrupamento
  - Ou para a inicialização dos parâmetros do procedimento de aprendizagem (pré-treinamento)
    - Bastante comum em redes neurais



# Aprendizado semi-supervisionado indutivo

- Métodos intrínsecos
  - Otimizam diretamente uma função objetivo para amostras rotuladas e não rotuladas
  - Não dependem de nenhuma etapa intermediária ou base supervisionada
  - Geralmente dependem de uma mais das suposições apresentadas anteriormente
    - Por exemplo, os métodos de margem máxima baseiam-se na suposição de baixa densidade,
    - A maioria das redes neurais semi-supervisionadas dependem da suposição de suavidade



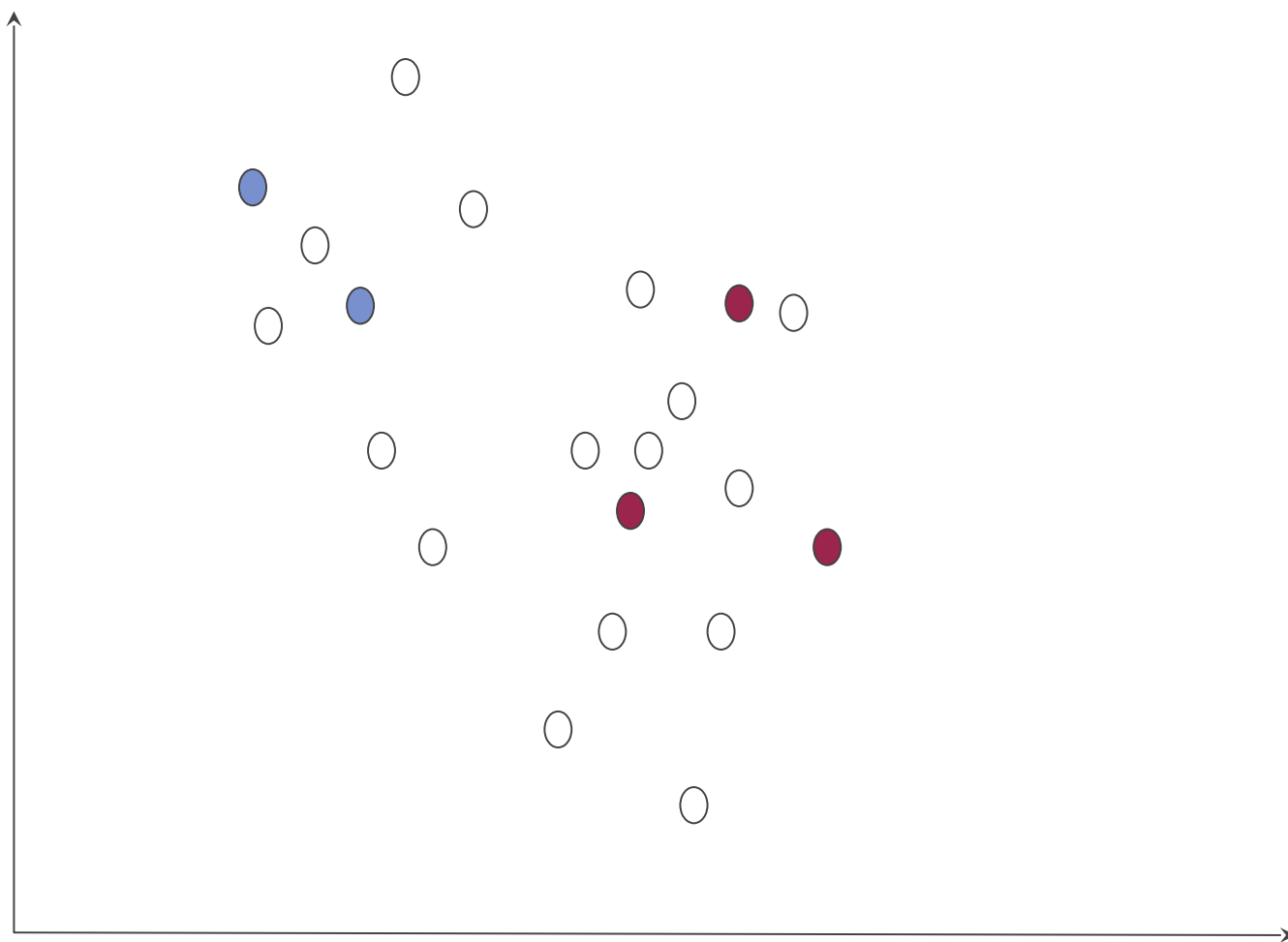
# Aprendizado semi-supervisionado transdutivo

- Ao contrário dos algoritmos indutivos, os algoritmos transdutivos não produzem um preditor
  - Em vez disso, produzem um conjunto de previsões para o conjunto de pontos de dados não rotulados fornecidos
  - Portanto, não podemos distinguir entre fases de treinamento e teste
- Eles recebem dados rotulados e não rotulados objetivando fornecer os rótulos dos não rotulados, quando possível



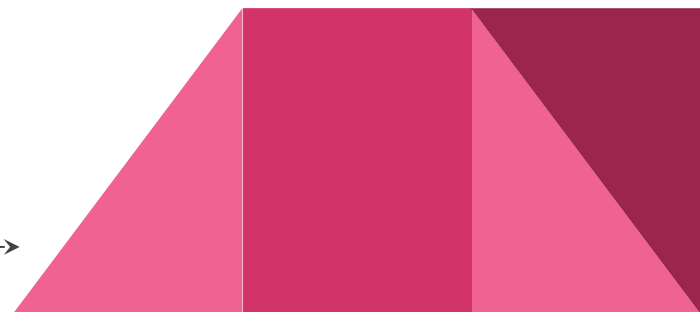
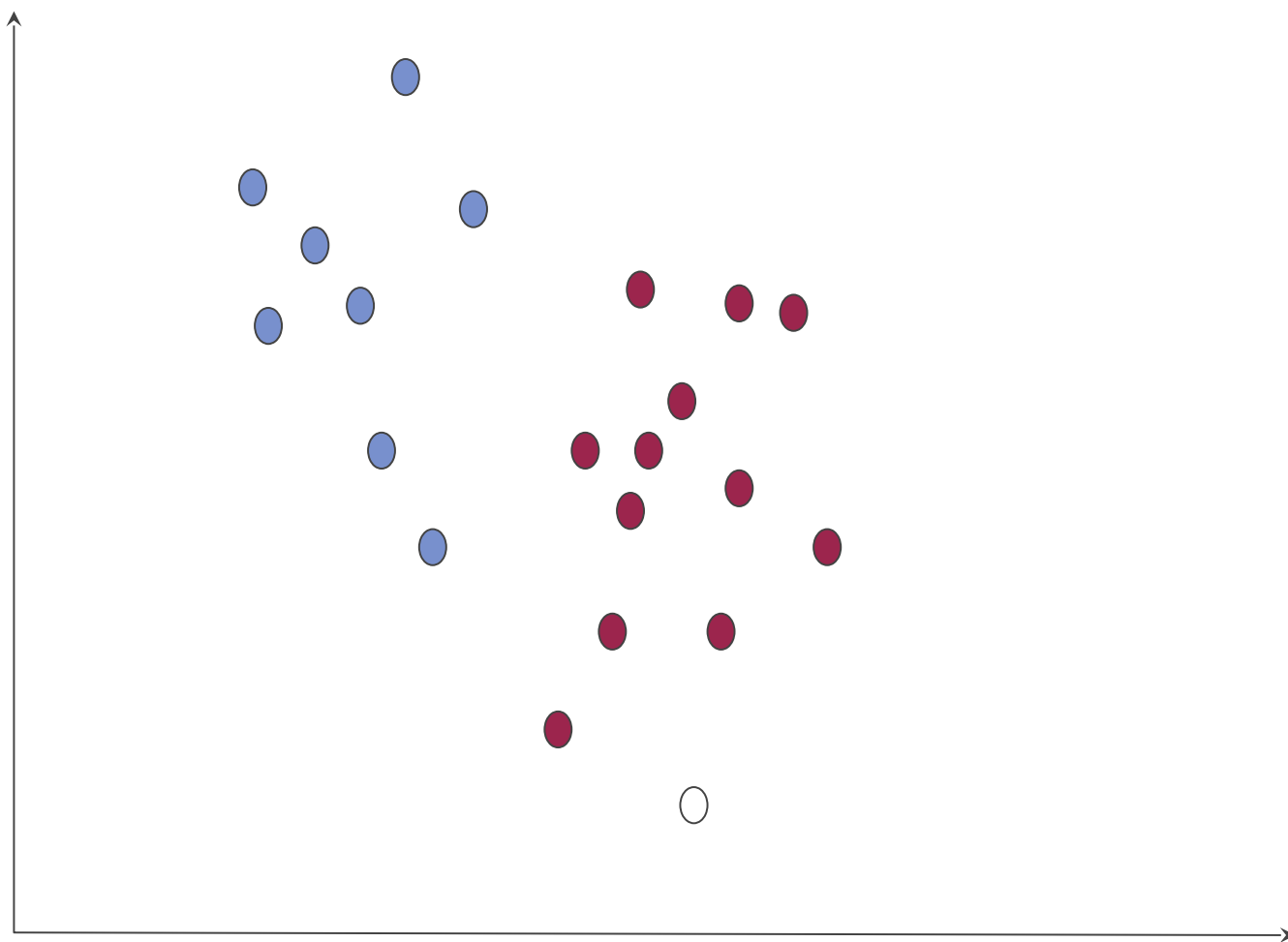
# Aprendizado semi-supervisionado

Aprendizado (semi-supervisionado) transdutivo



# Aprendizado semi-supervisionado

Aprendizado (semi-supervisionado) transdutivo



# Aprendizado semi-supervisionado transdutivo

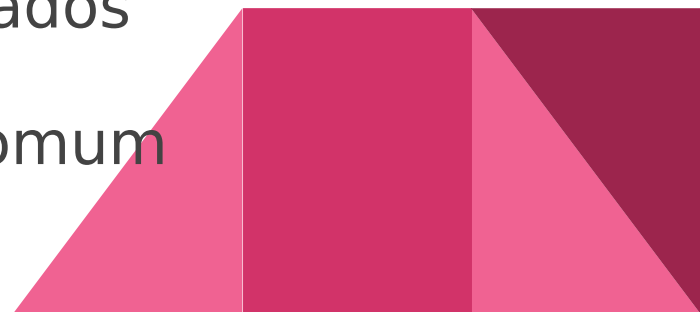
- Métodos transdutivos normalmente definem um gráfico sobre todos os dados,
  - Codificando a semelhança entre pares de pontos de dados com arestas possivelmente ponderadas
- Uma função objetivo é então definida e otimizada a fim de:
  - Para pontos de dados rotulados, os rótulos previstos devem corresponder aos rótulos verdadeiros.
  - Pontos de dados semelhantes, conforme definidos através do gráfico de similaridade, devem ter o mesmo rótulo.



# Label propagation

Dado um grafo,

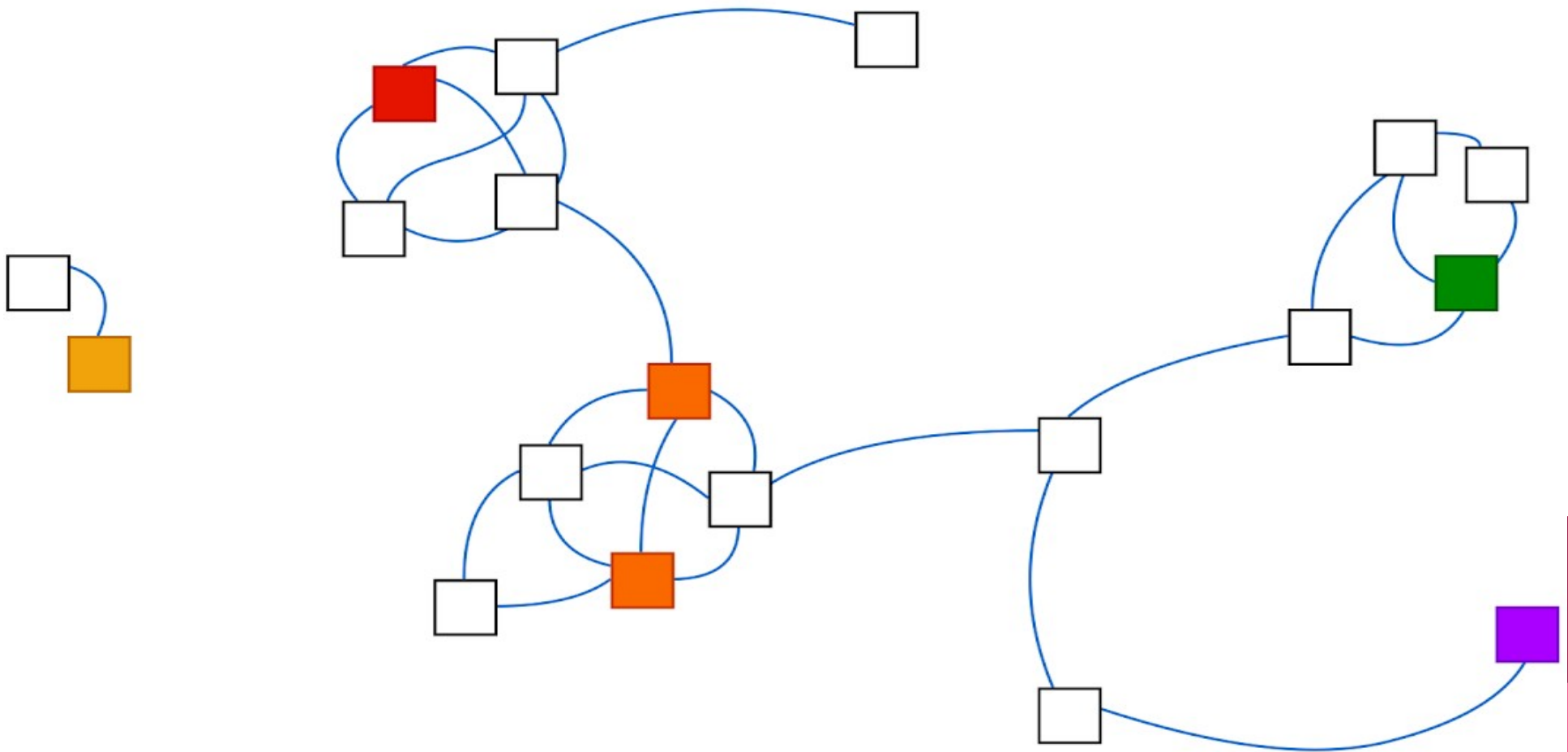
- 1) Cada nó tem seu rótulo correspondente
- 2) O rótulo denota a comunidade à qual esse nó pertence
- 3) Através da iteração, cada nó atualizará seu rótulo com base nos rótulos dos nós vizinhos
  - 1) O rótulo atualizado de cada nó será o mais presente dentre os vizinhos do nó
- 4) Eventualmente, nós densamente conectados alcançam uma comunidade de rótulos comum



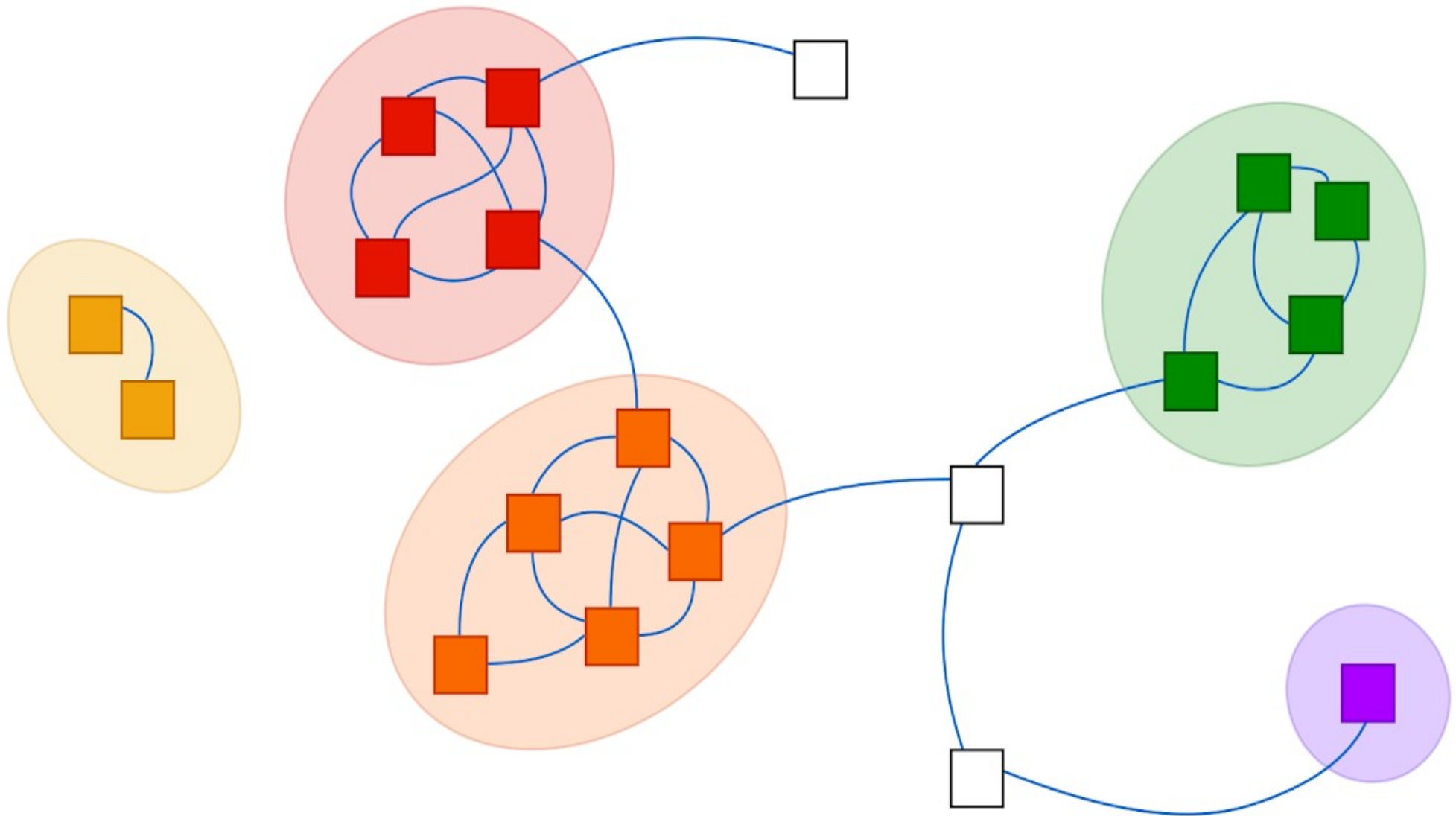


# Label propagation

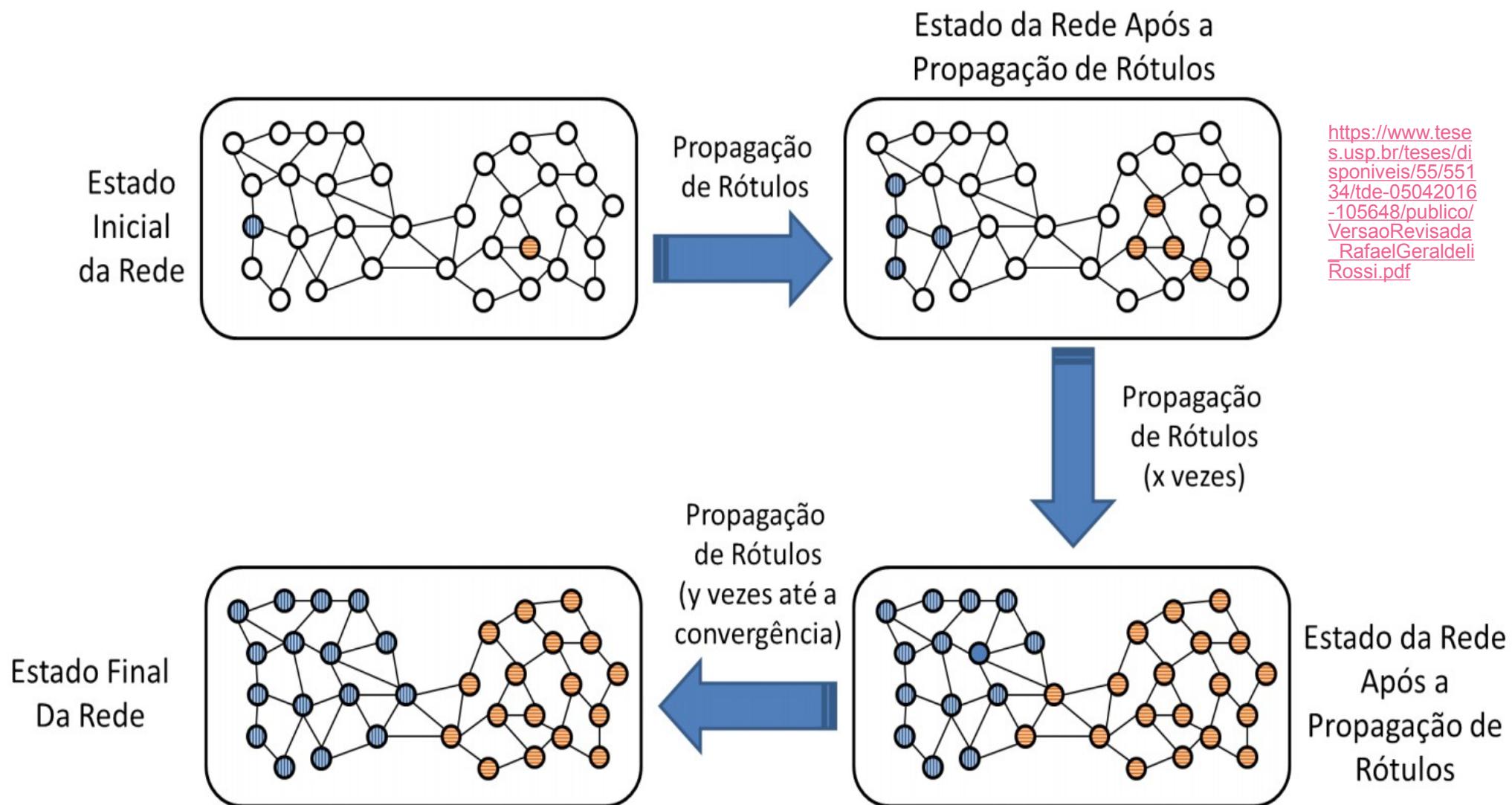
Vamos começar pelo conceito geral



# Label propagation



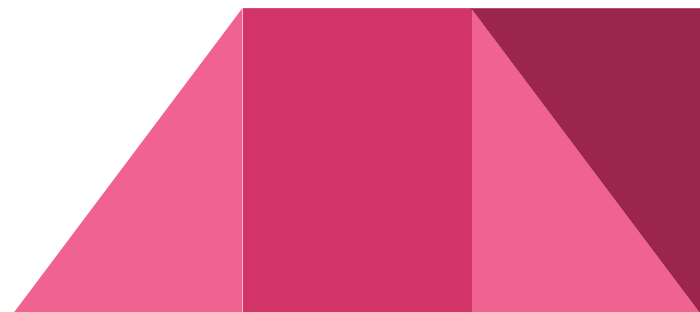
# Label propagation



# Label propagation

Construção do grafo – Gaussiana ou RBF sobre grafo completo

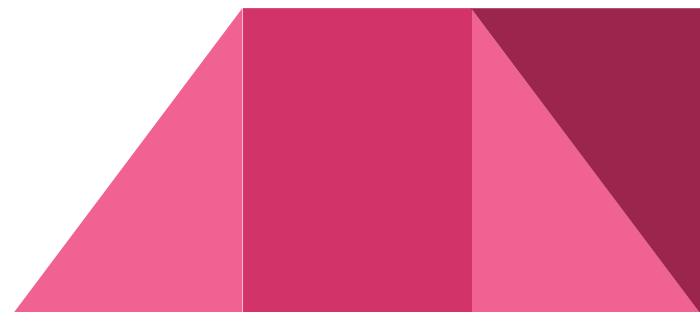
- Faz-se o grafo completo
- Ajusta-se Funções Gaussianas ou RBF nos dados
  - Arestas são ponderadas de acordo com a distribuição
- Aplica-se label propagation



# Label propagation

## Construção do grafo – $k$ NNG

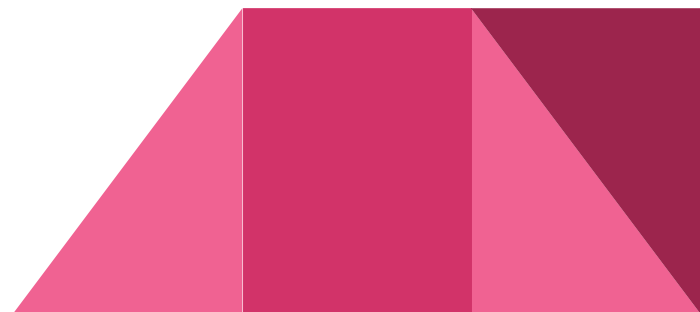
- Dado um valor de  $k$ , conecta-se os  $k$ -vizinhos mais próximos
- Depois label propagation
- Alternativamente,  $\epsilon$ NNG, onde  $\epsilon$  define uma dissimilaridade limite para conectar vizinhos (constante)



# Label propagation

## Construção do grafo – $k$ MNNG (mútuo)

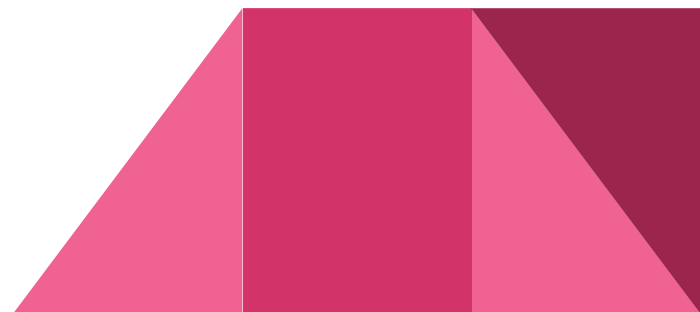
- Mesma ideia do  $k$ NNG, porém as conexões só ocorrem entre  $k$  vizinhos mais próximos que seja mútuos
  - Tende a gera menos *hubs*
- Alternativamente, também possui versão  $\epsilon$ MNNG



# Label propagation

## Algoritmo

- $l$  e  $u$  são o número de exemplos rotulados e não rotulados
- $Y$  é uma matriz  $(l+u) \times C$  com a distribuição de probabilidade dos labels



# Label propagation

## Algoritmo

- Definimos  $T$ , uma matriz de prob. de transição  $(l+u) \times (l+u)$

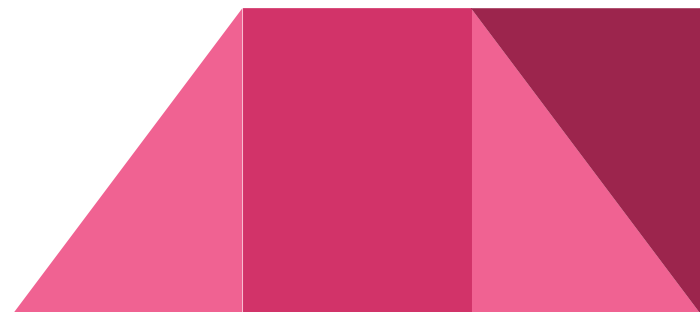




# Label propagation

## Algoritmo

1. Propagamos os rótulos:  $Y \leftarrow TY$
2. Normalizamos  $Y$  (por linha)
3. Asseguramos o rótulo dos inicialmente rotulados



# Label propagation

## Detalhes

- Convergência
- Parâmetro  $\sigma$  (RBF)
- Rebalanceamento das classes
- Zhu, Xiaojin & Ghahramani, Zoubin. (2003). Learning from Labeled and Unlabeled Data with Label Propagation.

<http://pages.cs.wisc.edu/~jerryzhu/pub/CMU-CALD-02-107.pdf>

- Variações: GFHF e LLGC
- 

# Agrupamento Semi-supervisionado

A forma que os rótulos são aplicados no agrupamento é diferente da forma da tarefa de classificação

- Na classificação os rótulos são usados para definir rótulos dos objetos não rotulados por transdução e melhorar a indução do modelo
- No agrupamento os rótulos servem para definir “o grupo” do objeto e só faz sentido se houver dois ou mais  
(porque?)



# Agrupamento Semi-supervisionado

Um objeto rotulado “define” o rótulo do grupo, portanto:

- Um grupo não deve ter dois ou mais objetos com rótulos distintos
- Um grupo deve possuir todos os objetos que possuem o mesmo rótulo




# Agrupamento Semi-supervisionado

Em outras palavras:

- Os rótulos servem para definir relações *must-link* e *cannot-link* no processo de agrupamento
- Os algoritmos devem ser adaptados para respeitar essas restrições durante a construção do modelo



## Exemplo: $k$ -médias

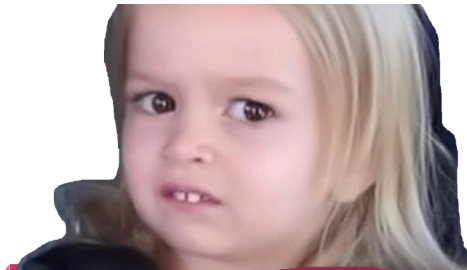
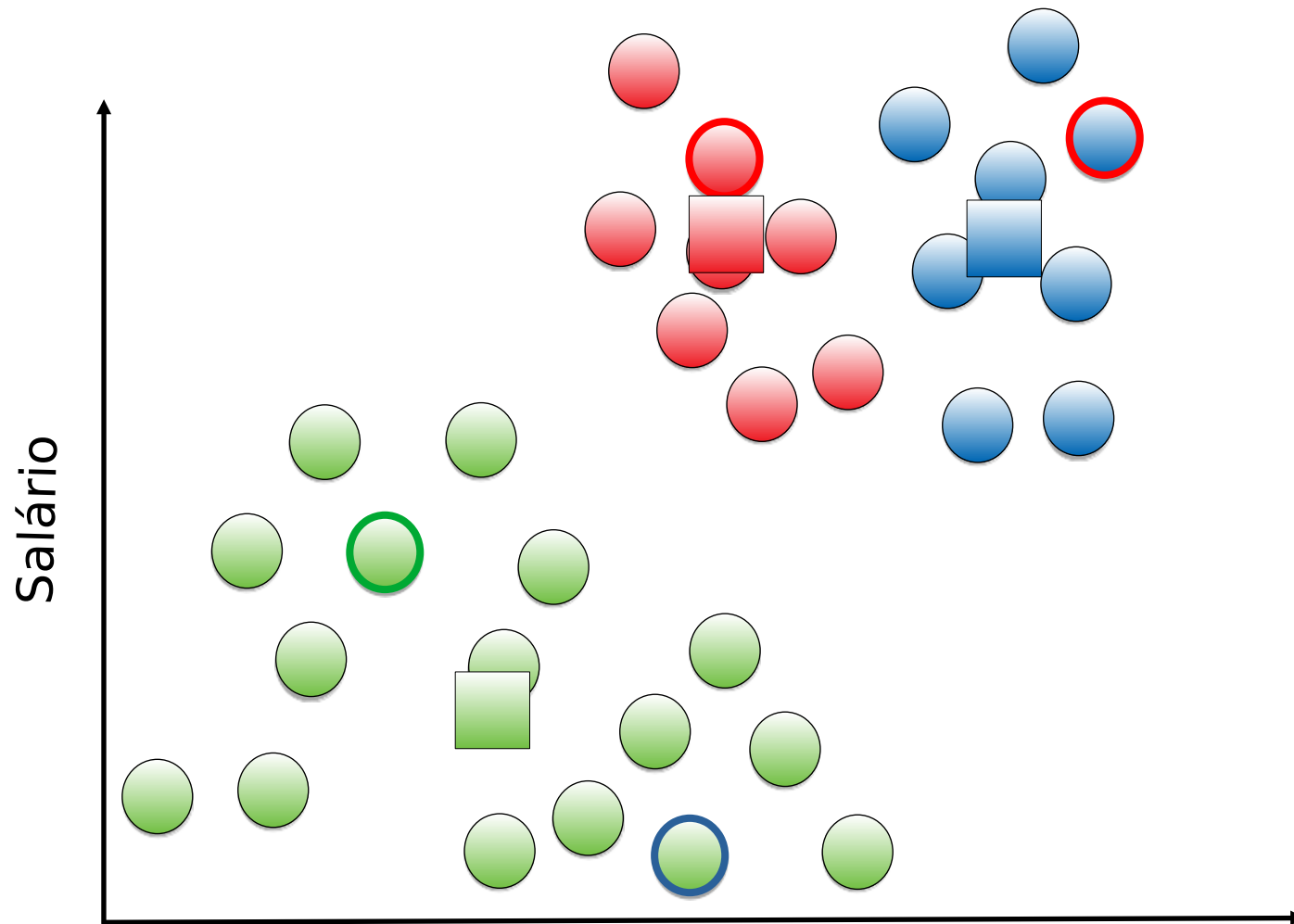
- 1) Escolher um número  $k$  de protótipos (centros) para os grupos
  - 2) Atribuir cada objeto para o grupo de centro mais próximo (segundo alguma distância, e.g. Euclidiana)
  - 3) Mover cada centro para a média (centróide) dos objetos do grupo correspondente
  - 4) Repetir os passos 2 e 3 até que algum critério de convergência seja obtido
- 

# Exemplo: $k$ -médias com restrições

- 1) Escolher um número  $k$  de protótipos (centros) para os grupos
- 2) Atribuir cada objeto para o grupo de centro mais próximo, obedecendo a lista de *cannot-link*
- 3) Mover cada centro para a média (centróide) dos objetos do grupo correspondente
- 4) Repetir os passos 2 e 3 até que algum critério de convergência seja obtido
- 5) Aglomerar grupos com objetos *must-link*

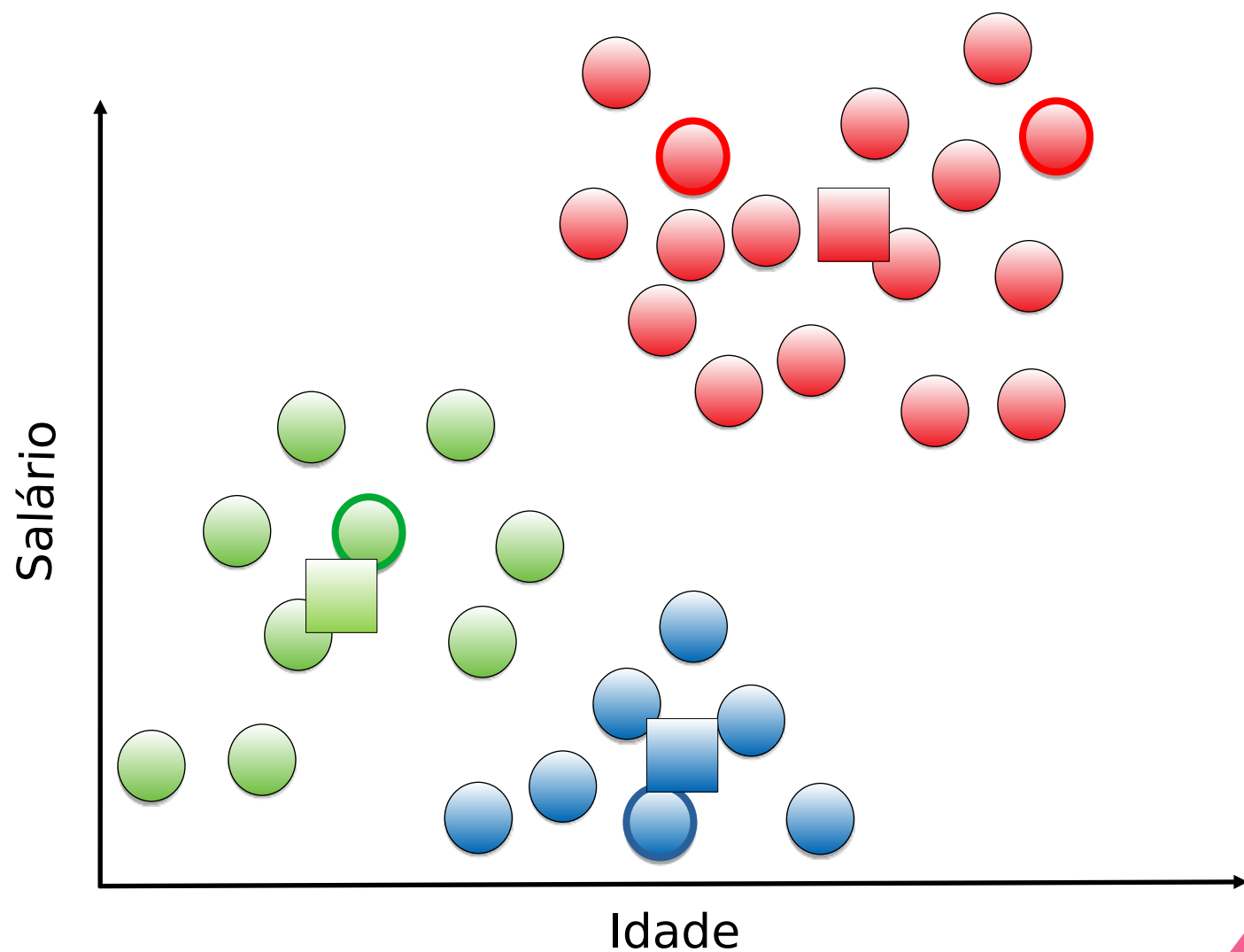


# Exemplo: $k$ -médias com restrições





## Exemplo: $k$ -médias com restrições



# SSDBSCAN

- Versão semi-supervisionada do DBSCAN (Lelis and Sanders 2009)
  - Baseada na distância de alcance mútuo por densidade
- Resolve o problema: para cada objeto não rotulado, rotular conforme o objeto rotulado mais próximo segundo uma perspectiva de densidade
  - Sem violar consistências



# SSDBSCAN

- $\epsilon$ -Vizinhança: conjunto de pontos com distância, no máximo,  $\epsilon$  para o ponto de referência  $p$

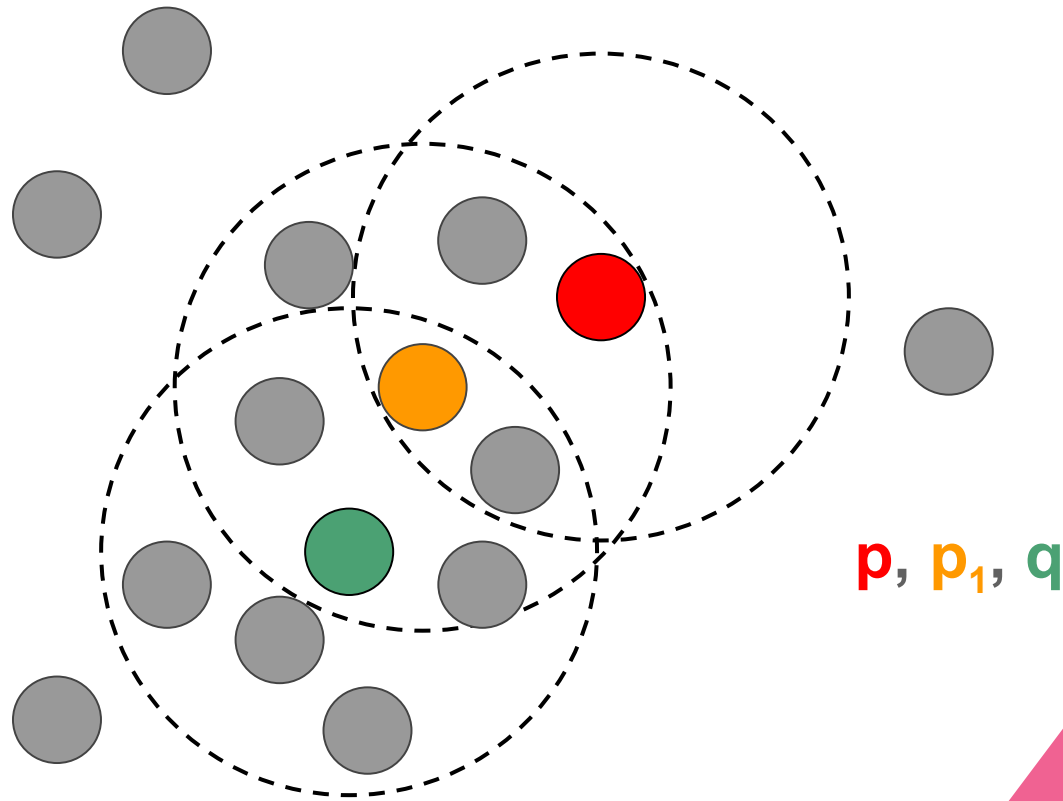
$$N_{\epsilon}(p) = \{q | d(p, q) \leq \epsilon\}$$

- $m_{pts}$ : Número mínimo de pontos na  $\epsilon$ -Vizinhança para considerar  $p$  como um ponto de região densa



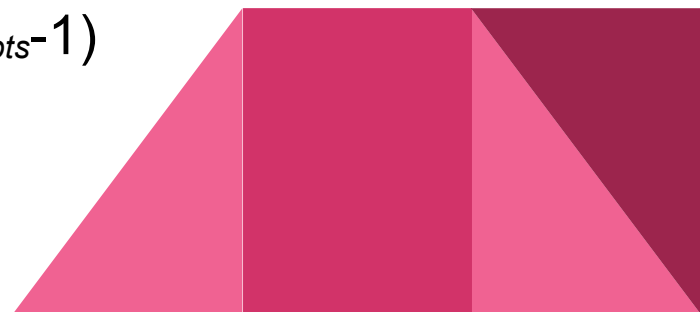
# SSDBSCAN

- $\epsilon$ -Vizinhança e a alcançabilidade



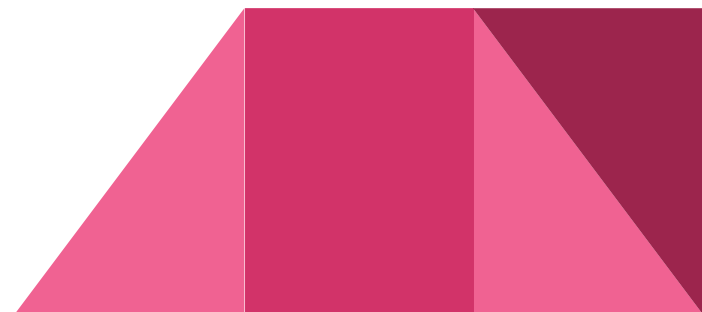
# SSDBSCAN

- No DBSCAN,  $p$  e  $q$  são diretamente alcançáveis em  $\varepsilon$  se:
  - $p$  e  $q$  são pontos de núcleo
    - ou seja,  $\varepsilon$  tem que ser grande o suficiente para que ambos tenham  $m_{pts}$  pontos na  $\varepsilon$ - vizinhança
  - $\varepsilon$  é grande o suficiente para  $p$  alcançar  $q$  (e vice versa)
    - Ou seja,  $d(p,q)$
- Distância de Alcançabilidade Mútua entre  $p$  e  $q$  é:
$$mrd(p,q) = \max(\text{core}(p), \text{core}(q), d(p,q) )$$
  - em que  $\text{core}(p)$  é a distância de  $p$  para seu  $(m_{pts}-1)$  vizinho mais próximo



# SSDBSCAN

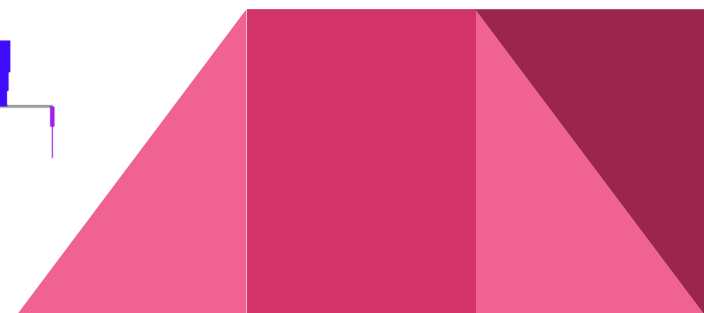
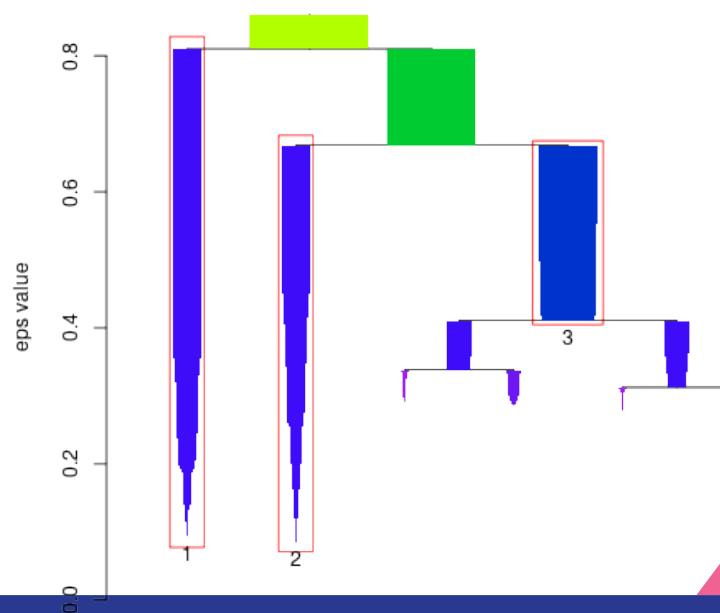
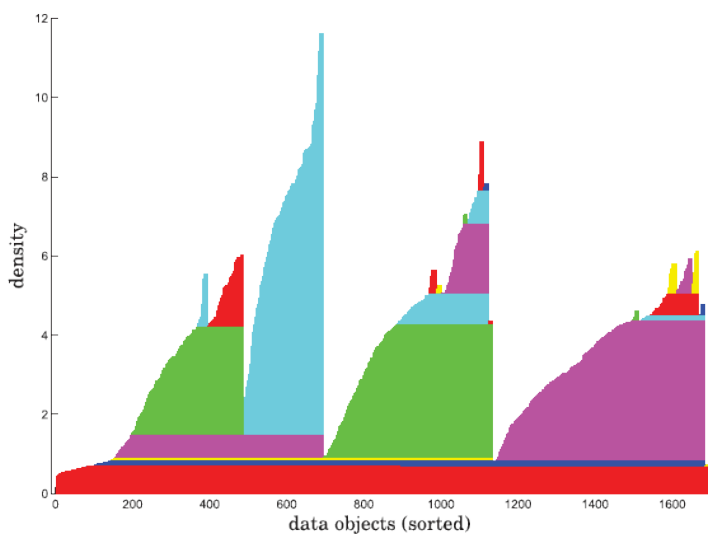
- Utiliza algoritmo de propagação de rótulos próprio:
  - 1) Encontra a **MST** do grafo completo ponderado pela distância de alcançabilidade mútua considerando os valores dados de  $m_{pts}$  e  $\epsilon$
  - 2) Encontre a maior aresta que conecte cada objeto não rotulado  $\mathbf{x}_i \in \mathbf{X}_U$  com qualquer objeto rotulado  $\mathbf{x}_j \in \mathbf{X}_L$
  - 3) Faça  $classe(\mathbf{x}_i) = classe(\mathbf{x}_j)$  em que  $\mathbf{x}_j$  é o objeto para qual a aresta selecionada pelo passo dois é mínima
  - 4) Repita o processo a partir do passo 3, se necessário



# HDBSCAN

- HDBSCAN é a versão hierárquica do DBSCAN
- Diferentemente do SSDBSCAN que utiliza uma partição gerada pelo DBSCAN, o HDBSCAN produz uma hierarquia com todas as partições de DBSCAN para todos os valores de  $\epsilon$ , dado um valor fixo de  $m_{pts}$ 
  - Ou seja, seu único parâmetro é o  $m_{pts}$

HDBSCAN\*

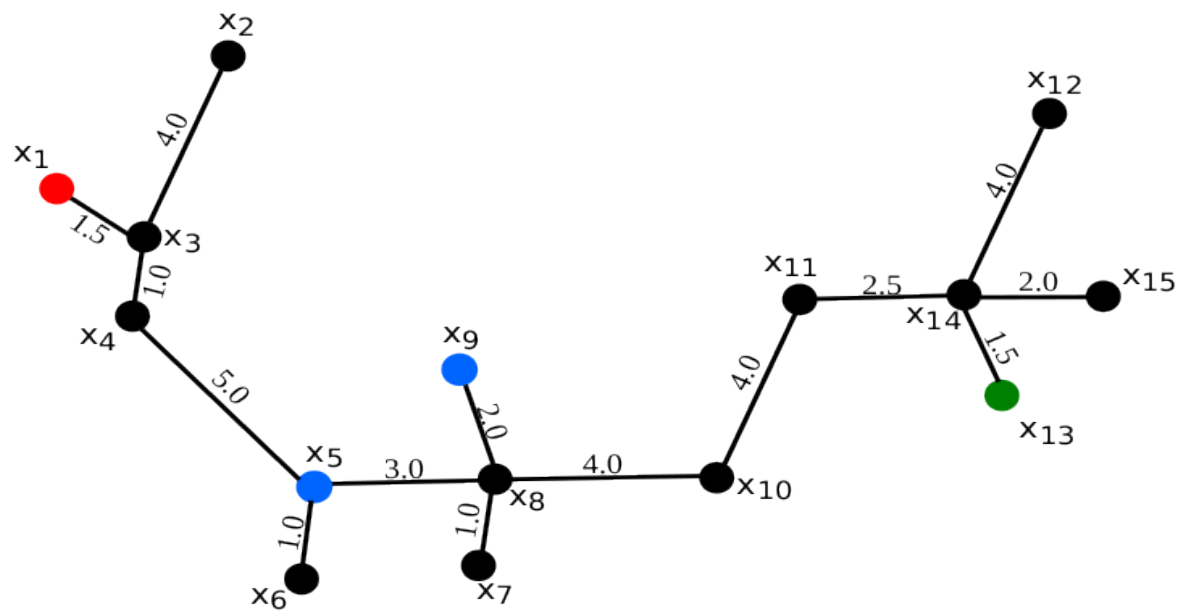


# HDBSCAN

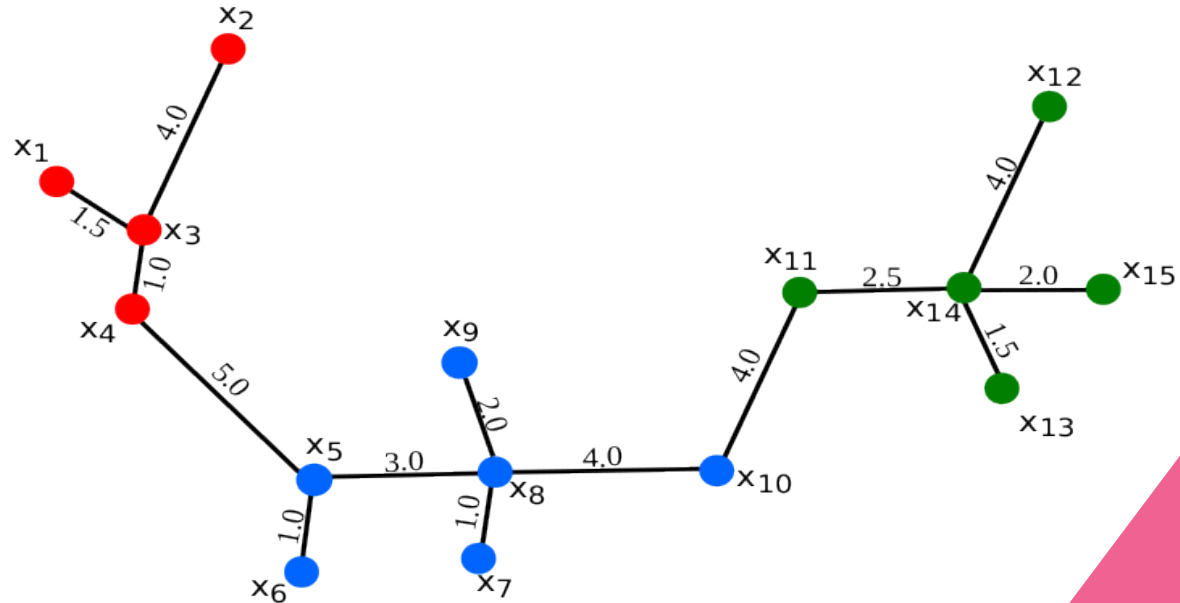
- Em Gerturdes et.al. 2018, foi proposto um arcabouço unificado para aprendizado semi-supervisionado baseado em densidade
  - Que utiliza a estrutura principal do HDBSCAN
- Dada a **MST** do HDBSCAN, feita sobre o grafo completo de  $\mathbf{X} = \{\mathbf{X}_U \cup \mathbf{X}_L\}$ , com arestas ponderadas pela distância de alcançabilidade mútua
  - Propaga rótulos baseado na seleção do caminho com a menor maior aresta que conecta um objeto rotulado com um não rotulado
  - A ideia é sair de cada objeto rotulado, propagando pela **MST**
    - Respeitando restrições







(a)  $MST_r$  with 4 pre-labeled objects ( $X_L = \{x_1, x_5, x_9, x_{13}\}$ ) from 3 classes (red, blue, green)



(b) Result of the label expansion

# Isso vai looooooooooooooonge

LIGUE OS PONTOS NA ORDEM CORRETA,  
E AJUDE A ARIEL A DESCOBRIR UM NOVO AMIGO!



# Referências

- Van Engelen, J.E., Hoos, H.H. A survey on semi-supervised learning. Mach Learn 109, 373–440 (2020). <https://doi.org/10.1007/s10994-019-05855-6>
- L. Lelis and J. Sander. 2009. Semi-supervised Density-Based Clustering. In Proc. IEEE ICDM. 842–847.
- Zhu, Xiaojin & Ghahramani, Zoubin. (2003). Learning from Labeled and Unlabeled Data with Label Propagation. School of Computer Science. Carnegie Mellon University.
- Jadson Castro Gertrudes, Arthur Zimek, Jörg Sander, and Ricardo J. G. B. Campello. 2018. A unified framework of density-based clustering for semi-supervised classification. In Proceedings of the 30th International Conference on Scientific and Statistical Database Management (SSDBM '18). Association for Computing Machinery, New York, NY, USA, Article 11, 1–12. <https://doi.org/10.1145/3221269.3223037>