Aprendizado Semi-Supervisionado Baseado em Grafos

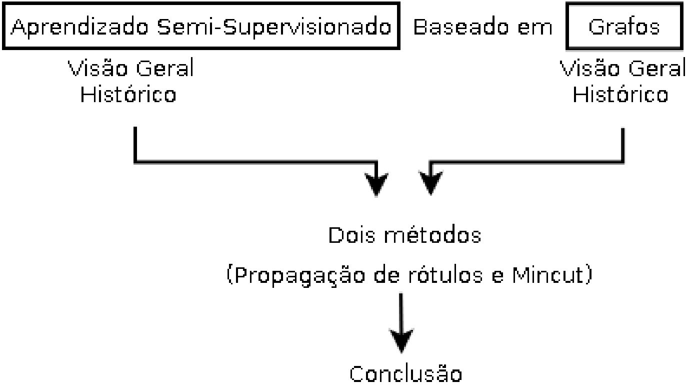
Seminário da Matéria SCC5882

Professores: Dr. Zhao Liang e Dr. Alneu de Andrade Lopes

Aluno: Renato Fabbri

6 de julho de 2010

Resumo


Aprendizado Semi-Supervisionado
Baseado em
Grafos
Visão Geral
Histórico
Visão Geral
Histórico
Dois métodos
(Propagação de rótulos e Mincut)
Conclusão

A - Definições (1)

- Aprendizado Supervisionado - treinamento somente com os dados rotulados

A - Definições (1)

- Aprendizado Supervisionado - treinamento somente com os dados rotulados
- Aprendizado não Supervisionado - utilizados somente dados não rotulados

A - Definições (1)

- Aprendizado Supervisionado - treinamento somente com os dados rotulados
- Aprendizado não Supervisionado - utilizados somente dados não rotulados
- Aprendizado Semi-Supervisionado - utilização tanto dos dados rotulados quanto dos não rotulados.

A - Definições (2)

- Classificação Semi-Supervisionada, CSS. (Clusterização forçada não é foco)

A - Definições (2)

- Classificação Semi-Supervisionada, CSS. (Clusterização forçada não é foco)
- Indutivo e Transdutivo.

A - Interesse

- Criança aprendendo as coisas (animais).

A - Interesse

- Criança aprendendo as coisas (animais).
- Dados rotulados podem ser poucos ou custosos.

A - Caso 1

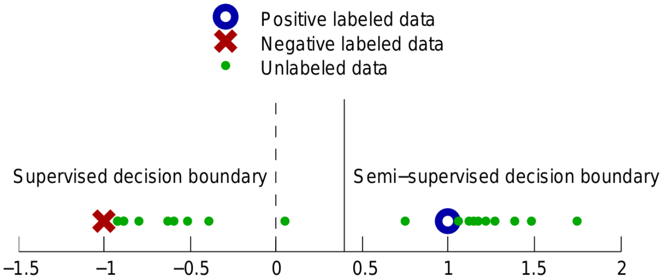
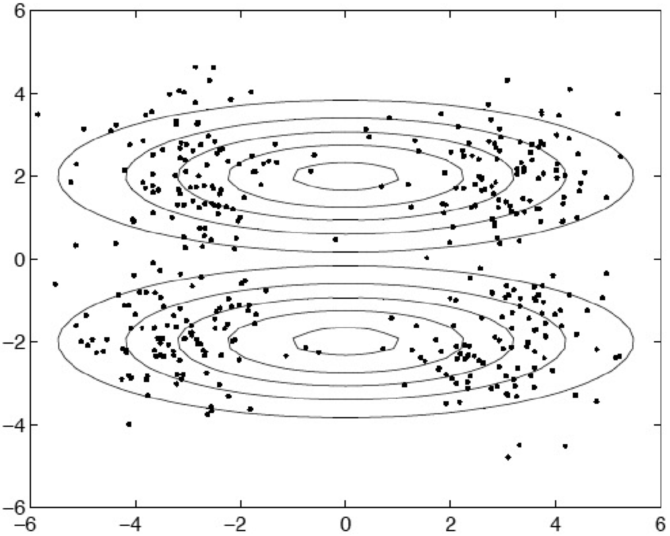


Figura: Forma em que a CSS auxilia na classificação. Fonte: Zhu, X. and Lafferty, J. and Rosenfeld, R., 2005

A - Caso 2



A - Breve Histórico

- Primeiros: Modelos de Mistura.

A - Breve Histórico

- Primeiros: Modelos de Mistura.
- Primeiros: Self-Training (self-teaching e bootstrapping).

A - Breve Histórico

- Primeiros: Modelos de Mistura.
- Primeiros: Self-Training (self-teaching e bootstrapping).
- Mais recentes: Co-training e SVM transdutivo.

A - Breve Histórico

- Primeiros: Modelos de Mistura.
- Primeiros: Self-Training (self-teaching e bootstrapping).
- Mais recentes: Co-training e SVM transdutivo.
- Mais recentes: Métodos baseados em grafos.

G & RC - Definição (1)

- Conjunto de objetos, alguns ligados.

G & RC - Definição (1)

- Conjunto de objetos, alguns ligados.
- Objetos $\rightarrow$ nós ou vértices. Ligações $\rightarrow$ arestas.

G & RC - Definição (1)

- Conjunto de objetos, alguns ligados.
- Objetos $\rightarrow$ nós ou vértices. Ligações $\rightarrow$ arestas.
- Visualmente $\rightarrow$ pontos e linhas.

G & RC - Variações

- Arestas Direcionadas.

G & RC - Variações

- Arestas Direcionadas.
- Pesos nas arestas.

G & RC - Variações

- Arestas Direcionadas.
- Pesos nas arestas.
- Mais de um tipo de aresta.

G & RC - Variações

- Arestas Direcionadas.
- Pesos nas arestas.
- Mais de um tipo de aresta.
- Atributos quaisquer associados às arestas e aos nós.

G & RC - Exemplos Clássicos

- Regular.

G & RC - Exemplos Clássicos

- Regular.
- Completo (*click*).

G & RC - Exemplos Clássicos

- Regular.
- Completo (*click*).
- Conectado.

G & RC - Exemplos Clássicos

- Regular.
- Completo (*click*).
- Conectado.
- Randômico.

G & RC - Exemplos Clássicos

- Regular.
- Completo (*click*).
- Conectado.
- Randômico.
- trivial, nulo, etc.

G & RC - Medidas mais Comuns

- Grau.

G & RC - Medidas mais Comuns

- Grau.
- Força.

G & RC - Medidas mais Comuns

- Grau.
- Força.
- Entrada e Saída.

G & RC - Rede Complexa

- Grafo de grandes dimensões.
- Casos triviais - RC randômica e regular.

G & RC - Breve Histórico

- Euler - "Konigsberg Bridge problem".

G & RC - Breve Histórico

- Euler - "Konigsberg Bridge problem".
- Cayley - Química.

G & RC - Breve Histórico

- Euler - "Konigsberg Bridge problem".
- Cayley - Química.
- "Grafo" Nature - Silvester 1878.

G & RC - Breve Histórico

- Euler - "Konigsberg Bridge problem".
- Cayley - Química.
- "Grafo" Nature - Silvester 1878.
- Grafos, topologia, álgebra - Lei de Kirchoff

G & RC - Breve Histórico

- Euler - "Konigsberg Bridge problem".
- Cayley - Química.
- "Grafo" Nature - Silvester 1878.
- Grafos, topologia, álgebra - Lei de Kirchoff
- Teorias probabilísticas

G & RC - Breve Histórico

- Euler - "Konigsberg Bridge problem".
- Cayley - Química.
- "Grafo" Nature - Silvester 1878.
- Grafos, topologia, álgebra - Lei de Kirchoff
- Teorias probabilísticas
- RC naturais - propriedades não triviais - LE, PM.

G & RC - Breve Histórico

- Euler - "Konigsberg Bridge problem".
- Cayley - Química.
- "Grafo" Nature - Silvester 1878.
- Grafos, topologia, álgebra - Lei de Kirchoff
- Teorias probabilísticas
- RC naturais - propriedades não triviais - LE, PM.
- Neurociência, Física, Linguística, Computação, etc.

CSS em G

- Algorítmo de regularização.
- Mincut
- Propagação de Rótulo.
- Modelos de Mistura.
- Máxima Entropia.
- ETC.

PR - Concepção

- Vizinho mais próximo $\rightarrow$ propaga mais fácil.

PR - Concepção

- Vizinho mais próximo $\rightarrow$ propaga mais fácil.
- Rótulos fixos nos rotulados de antemão.

PR - Concepção

- Vizinho mais próximo $\rightarrow$ propaga mais fácil.
- Rótulos fixos nos rotulados de antemão.
- Estes "emanam" os seus rótulos.

Convenções e Exposição do Problema (1)

- $(x_1, y_1)...(x_l, y_l)$ os dados rotulados.

Convenções e Exposição do Problema (1)

- $(x_1, y_1)...(x_l, y_l)$ os dados rotulados.
- $y \in 1...C$ com C o número de classes.

Convenções e Exposição do Problema (1)

- $(x_1, y_1)...(x_l, y_l)$ os dados rotulados.
- $y \in 1...C$ com C o número de classes.
- $x_{l+1}...x_{l+u}$.

Convenções e Exposição do Problema (1)

- $(x_1, y_1)...(x_l, y_l)$ os dados rotulados.
- $y \in 1...C$ com C o número de classes.
- $x_{l+1}...x_{l+u}$.
- Geralmente $l \ll u$.

Convenções e Exposição do Problema (1)

- $(x_1, y_1)...(x_l, y_l)$ os dados rotulados.
- $y \in 1...C$ com C o número de classes.
- $x_{l+1}...x_{l+u}$.
- Geralmente $l \ll u$.
- Seja $n = l + u$

Convenções e Exposição do Problema (1)

- $(x_1, y_1)...(x_l, y_l)$ os dados rotulados.
- $y \in 1...C$ com C o número de classes.
- $x_{l+1}...x_{l+u}$.
- Geralmente $l \ll u$.
- Seja $n = l + u$
- L e U denotam os dados rotulados e não rotulados, respectivamente

Convenções e Exposição do Problema (2) - Notas

- Cada classe de C presente em ao menos 1 elemento.

Convenções e Exposição do Problema (2) - Notas

- Cada classe de C presente em ao menos 1 elemento.
- Método Transdutivo.

Convenções e Exposição do Problema (3) - Obtenção do Grafo

- Completo.

Convenções e Exposição do Problema (3) - Obtenção do Grafo

- Completo.
- $w_{ij} = exp(-\frac{\|x_i - x_j\|^2}{\alpha^2})$

Convenções e Exposição do Problema (3) - Obtenção do Grafo

- Completo.
- $w_{ij} = exp(-\frac{\|x_i - x_j\|^2}{\alpha^2})$
- α é um *hiperparâmetro*

Convenções e Exposição do Problema (4) - Transição

- P_{ij} é a probabilidade de transição do nó i para o nó j (passagem do rótulo)

Convenções e Exposição do Problema (4) - Transição

- P_{ij} é a probabilidade de transição do nó i para o nó j (passagem do rótulo)
- $P_{ij} = P(i \to j) = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}}$

Convenções e Exposição do Problema (5) - Últimas Definições

- Matriz de rótulos Y_L, $l \times C$, cuja *iésima* linha possui 1 na coluna correspondente à classe do dado x_i.

Convenções e Exposição do Problema (5) - Últimas Definições

- Matriz de rótulos Y_L, $l \times C$, cuja *iésima* linha possui 1 na coluna correspondente à classe do dado x_i.
- f uma matriz $n \times C$ em que cada linha pode ser interpretada como uma distribuição de probabilidade sobre os rótulos.

Convenções e Exposição do Problema (5) - Tarefa!

A tarefa consiste em computar os rótulos dos nós todos (na verdade somente dos não rotulados)

O Algoritmo de Propagação de Rótulo

- Propague $f \leftarrow Pf$. (todos os nós propagam o seu rótulo para os seus vizinhos)
- Mantenha os dados rotulados iniciais $f_L = Y_L$. (assegura que os rótulos dos nós inicialmente rotulados não sejam sobrescritos)
- Repita do passo 1 até que o algoritmo convirja.

PR - Convergência (1)

$$f = \begin{pmatrix} f_L \\ f_U \end{pmatrix} \tag{1}$$

$$\begin{bmatrix} P_{LL} & P_{LU} \\ P_{UL} & P_{UU} \end{bmatrix} \tag{2}$$

$$f_U \leftarrow P_{UU} f_U + P_{UL} Y_L \tag{3}$$

$$f_u = \lim_{n \to \infty} [(P_{UU})^n f_U^0 + (\sum_{i=1}^{n} (P_{UU})^{(i-1)}) P_{UL} Y_L] \tag{4}$$

PR - Convergência (2)

$$f_u = \lim_{n\to\infty}[(P_{UU})^n f_U^0 + (\sum_{i=1}^{n}(P_{UU})^{(i-1)})P_{UL}Y_L] \tag{5}$$

$$(P_{UU})^n f_U^0 \to 0 \tag{6}$$

$$\exists\, \gamma < 1 : \sum_{j=1}^{u}(P_{UU})_{ij} \leq \gamma, \forall i = 1...u \tag{7}$$

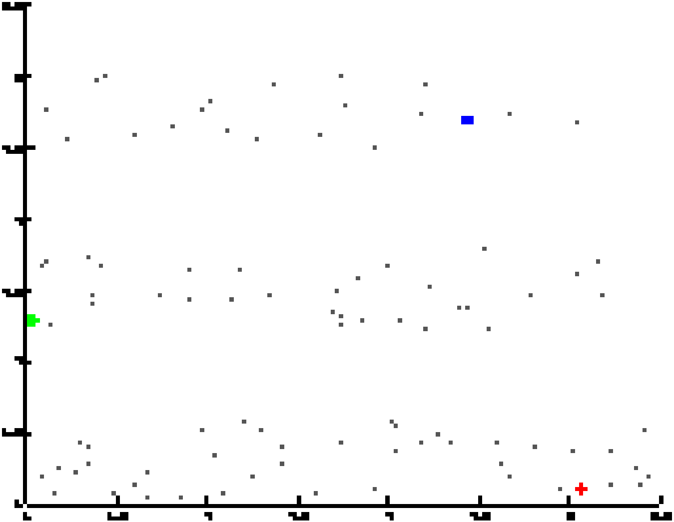
Portanto:

$$\sum_{j}(P_{UU})_{ij}^{n} \leq \gamma^n \tag{8}$$

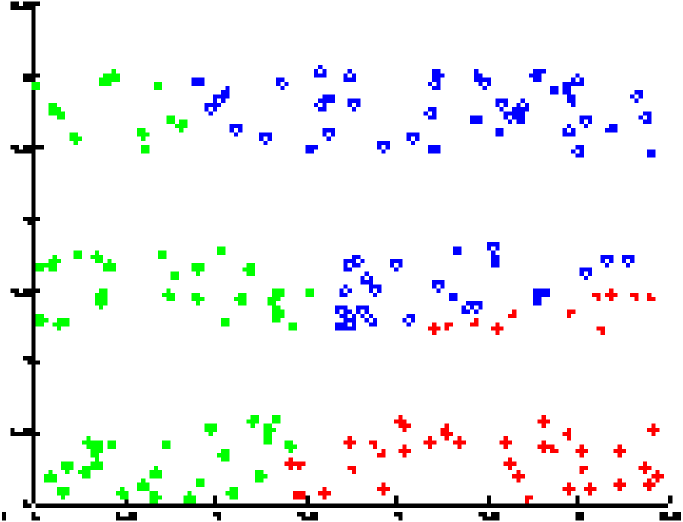
Então:

$$f_u = (I - P_{UU})^{-1}P_{UL}Y_L \tag{9}$$

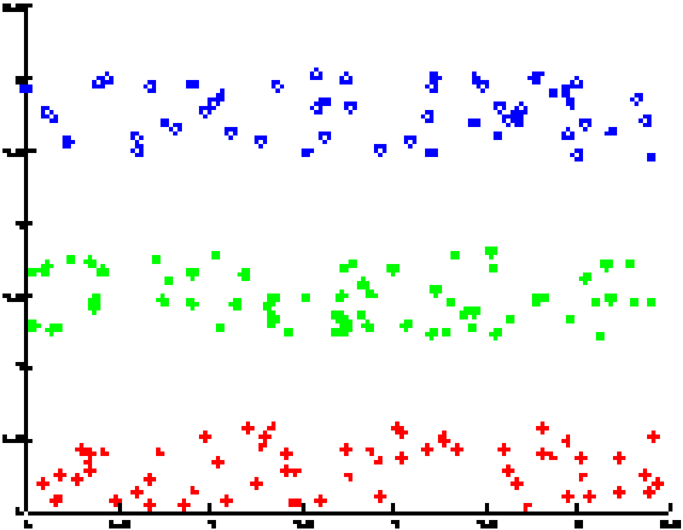
PR - Exemplo 1 (1)



PR - Exemplo 1 (2)



PR - Exemplo 1 (3)



PR - Exemplo 2 (1)

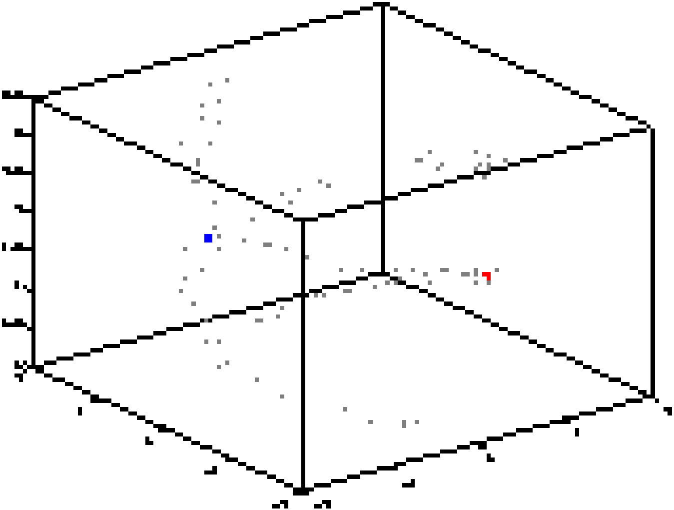


Figura: Dados originais, 2 rotulados, 184 não rotulados

PR - Exemplo 1 (2)

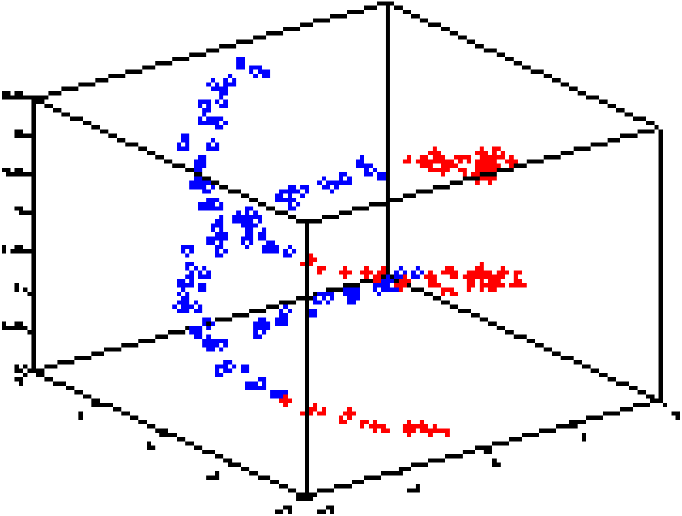


Figura: Dados espiralados rotulados utilizando-se knn.

PR - Exemplo 1 (2)

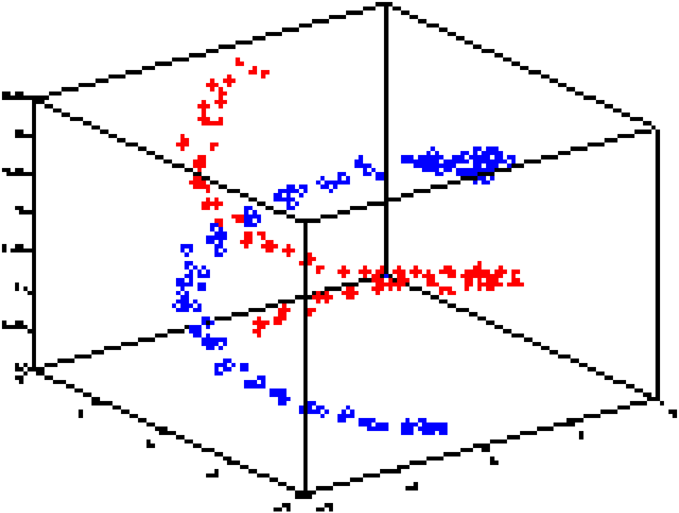


Figura: Dados espiralados rotulados utilizando propagação de rótulos.

Mincut - Concepção

- Objetos parecidos devem ficar juntos.
- Seccionamento por separações de custos mínimos.

Mincut - Convenções e Exposição do Problema

- rotulados L e não rotulados U.

Mincut - Convenções e Exposição do Problema

- rotulados L e não rotulados U.
- L_+ os positivos; L_- os negativos.

Mincut - Convenções e Exposição do Problema

- rotulados L e não rotulados U.
- L_+ os positivos; L_- os negativos.
- v_+ e v_-

O Algoritmo de Propagação de Mincut

- $G = (V, E)$, onde $V = L \cup U$ e $E \subseteq V \times V$
- $e \in E$ é associado um peso $w(e)$
- $w(v_1, v_2) = \infty, \forall\ v_1, v_2 \in L_+$ e $w(v_1, v_2) = \infty, \forall\ v_1, v_2 \in L_-$.
- Função de atribuição de pesos, denotada por w.
- Determinamos o conjunto de arestas com a menor soma de pesos que, se removidas, separam todos os v_+ dos v_-. (Corte mínimo)
- Por fim, rotulamos como positivos todos os vértices em V_+ e como negativos todos os vértives em V_-.

Calibragem do Mincut

- w é sem dúvida o crucial neste algorítmo.
- **Mincut-3**
- **Mincut-**δ, **Mincut-**δ_0, **Mincut-**$\delta_{\frac{1}{2}}$, **Mincut-**δ_{opt}

Mincut - Problemas

- Difícil Calibragem.
- Casos de poucos exemplos rotulados

Mincut - Desempenho Comparativo (1)

Dataset	$\lvert L\rvert \& \lvert U\rvert$	number of features	Mincut				ID3	3-NN
			Mincut-3	Mincut-δ_{opt}	Mincut-δ_0	Mincut-$\delta_{1/2}$		
Mush	20+1000	22	82.1	**97.7**	**97.7**	97.0	93.3	91.1
Mush*	20+1000	22	74.2	**88.7**	56.9	**87.0**	80.8	83.3
Tae	10+100	5	86.0	**99.0**	96.0	**97.0**	86.0	80.0
Tae*	10+100	5	76.0	**96.0**	86.0	**94.0**	76.0	62.0
Voting	45+390	16	89.1	**91.3**	66.1	83.3	86.4	**89.6**
Musk	40+200	166	73.0	**92.5**	91.0	**92.5**	83.5	87.0
Pima	50+718	8	63.8	**72.3**	48.8	**72.3**	70.0	68.1
Iono	50+300	34	71.0	81.6	78.0	77.6	**88.6**	69.6
Bupa	45+300	6	53.3	**59.3**	48.0	41.7	**55.3**	52.7
MI	124+432	6	70.0	64.4	64.4	64.4	**98.6**	81.1
MII	169+432	6	**68.6**	67.2	57.2	67.2	67.9	63.6
MIII*	122+432	6	79.1	80.6	64.8	80.6	**94.4**	83.6

Figura: Mincut com alguns valores para δ e comparação com outros métodos.

Mincut - Desempenho Comparativo (2)

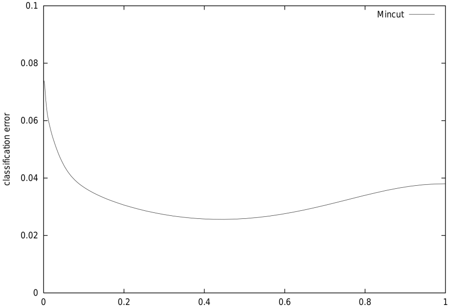


Figura: Variação do erro com o Delta.

Mincut - Desempenho Comparativo (3)

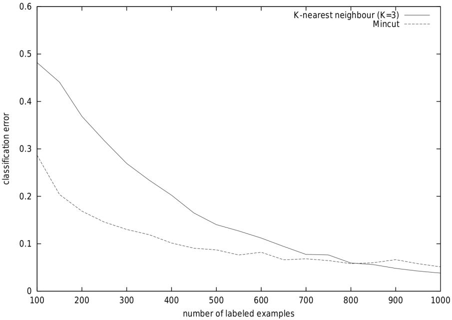


Figura: Comparação do erro entre mincut e kNN com a variação de exemplos rotulados δ_{opt}.

Mincut - Desempenho Comparativo (4)

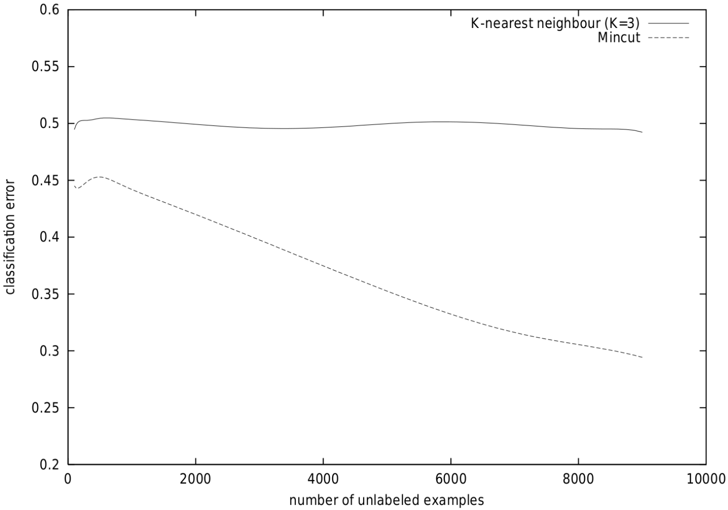


Figura: Comparação do erro entre mincut e kNN com a variação de exemplos não

Conclusão

- Com segurança se as classes são clusterizadas entre si.
- Na falta de dados rotulados
- Seguindo a distribuição dos dados evitam o critério único da proximidade dos dados rotulados.
- Melhor que o kNN para os casos que vimos.
- Complexidade simples-moderada.

Referências Principais

- Zhu, X. and Lafferty, J. and Rosenfeld, R. "Semi-supervised learning with graphs.", 2005
- Zhu, X. and Ghahramani, Z., "Learning from labeled and unlabeled data with label propagation.", 2002
- Blum, A. and Chawla, S., "Learning from labeled and unlabeled data using graph mincuts", 2001

FIM.

FIM.