Multi-task Regularization of Generative Similarity Models

International Workshop on Similarity-based Pattern Analysis and Recognition – SIMBAD '11

September 28-30, 2011

Dr. Luca Cazzanti
Applied Physics Lab
Univ. Washington
Seattle, USA

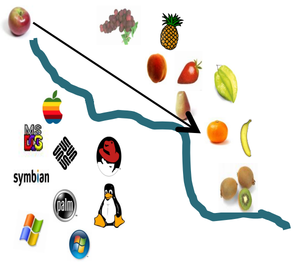
Prof. Maya Gupta
Dept. EE
Univ. Washington
Seattle, USA

Mr. Sergey Feldman
Dept. EE
Univ. Washington
Seattle, USA

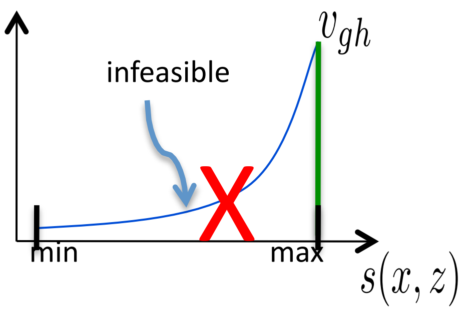
Dr. Michael Gabbay
Applied Physics Lab
Univ. Washington
Seattle, USA

Work supported by U.S Office of Naval Research – PM Dr. Ivy Estabrooke

Outline



1. Review local similarity discriminant analysis (local SDA)



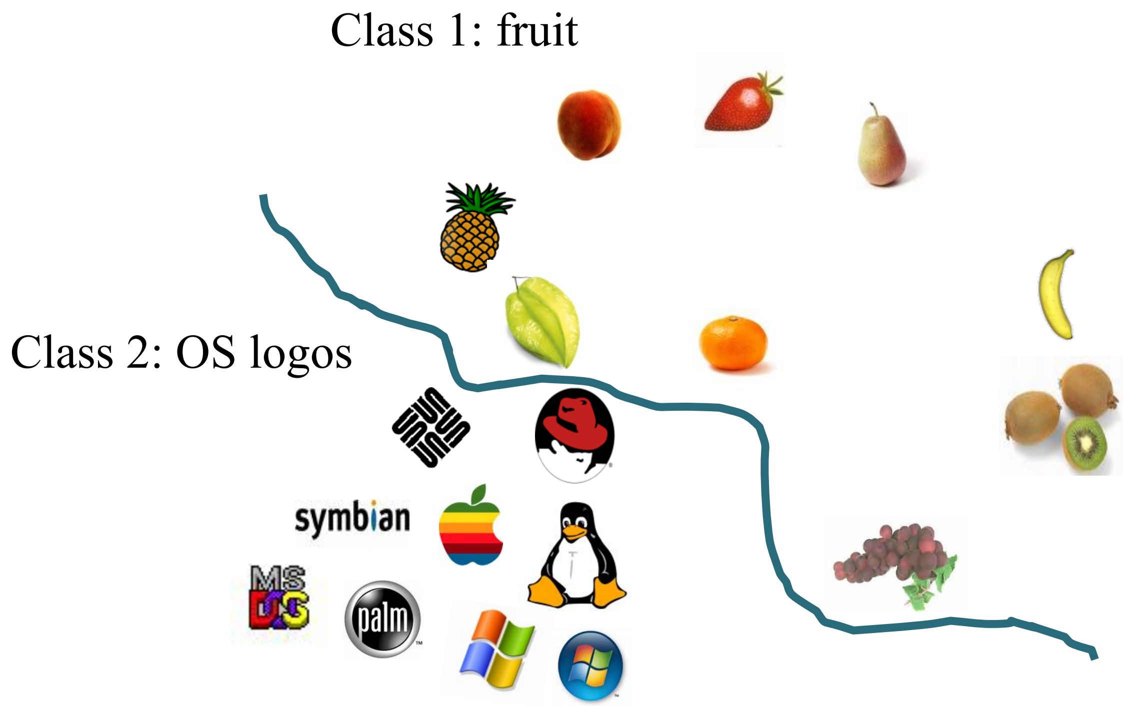
2. Need for regularization

$$\{v^*_{gh}\}_{g,h=1}^G = \arg\min_{\{\hat{v}_{gh}\}_{g,h=1}^G} \sum_{g,h=1}^G \sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} (s(z_a,z_b) - \hat{v}_{gh})^2) + \eta \sum_{j,k=1}^G \sum_{l,m=1}^G A(v_{jk},v_{lm})(\hat{v}_{jk} - \hat{v}_{lm})^2.$$

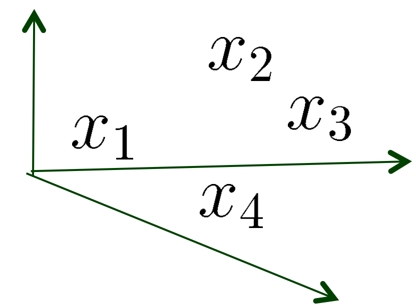
3. Multi-task regularization for local SDA

4. Computer experiments and discussion

Euclidean Features



x_i described by
four numbers, $x_i \in \mathbb{R}^4$:
RGB color triplet
image texture



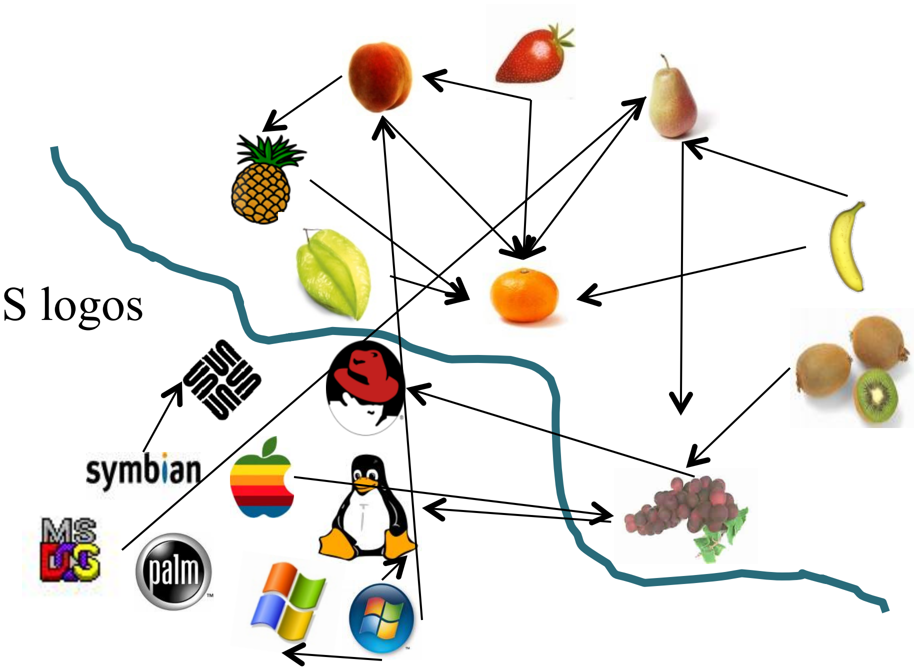
Goal: classify as either fruit of OS logo
This is the conventional statistical learning set-up

$$P(x_i | Y = g)$$

Similarities

Class 1: fruit

Class 2: OS logos



Information about the relationship between samples:

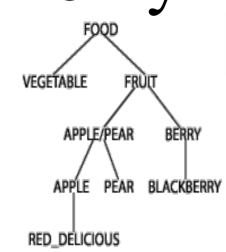
If you like 

then you like 

Human judgments of similarity

eduicons.com

Given taxonomy of objects

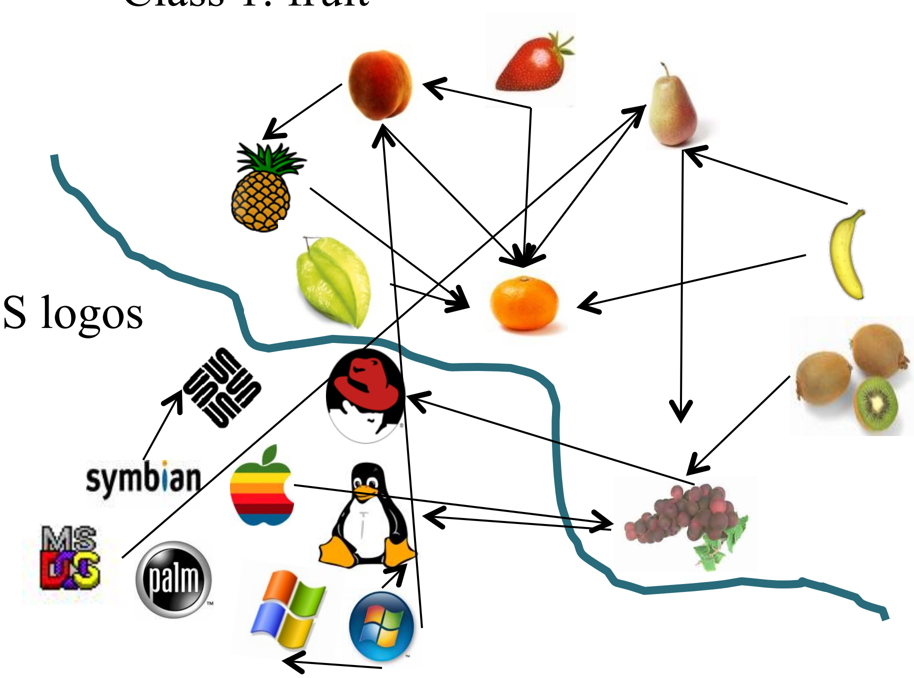


www.indiana.edu/~hlw/Meaning/appleTaxonomy2.gif

Similarities

Class 1: fruit

Class 2: OS logos



Information about the relationship between samples:

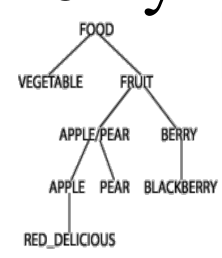
If you like 

then you like 

Human judgments of similarity

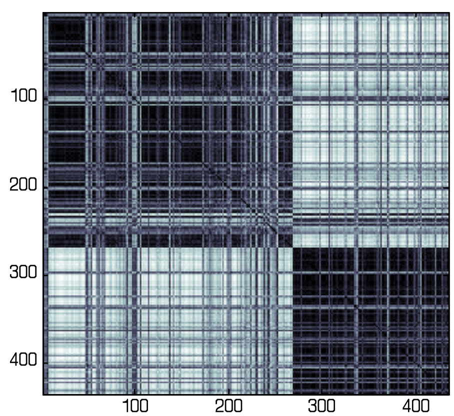
eduicons.com

Given taxonomy of objects



www.indiana.edu/~hlw/Meaning/appleTaxonomy2.gif

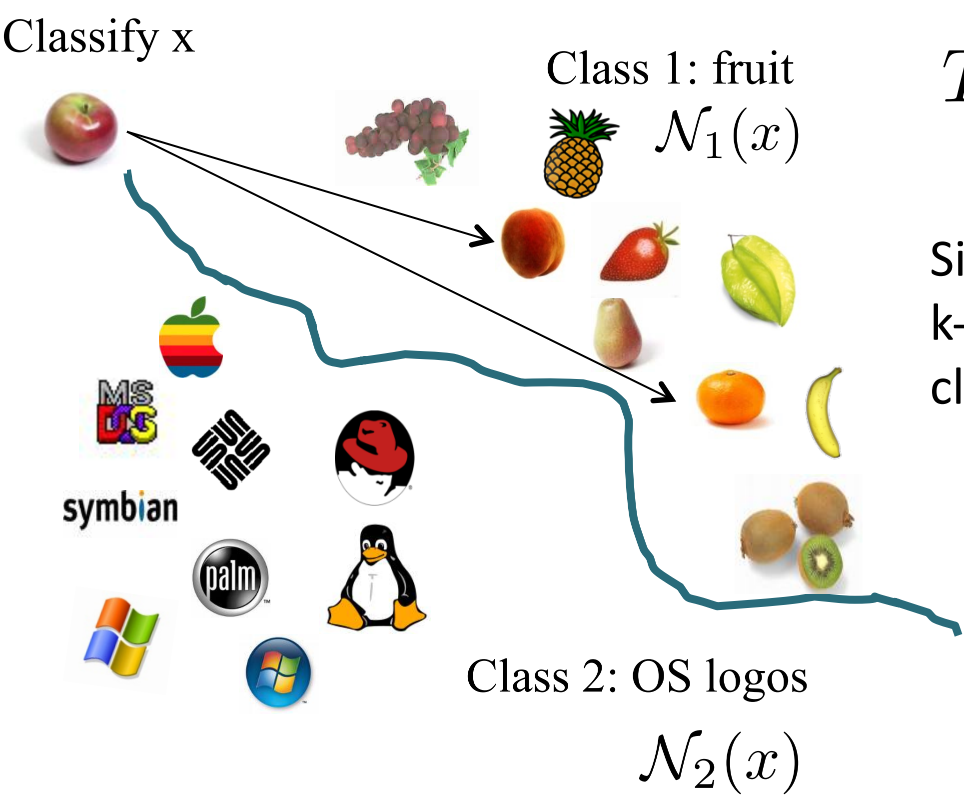
Matrix of pairwise similarities



$$P(s(x_i, x_j)|Y = g)$$

Local Similarity Discriminant Analysis (local SDA)

(Cazzanti '07, Cazzanti & Gupta '07)



$$T_1(x) = \{s(x,z)\}, z \in \mathcal{N}_1(x)$$

Similarities of test sample *x* to all its k-most similar training samples from class 1.

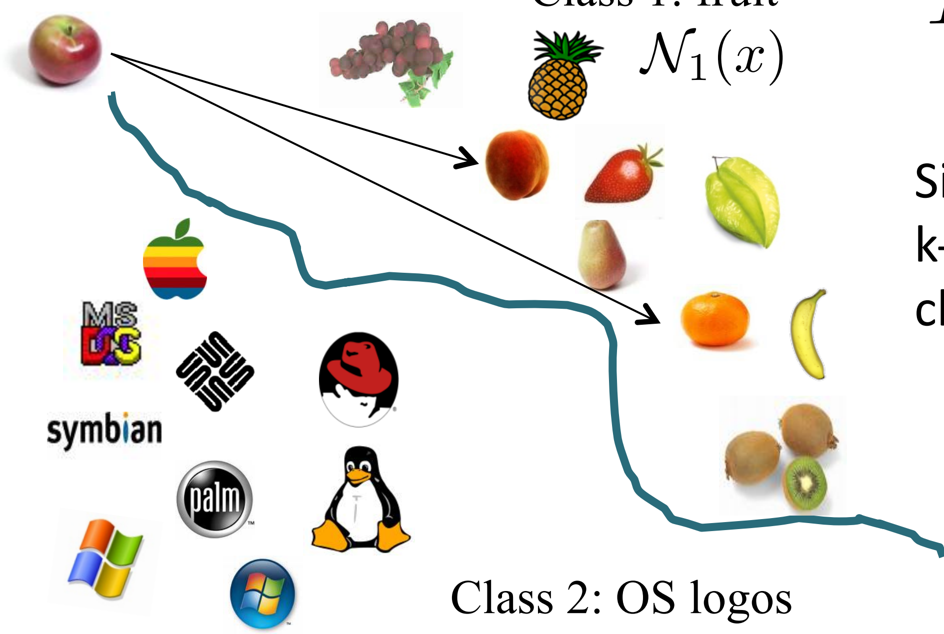
Local Similarity Discriminant Analysis (local SDA)

(Cazzanti '07, Cazzanti & Gupta '07)

Classify x

Class 1: fruit

$\mathcal{N}_1(x)$



Class 2: OS logos

$\mathcal{N}_2(x)$

$$T_1(x) = \{s(x,z)\}, z \in \mathcal{N}_1(x)$$

Similarities of test sample *x* to all its k-most similar training samples from class 1.

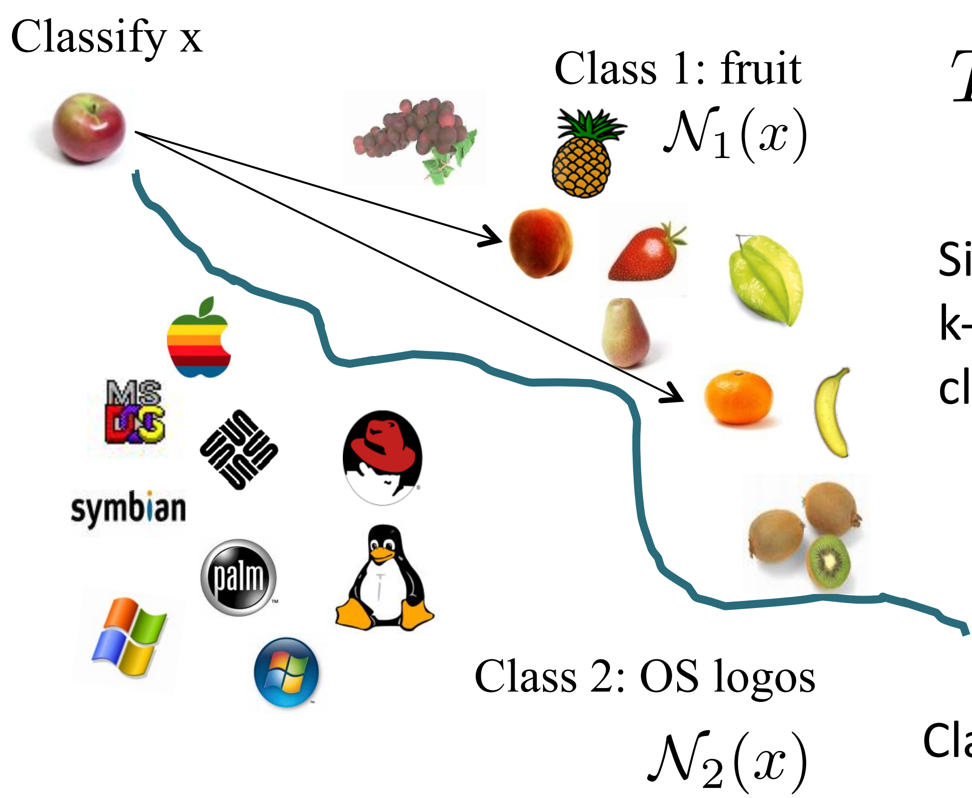
$$P_1(T_1(x) | Y = 1)$$

Assumed class for *x*

Class x is compared to

Local Similarity Discriminant Analysis (local SDA)

(Cazzanti '07, Cazzanti & Gupta '07)



$$T_1(x) = \{s(x,z)\}, z \in \mathcal{N}_1(x)$$

Similarities of test sample *x* to all its k-most similar training samples from class 1.

$$P_1(T_1(x)|Y = 1)$$

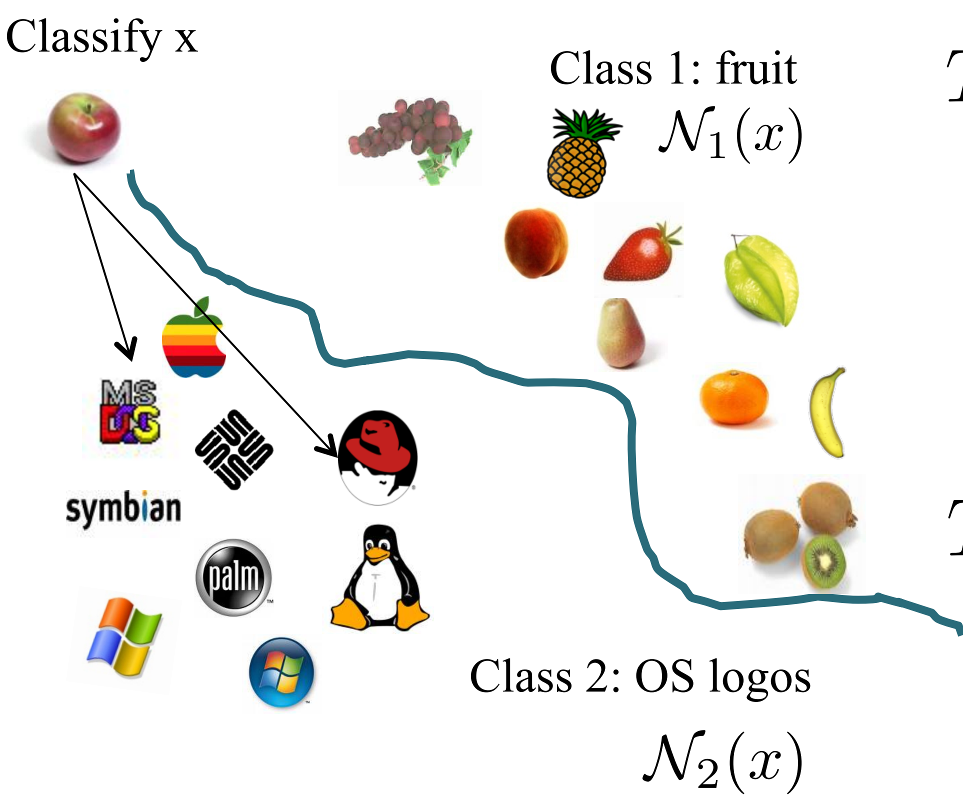
Assumed class for *x*

Class x is compared to

$$P_1(T_1(x)|Y = 2)$$

Local Similarity Discriminant Analysis (local SDA)

(Cazzanti '07, Cazzanti & Gupta '07)



$$T_1(x) = \{s(x,z)\}, z \in \mathcal{N}_1(x)$$

$$P_1(T_1(x)|Y=1)$$

$$P_1(T_1(x)|Y=2)$$

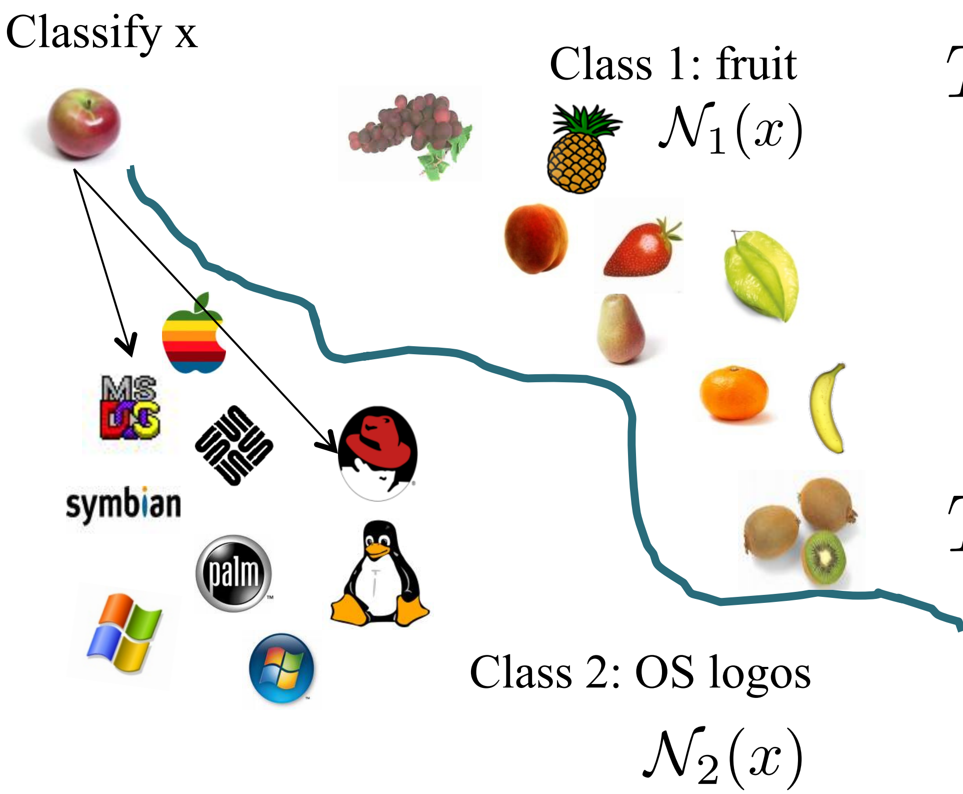
$$T_2(x) = \{s(x,z)\}, z \in \mathcal{N}_2(x)$$

$$P_2(T_2(x)|Y=1)$$

$$P_2(T_2(x)|Y=2)$$

Local Similarity Discriminant Analysis (local SDA)

(Cazzanti '07, Cazzanti & Gupta '07)



$$T_1(x) = \{s(x,z)\}, z \in \mathcal{N}_1(x)$$

$$P_1(T_1(x)|Y = 1)$$

$$P_1(T_1(x)|Y = 2)$$

$$T_2(x) = \{s(x,z)\}, z \in \mathcal{N}_2(x)$$

$$P_2(T_2(x)|Y = 1)$$

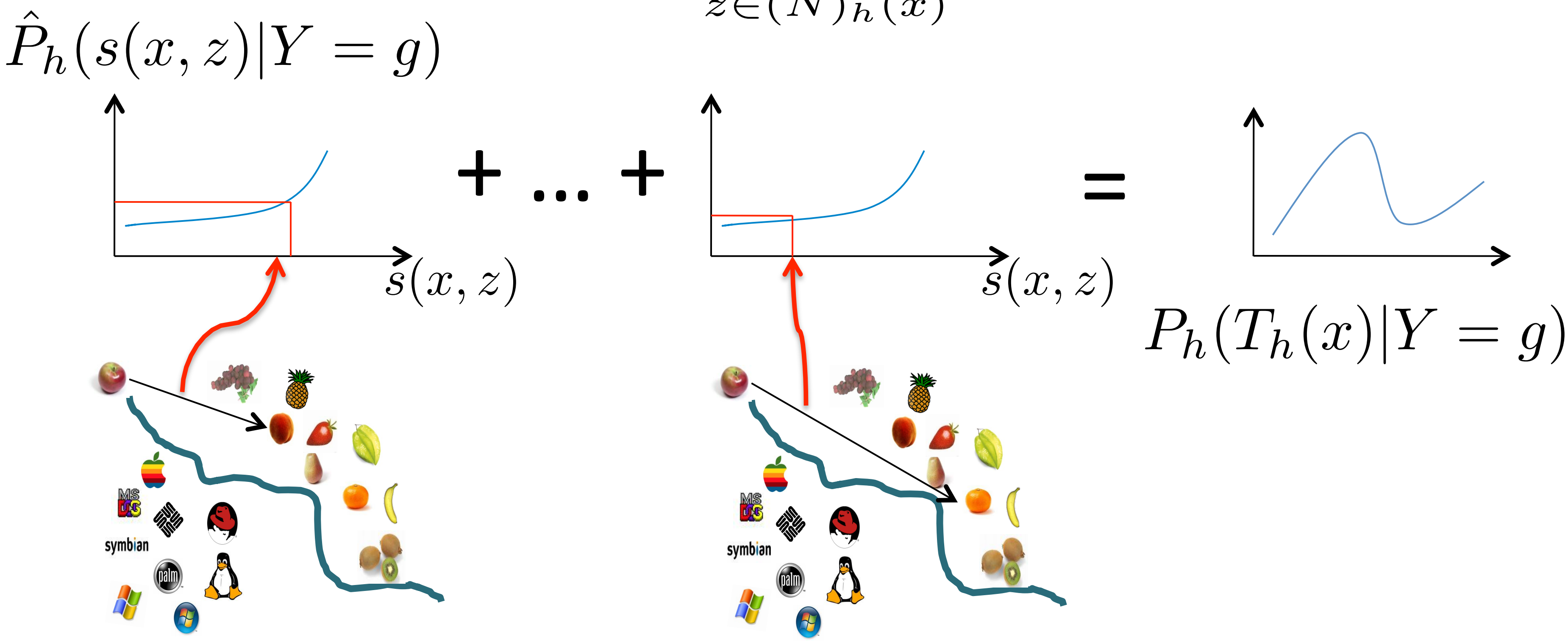
$$P_2(T_2(x)|Y = 2)$$

$$y = \arg\max_g \prod_{h=1}^{G} P_h(T_h(x)|Y = g)P(Y = g)$$

Local Similarity Discriminant Analysis (local SDA)

$$P_h(T_h(x)|Y=g) \triangleq \frac{1}{k_h} \sum_{z \in (N)_h(x)} \hat{P}_h(s(x,z)|Y=g)$$

$$= \frac{1}{k_h} \sum_{z \in (N)_h(x)} \gamma_{gh} e^{\lambda_{gh} s(x,z)}$$



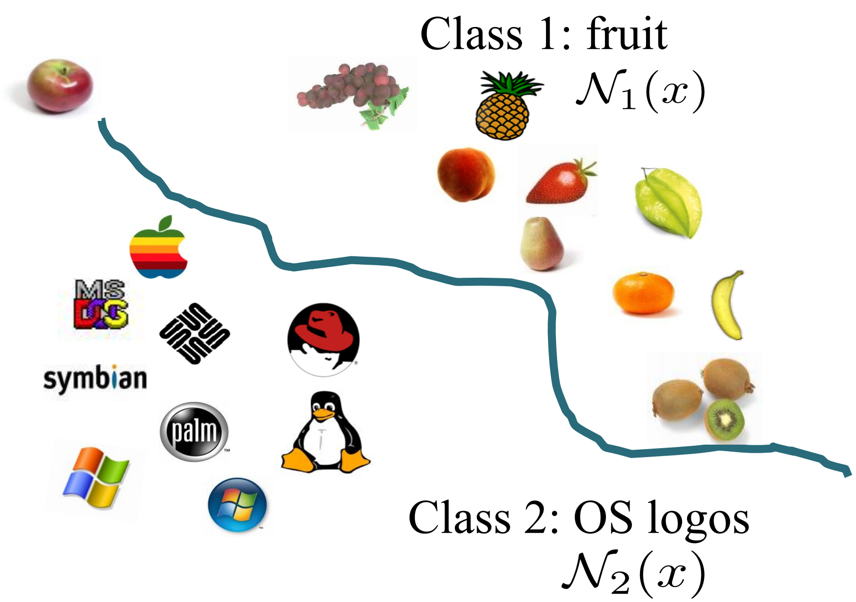
Estimating the Local SDA Parameters

$$\gamma_{gh} e^{\lambda_{gh} s(x,z)}$$

normalizing constant

solution to mean-similarity constraint

$$E_{P_h(\mathcal{T}_h(x)|Y=g)}[s(X,z)] = \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h}.$$



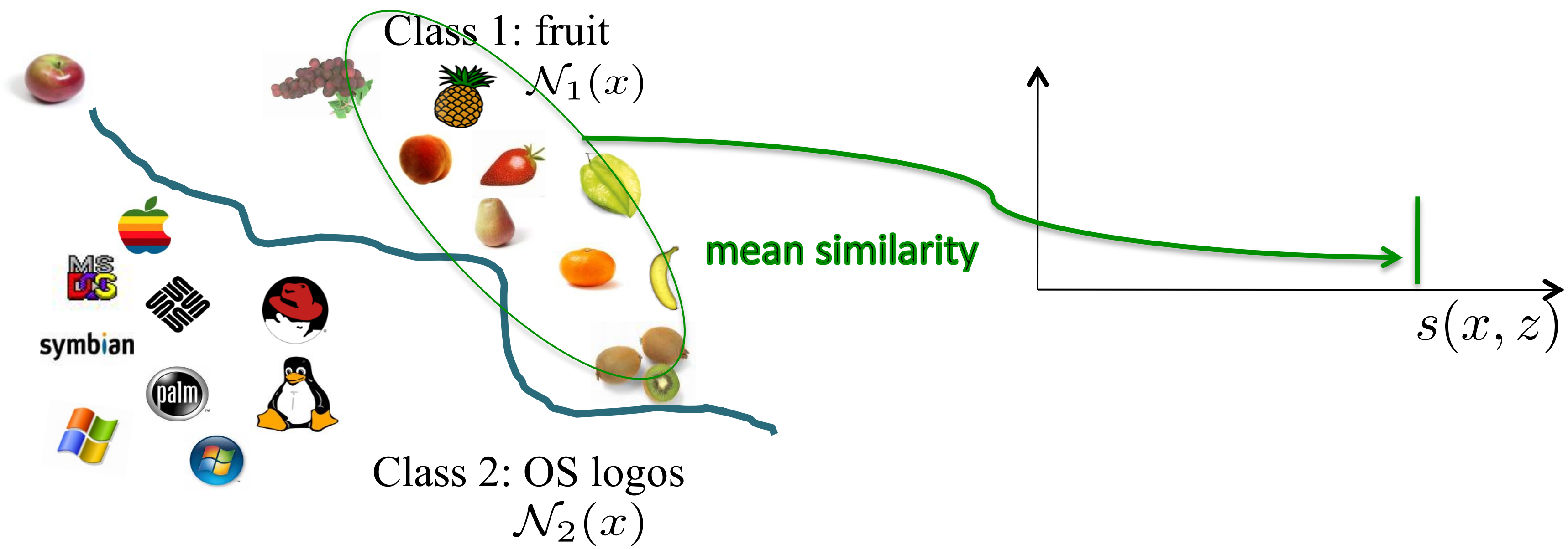
Estimating the Local SDA Parameters

$$\gamma_{gh} e^{\lambda_{gh} s(x,z)}$$

normalizing constant

solution to mean-similarity constraint

$$E_{P_h(\mathcal{T}_h(x)|Y=g)}[s(X,z)] = \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h}.$$

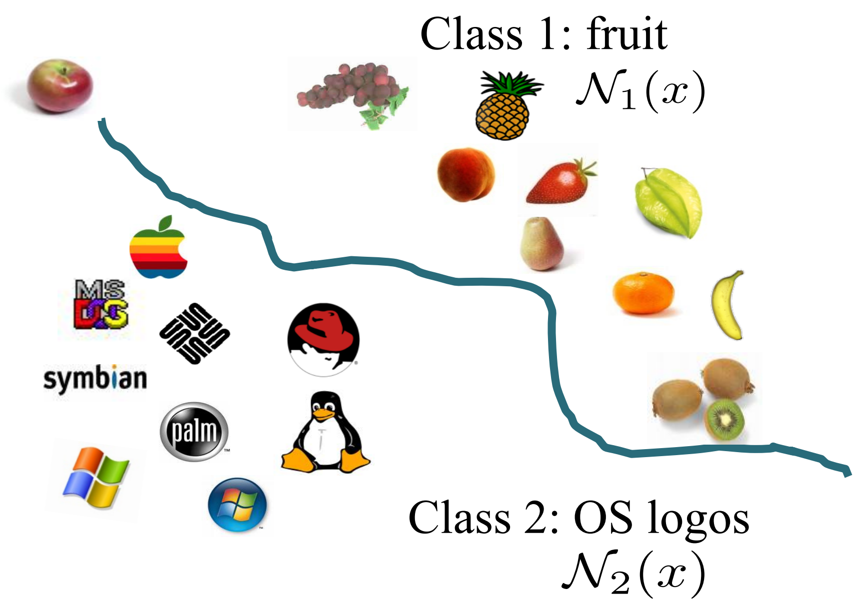


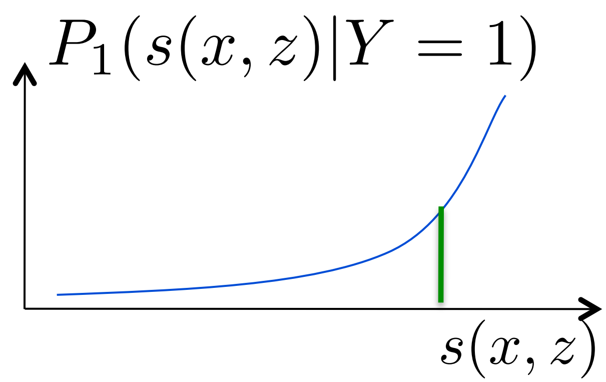
Estimating the Local SDA Parameters

normalizing constant $\longleftarrow \gamma_{gh} e^{\lambda_{gh} s(x,z)}$

solution to mean-similarity constraint

$$E_{P_h(\mathcal{T}_h(x)|Y=g)}[s(X,z)] = \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h}.$$





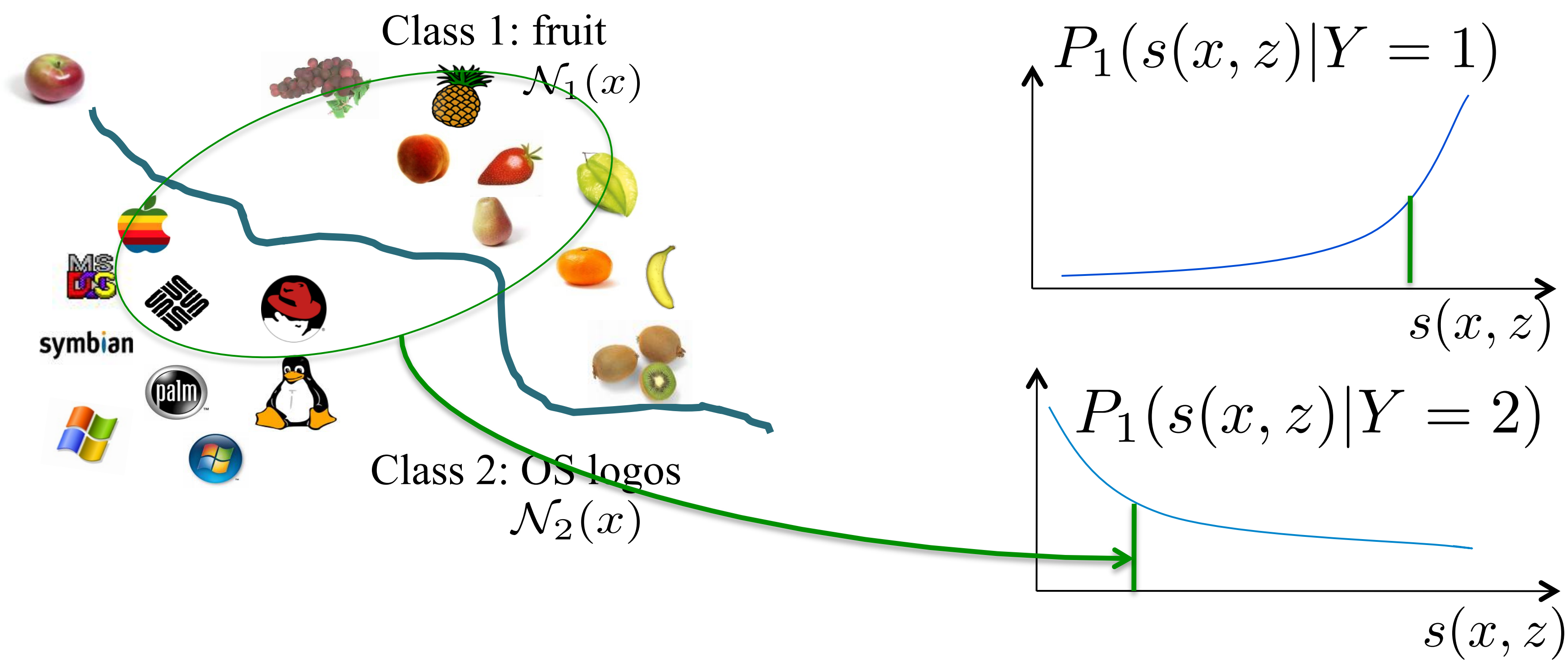
Estimating the Local SDA Parameters

$$\gamma_{gh} e^{\lambda_{gh} s(x,z)}$$

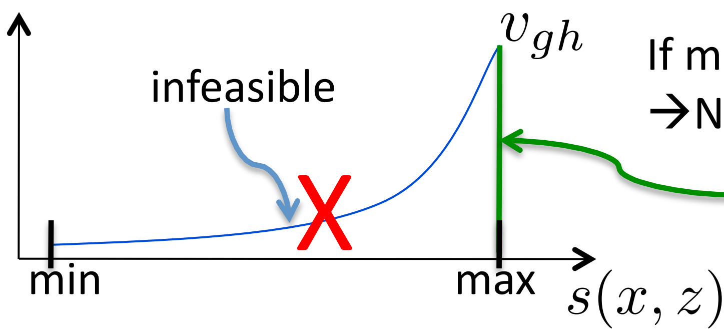
normalizing constant

solution to mean-similarity constraint

$$E_{P_h(\mathcal{T}_h(x)|Y=g)}[s(X,z)] = \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h}.$$



Need for Regularization



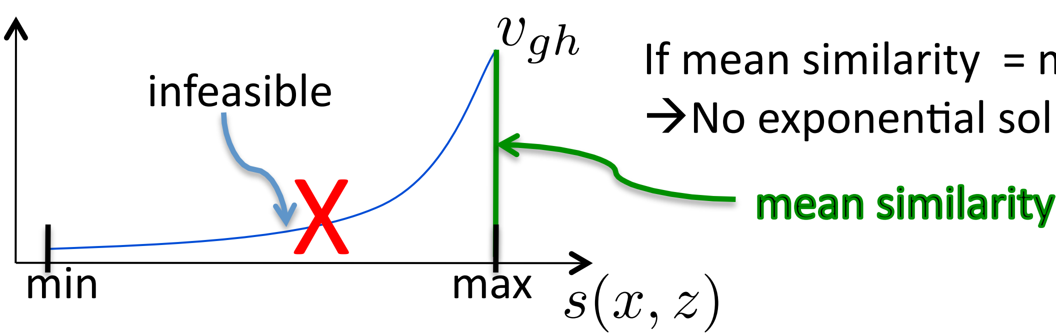
If mean similarity = maximum (or minimum) similarity
→No exponential solution exists!

mean similarity

$\boxed{\text{Approach: mean class-conditional similarities regularize each other}}$

$$v_{11} \leftrightarrow v_{12} \Longrightarrow v_{11}^*, v_{12}^*$$

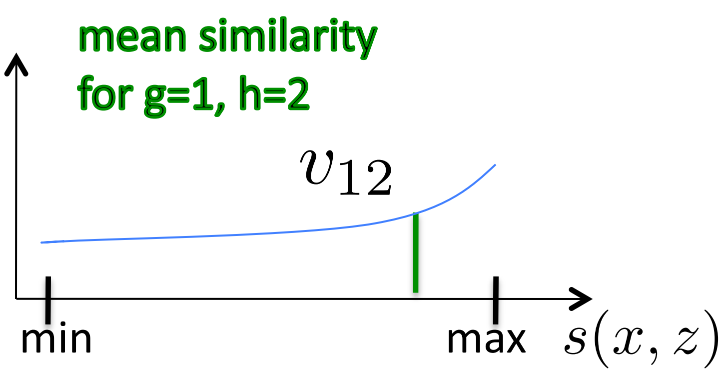
Need for Regularization

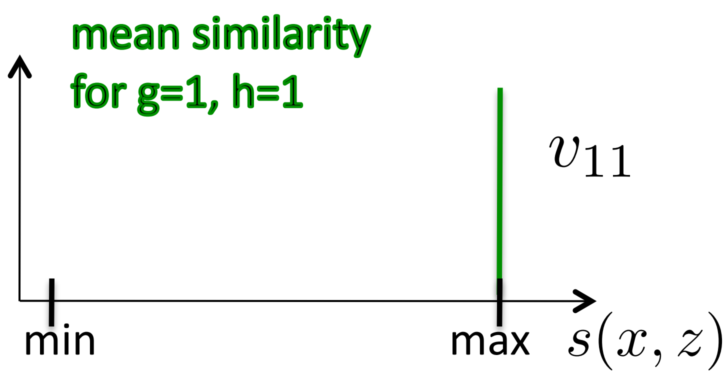


If mean similarity = maximum (or minimum) similarity
→No exponential solution exists!

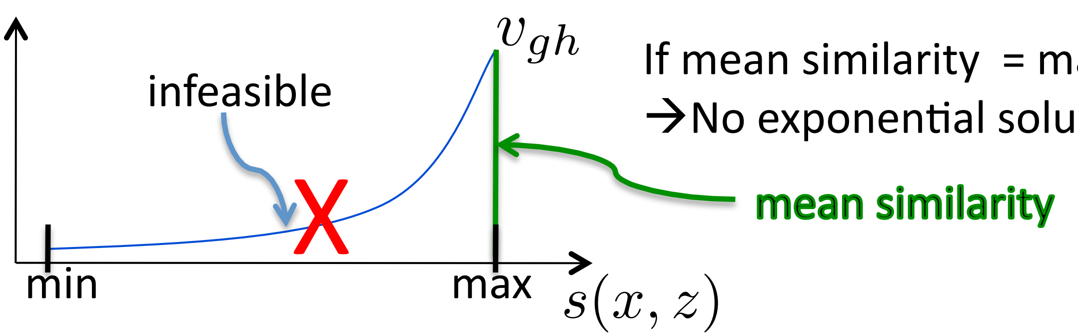
Approach: mean class-conditional similarities regularize each other

$$v_{11} \leftrightarrow v_{12} \Longrightarrow v^*_{11}, v^*_{12}$$





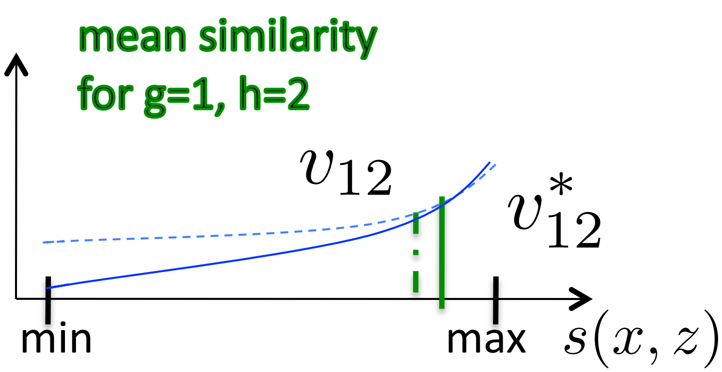
Need for Regularization

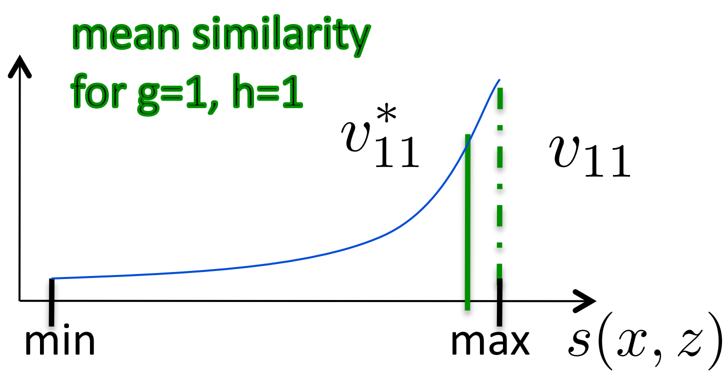


If mean similarity = maximum (or minimum) similarity
→No exponential solution exists!

Approach: mean class-conditional similarities regularize each other

$$v_{11} \leftrightarrow v_{12} \Longrightarrow v^*_{11}, v^*_{12}$$





Multi-task Regularization

Single task: estimate mean similarity v_{gh}

Multi-task:

$$\{v^*_{gh}\}_{g,h=1}^G = \arg \min_{\{\hat{v}_{gh}\}_{g,h=1}^G} \sum_{g,h=1}^G \overbrace{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} (s(z_a, z_b) - \hat{v}_{gh})^2)}^{\text{empirical mean}} +$$

$$\underbrace{\eta \sum_{j,k=1}^G \sum_{l,m=1}^G A(v_{jk}, v_{lm})(\hat{v}_{jk} - \hat{v}_{lm})^2}_{\text{regularizing term}}.$$

Multi-task Regularization

Single task: estimate mean similarity v_{gh}

Multi-task:

$$\{v^*_{gh}\}_{g,h=1}^G = \arg \min_{\{\hat{v}_{gh}\}_{g,h=1}^G} \sum_{g,h=1}^G \sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} (s(z_a, z_b) - \hat{v}_{gh})^2) +$$

$$\eta \sum_{j,k=1}^G \sum_{l,m=1}^G A(v_{jk}, v_{lm})(\hat{v}_{jk} - \hat{v}_{lm})^2.$$

Controls how much to regularize

G^2-by-G^2 task-relatedness matrix

Multi-task Regularization – Closed Form Solution

For A symmetric and invertible:

$$v^* = (I - \tilde{A})^{-1} \tilde{v}\,,$$

$$\begin{aligned}
\tilde{v}_{gh} &= \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h + \eta \sum_{l,m \neq g,h} A(v_{gh}, v_{lm})} \text{ and} \\
\tilde{A}(v_{gh}, v_{lm}) &= \begin{cases} \frac{\eta A(v_{gh}, v_{lm})}{k_g k_h + \eta \sum_{g,h \neq l,m} A(v_{gh}, v_{lm})} & \text{for} \{g,h\} \neq \{j,k\} \\ 0 & \text{for} \{g,h\} = \{j,k\} \end{cases}
\end{aligned}$$

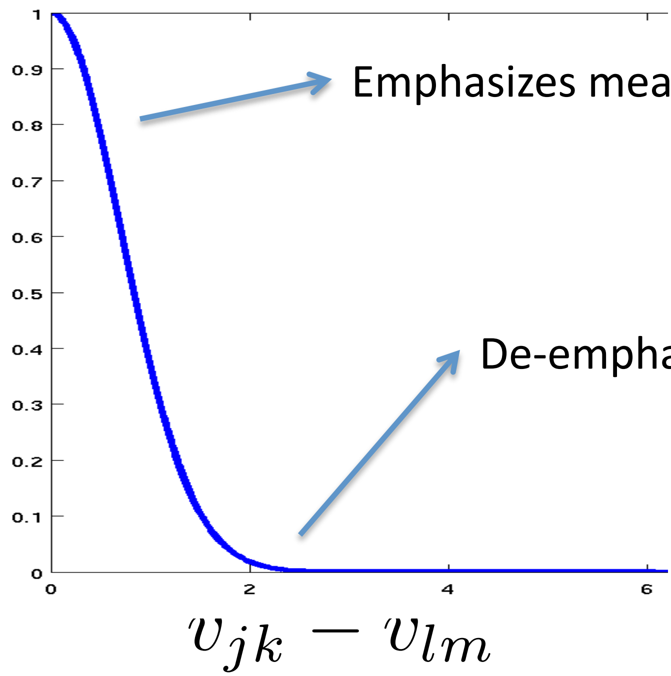
Then solve the G^2 regularized mean-similarity constraints:

$$E_{P_h(\mathcal{T}_h(x)|Y=g)}[s(X,z)] = v^*_{gh}$$

$$\rightarrow \lambda^*_{gh}$$

Choice of Task Relatedness Matrix A

Symmetric and invertible → $A(v_{jk}, v_{lm}) = e^{-(v_{jk} - v_{lm})^2/\sigma}$

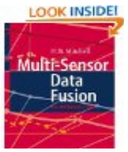


Emphasizes mean similarities close to each other

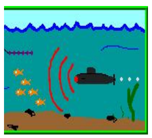
De-emphasizes distant mean similarities

Can use any problem-relevant task relatedness.
Side information easily incorporated into problem.

Benchmark Datasets



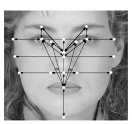
AMAZON (fiction & nonfiction): similarities between books based on user statistics from **`amazon.com`**



SONAR (target & clutter): similarities between sonar signals rated by human subjects.

PATROL (8 patrol units): membership in patrol unit reported by other patrol members.

VOTING (2 political parties): value difference metric on congressional votes.



FACE RECOGNITION (139 faces): cosine similarity between features from 3D face data.

Benchmark Datasets

	Amazon 2 classes	Sonar 2 classes	Patrol 8 classes	Protein 4 classes	Voting 2 classes	FaceRec 139 classes
Multi-task Local SDA	**8.95**	14.50	**11.56**	**9.77**	5.52	**3.44**
Local SDA	11.32	15.25	**11.56**	10.00	6.15	4.23
Similarity k-NN	12.11	15.75	19.48	30.00	5.69	4.29
SVM-KNN (sims-as-features)	13.68	**13.00**	14.58	29.65	**5.40**	4.23

Percent test error averaged over 20 random train/test splits.
RBF task relatedness for multi-task local SDA

Multi-task local SDA at least as good as local SDA.

Benchmark Datasets

	Amazon 2 classes	Sonar 2 classes	Patrol 8 classes	Protein 4 classes	Voting 2 classes	FaceRec 139 classes
Multi-task Local SDA	**8.95**	14.50	**11.56**	**9.77**	5.52	**3.44**
Local SDA	11.32	15.25	**11.56**	10.00	6.15	4.23
Similarity k-NN	12.11	15.75	19.48	30.00	5.69	4.29
SVM-KNN (sims-as-features)	13.68	**13.00**	14.58	29.65	**5.40**	4.23

Percent error averaged over 20 random train/test splits.
RBF task relatedness for multi-task local SDA

Multi-task local SDA at least as good as local SDA.
Multi-task local SDA competitive with other similarity-based classifiers.

Insurgent Rhetoric Experiment

1924 documents (press releases)

Which of 8 Iraqi insurgent groups authored the document?

Document similarity: KL divergence of pmfs over 173 keywords

Num. docs jointly released by groups (j,k)

Num. docs jointly released by groups (l,m)

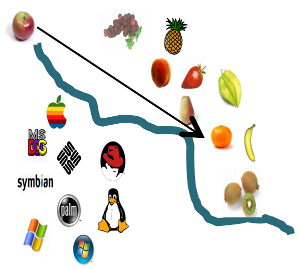
$$A(v_{jk}, v_{lm}) = e^{-(Q_{jk}-Q_{lm})^2/\sigma}$$

Multi-task Local SDA (w/ joint statements task relatedness)	**52.34**
Multi-task Local SDA (w/ Gaussian kernel task relatedness)	52.75
Local SDA	54.52
Similarity k-NN	53.53
Guessing Using Class Priors	77.91

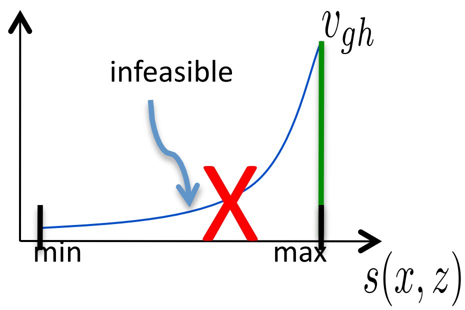
Leave-one-out cross validation error

Summary

Reviewed local SDA



Need for regularization



Multi-task regularization

$$\{v^*_{gh}\}_{g,h=1}^G = \arg \min_{\{\hat{v}_{gh}\}_{g,h=1}^G} \sum_{g,h=1}^G \sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} (s(z_a, z_b) - \hat{v}_{gh})^2) + \eta \sum_{j,k=1}^G \sum_{l,m=1}^G A(v_{jk}, v_{lm})(\hat{v}_{jk} - \hat{v}_{lm})^2.$$

Illustrated different choices for task relatedness A with benchmark and real datasets.

Submitted to JMLR: "Multi-Task Output Space Regularization," S. Feldman, B. A. Frigyik, M. R. Gupta, L. Cazzanti, P. Sadowski, available at http://arxiv.org/abs/1107.4390.

Software and data available: http://staff.washington.edu/lucagc

To Learn More

"Bayesian and Pairwise Local Similarity Discriminant Analysis," P. Sadowski, L. Cazzanti and M. R. Gupta, Proc. Intl. Workshop on Cognitive Information Processing (CIP), Isola d'Elba, Italy, June 2010.

"Regularizing the Local Similarity Discriminant Analysis Classifier," L. Cazzanti and M. R. Gupta, Proc. Intl. Conf. on Machine Learning and Applications (ICMLA), Miami Beach, December 2009.

“Fusing Similarities and Euclidean Features with Generative Classifiers," L. Cazzanti, M.R. Gupta, and S. Srivastava, Proc. Intl. Conf. on Information Fusion (FUSION), Seattle, July, 2009.

"Similarity-based Classification: Concepts and Algorithms," Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi, and L. Cazzanti, Journal of Machine Learning Research, March 2009.

“Generative Models for Similarity-Based Classification," L. Cazzanti, M. R. Gupta, and A. J. Koppal, Pattern Recognition, vol. 41, no. 7, 2289-2297, 2008.

"Local Similarity Discriminant Analysis," L. Cazzanti and M. R. Gupta, Intl. Conf. Machine Learning (ICML), 2007.