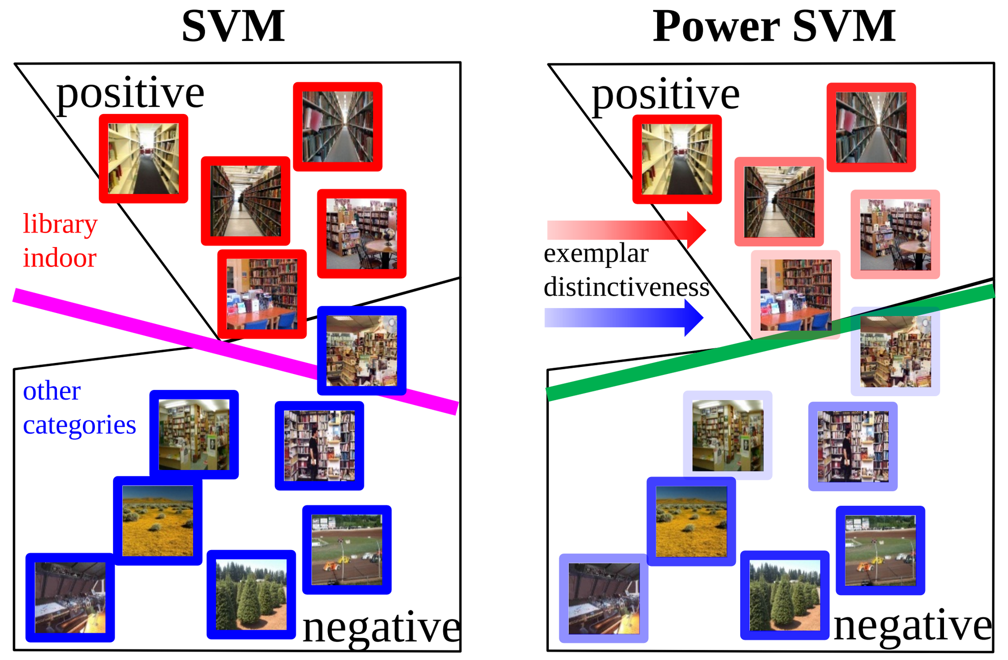
Power SVM: Generalization with Exemplar Classification Uncertainty

Weiyu Zhang — University of Pennsylvania
Stella X. Yu — UC Berkeley / ICSI
Shang-Hua Teng — University of Southern California

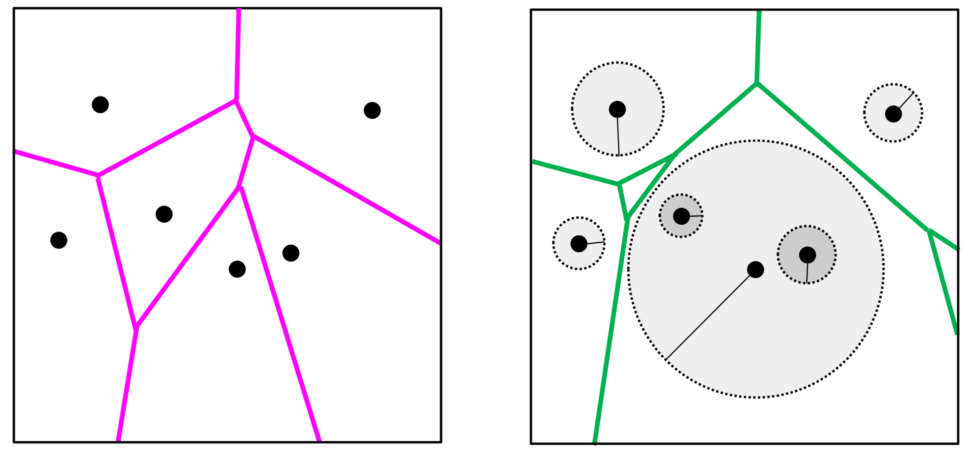
Motivation: Visual generalization is better with distinctive exemplars

1. Human vision recognizes more variants of a distinctive exemplar.
2. An exemplar's generalization capacity is relative in terms of what it is discriminated against.

Basic Idea: Distinctive exemplars lie further away from the desired decision boundary



1. Exemplars are treated equally in SVM, but differently in Power SVM.
2. By acknowledging the distinction in exemplar discrimination capacity, Power SVM generalizes better from fewer exemplars.
3. SVM :: Power SVM = Voronoi Diagram :: Power Diagram

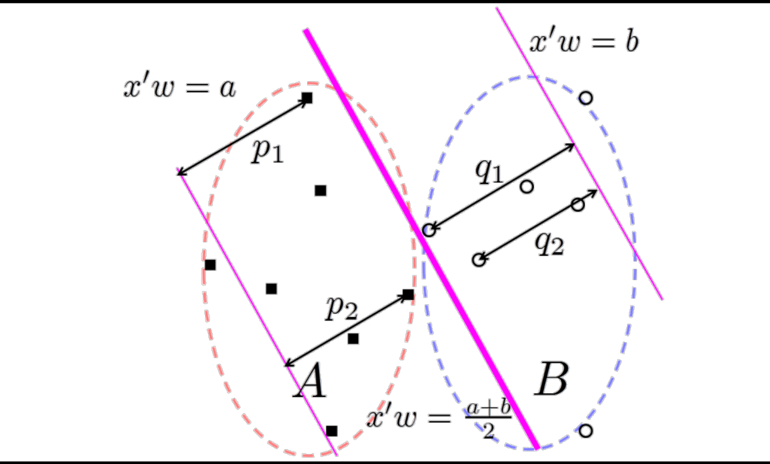
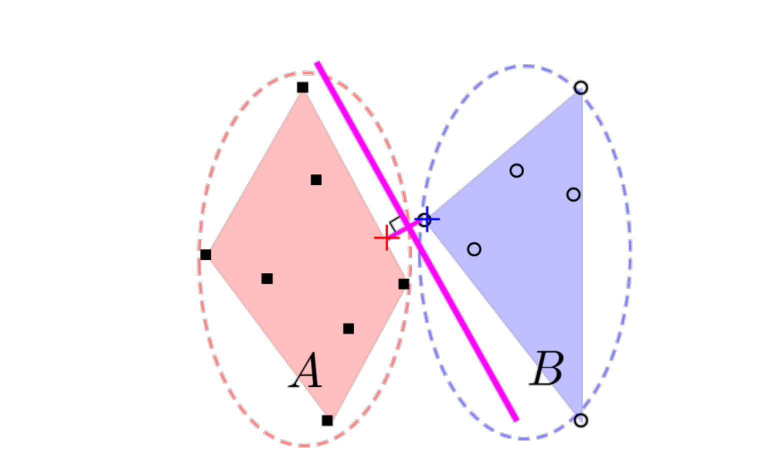
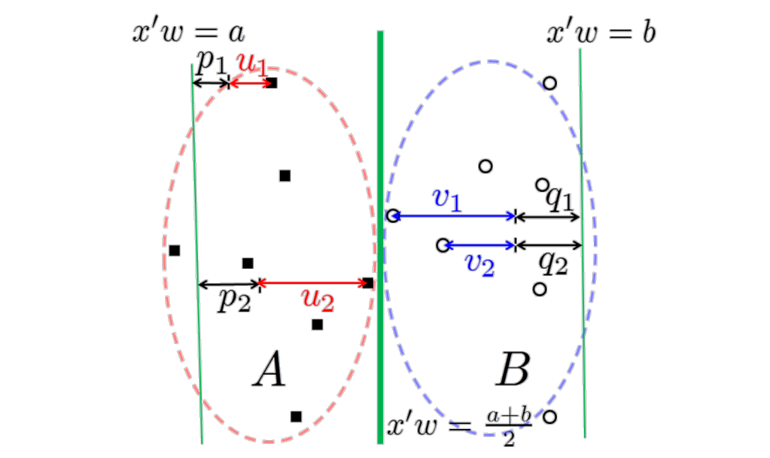
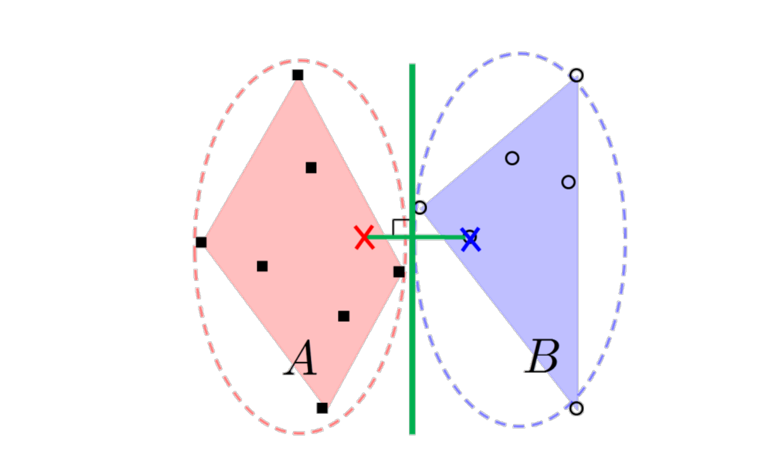


Power SVM: Seek a global classifier constrained by individual exemplar uncertainty

Known : (feature x_i, label y_i, uncertainty u_i) for exemplar i in positive and negative classes.

Solve : binary classifier represented by two parallel bounding planes (normal w, offsets a, b).

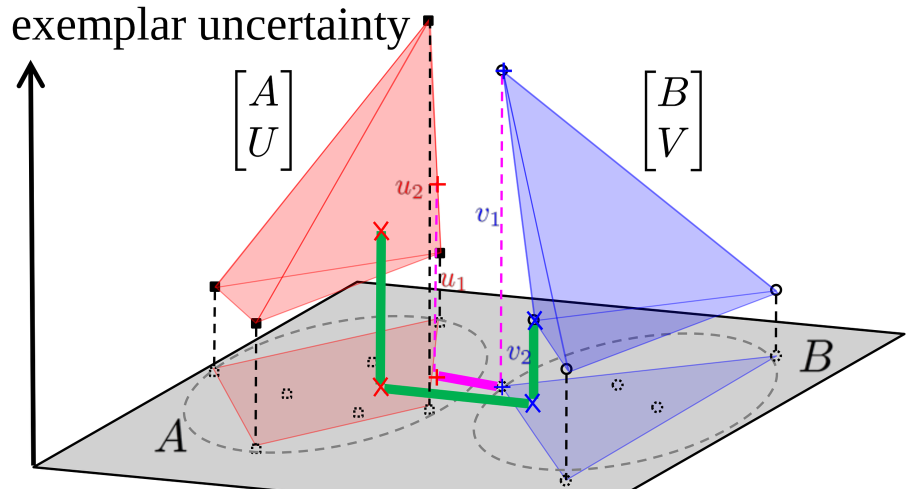
Matrix Notation: positive features : $A_{d\times m} = [x_1, \dots, x_m]$, negative features : $B_{d\times n} = [x_{m+1}, \dots, x_{m+n}]$, A's uncertainty : $U_{m\times 1} = [u_1, \dots, u_m]'$, B's uncertainty : $V_{n\times 1} = [u_{m+1}, \dots, u_{m+n}]'$.

	Primal	Dual
Goal	Find parallel bounding planes of maximal separation	Find the shortest path between convex hulls
SVM		
Criterion	SVM, when $U = V = 0$	
	$\min_{w,a,b,p,q} \ \frac{1}{2}w'w - (a-b) + D(p'1_m + q'1_n)$ s. t. $A'w \boxed{+U} \ge a1_m - p,\ p \ge 0_m$; $B'w \boxed{-V} \le b1_n + q,\ q \ge 0_n$	$\min_{\alpha,\beta} \ \frac{1}{2}\Vert A\alpha - B\beta\Vert^2 + \boxed{(U'\alpha + V'\beta)}$ s. t. $\alpha'1_m = 1,\ \beta'1_n = 1$, $0_m \le \alpha \le D1_m,\ 0_n \le \beta \le D1_n$.
	Power SVM, for general U, V	
Power SVM		

SVM seeks the **shortest path** between 2D convex hulls in the original feature space.

Power SVM seeks the **shortest path** between 3D convex hulls in the augmented feature space.

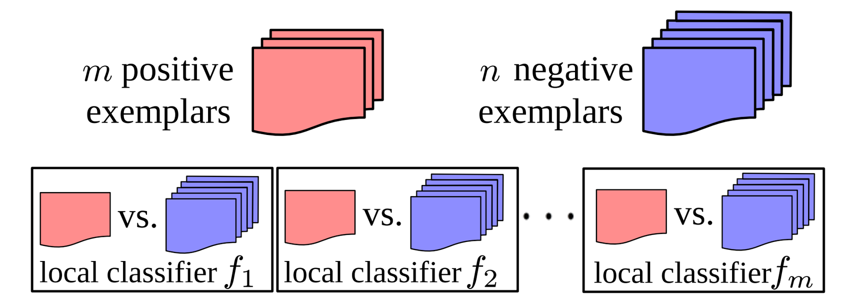
path cost = $\underbrace{\frac{1}{2}\Vert A\alpha - B\beta\Vert^2}_{\text{horizontal paths}} + \underbrace{(U'\alpha + V'\beta)}_{\text{vertical paths}}$



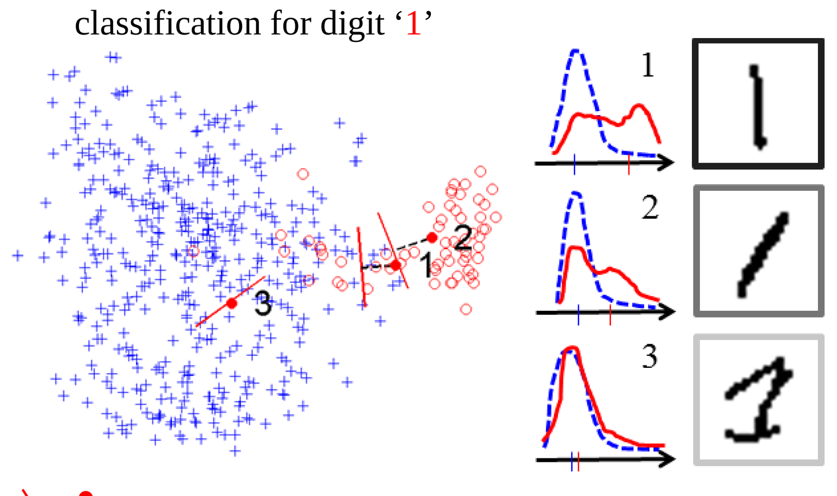
Data-driven exemplar uncertainty

- Classification uncertainty u_i indicates how easily an exemplar might be confused with those from the opposite class.
- We can obtain informative estimate of how discriminative an exemplar is without knowing the desired classifier in advance.
- $u_i \neq$ data uncertainty.

1. Learn exemplar-centric local classifiers



2. Derive positive exemplars' uncertainty



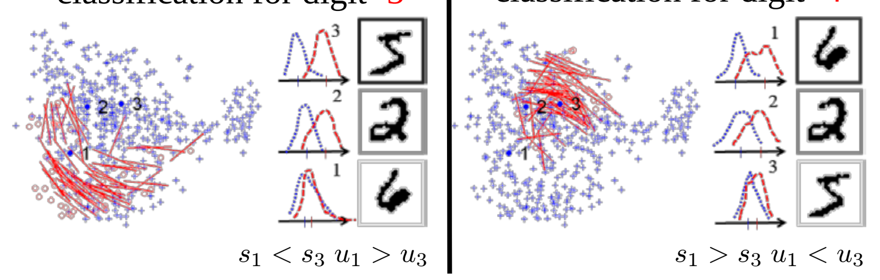
local classifier f_i of exemplar x_i $\quad s_1 > s_3 \quad u_1 < u_3$

s_i : difference between mean positive/negative response of f_i

$$s_i = \sum_{t=1}^{m} \frac{f_i(x_t)}{m} - \sum_{t=m+1}^{m+n} \frac{f_i(x_t)}{n}, \qquad u_i \propto \max_{1\le t\le m} s_t - s_i$$

A large s_i (small u_i) means its local classifier separates the entire positive class from the negative class.

3. Derive negative exemplars' uncertainty



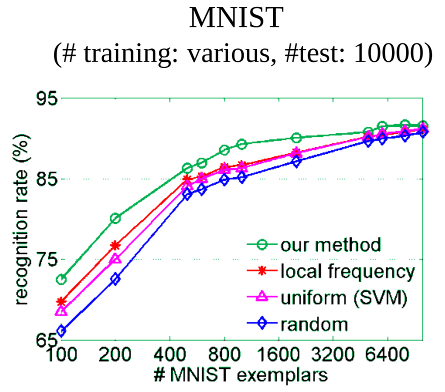
s_j: mean diff over all f_i btw. their response on positive class and x_j

$$s_j = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{t=1}^{m}\frac{f_i(x_t)}{m} - f_i(x_j)\right), u_j \propto \max_{m+1\le t\le m+n} s_t - s_j$$

A large s_j (small u_j) means all local classifiers tend to put x_j in the negative class.
Same negative exemplars may assume different uncertainties.

Experiments: Power SVM with exemplar uncertainty in multiclass visual discrimination

Our classification uncertainty is more informative

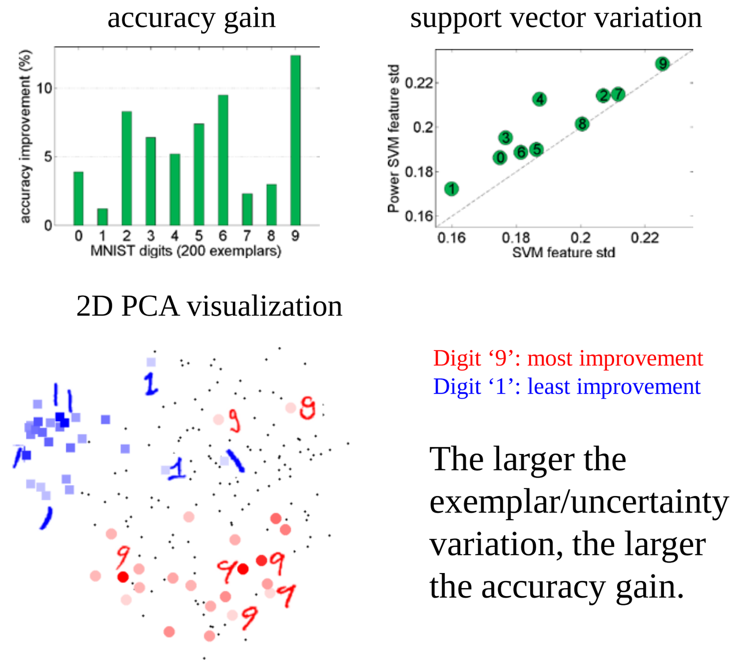


Indoor-outdoor categorization accuracy (%)
(# training 100, # test 1000)

uncertainty type	GIST	Sparse SIFT	HOG
our method	**81.9**	83.4	**89.0**
human	80.2	**84.5**	87.1
local frequency	81.0	83.3	88.0
uniform (SVM)	77.6	82.9	87.3
random	75.2	81.2	85.2

1. Our classification uncertainty outperforms local frequency and uniform uncertainty (SVM) in various feature spaces.
2. Our uncertainty outperforms human accuracy since it is tuned to the feature specific for final classification and has finer exemplar separation.
3. The benefit of good uncertainty diminishes as # exemplars increases.

due to larger support vector variation



Digit '9': most improvement
Digit '1': least improvement

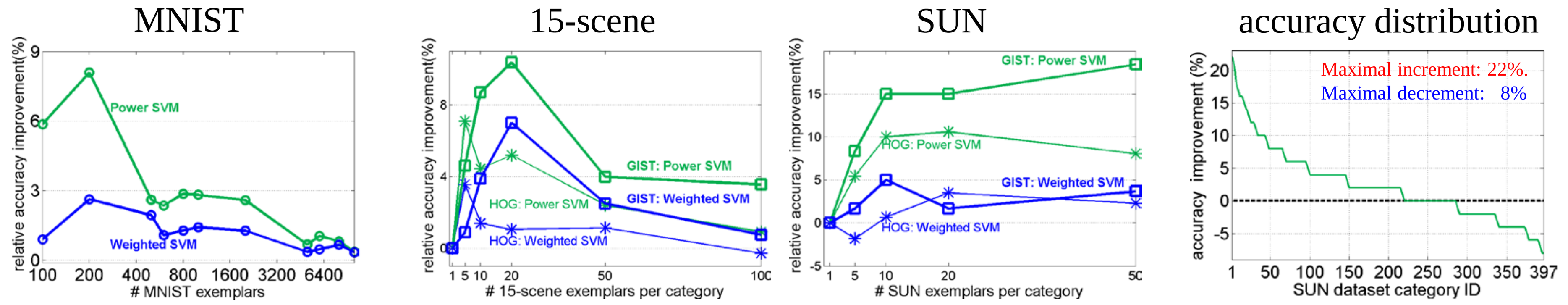
The larger the exemplar/uncertainty variation, the larger the accuracy gain.

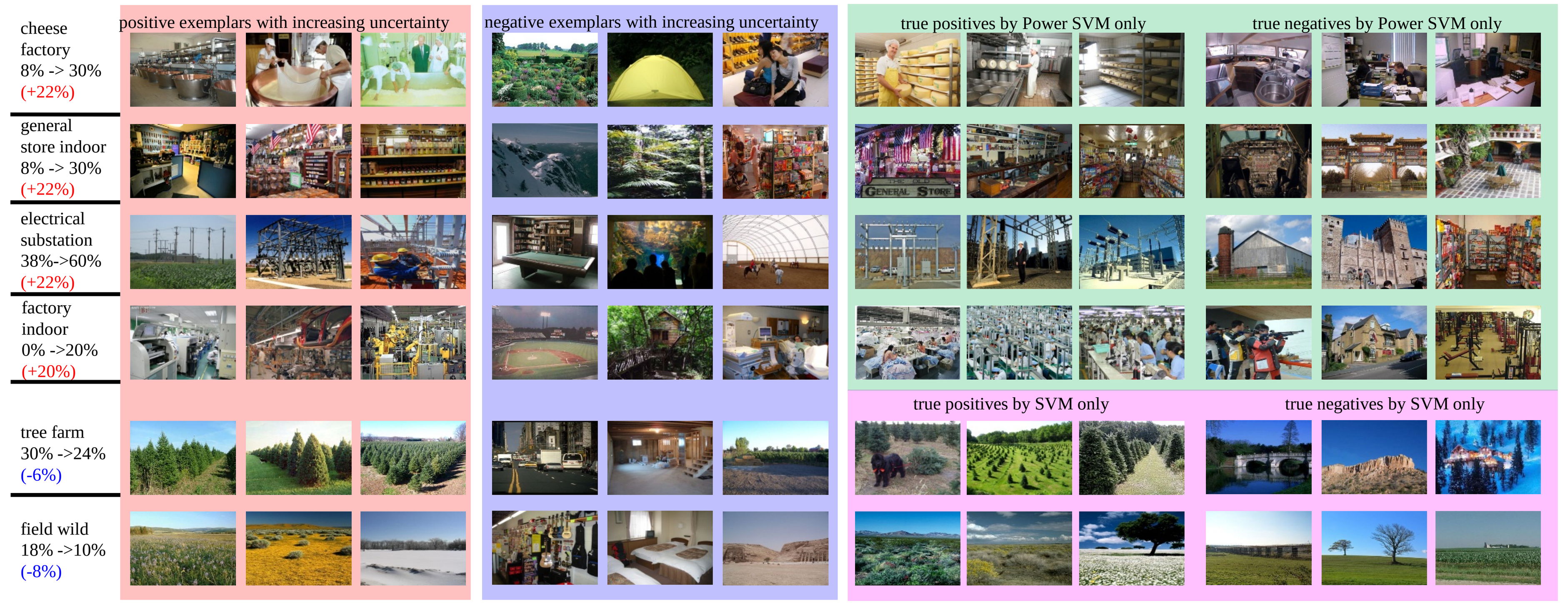
Power SVM is more effective than Weighted SVM at utilizing exemplar uncertainty

Weighted SVM

$$\min_{w,t,p,q} \ \varepsilon = \frac{1}{2}w'w + C(p'(1_m - U) + q'(1_n - V))$$
$$\text{s. t.} \ A'w - t \ge 1_m - p, \quad p \ge 0_m, \qquad B'w - t \le -1_n + q, \quad q \ge 0_n.$$

Theoretical distinction. In the **Primal**, Weighted SVM duplicates higher certainty exemplars (weak impact),while Power SVM pushes higher certainty exemplars away from the decision boundary. In the **Dual**, Weighted SVM changes the shape of convex hulls (often slightly), while Power SVM modifies the distance measure for the shortest path directly.





1. Indoor scenes with larger intra-category variation get bigger improvement than outdoor natural scene with smaller variation.
2. Positive exemplars of larger uncertainty often have extreme lighting or small fields of view.

This research is funded by NSF CAREER IIS-0644204 and by Boston College to Stella Yu, and by NSF CCF 1111270 to Teng.