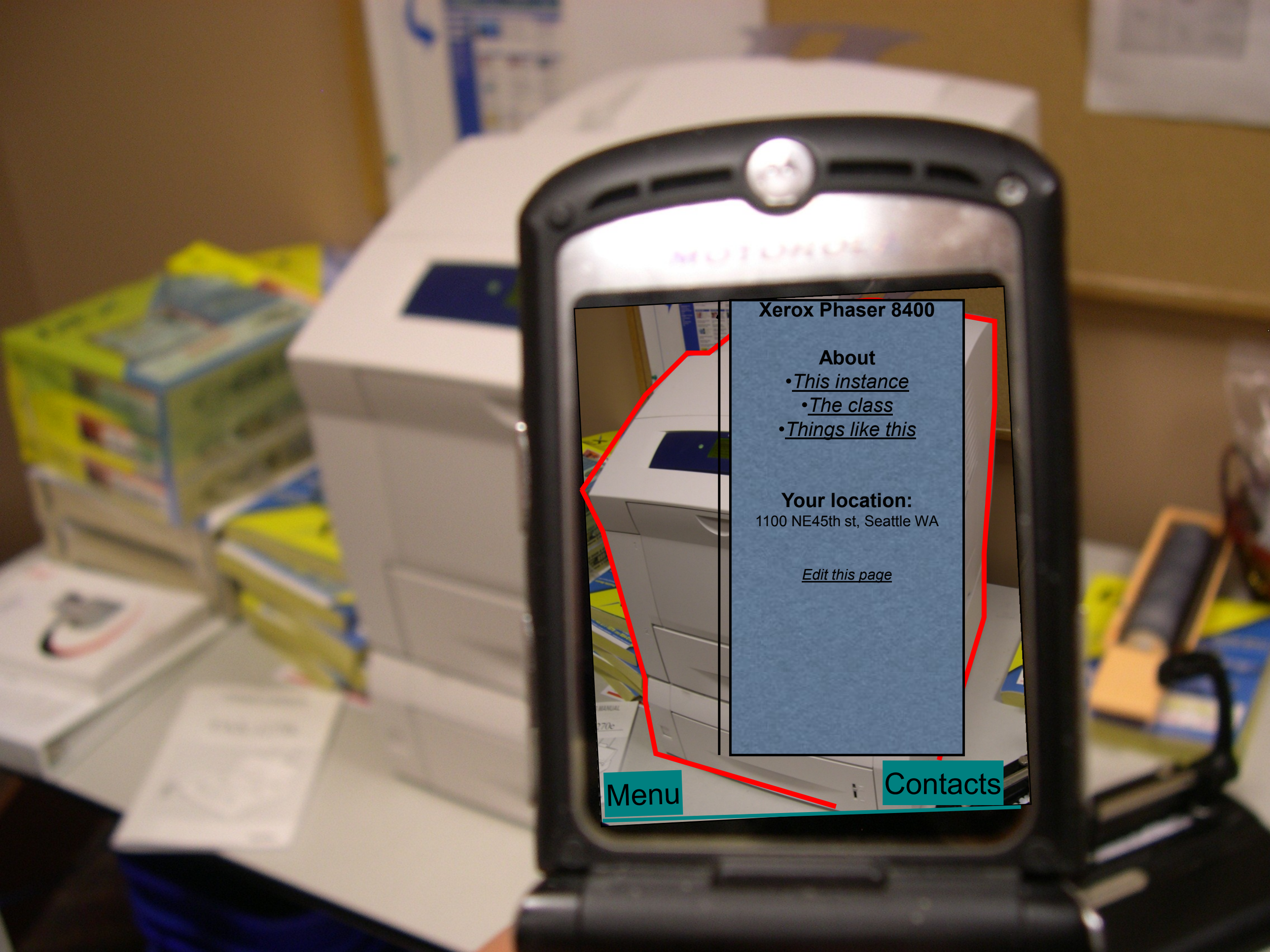
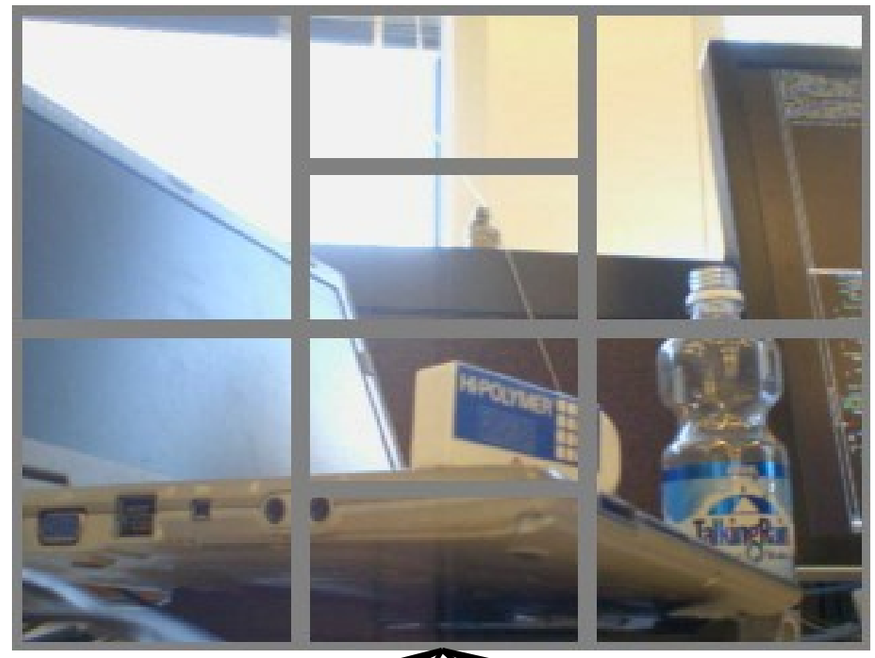
Generalization Bounds for Indefinite Kernel Machines

Shai Ben-David, Ali Rahimi, Nati Srebro


Xerox Phaser 8400
About
•This instance
•The class
•Things like this
Your location:
1100 NE45th st, Seattle WA
Edit this page
Menu
Contacts

Input image

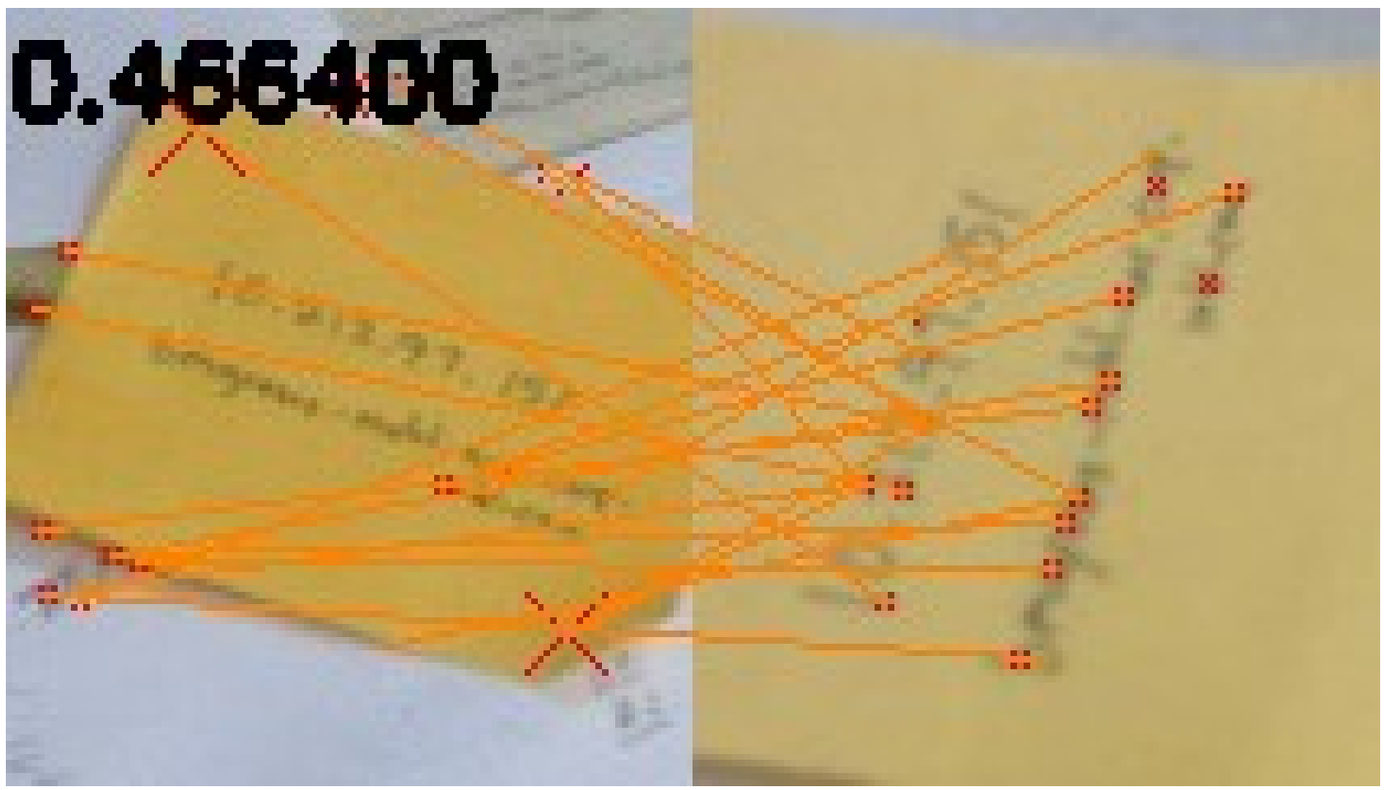


Compute similarities with training images

. . .

Some kind of classifier

x x_i



$k(x, x_i)$

K is not...

- Positive definite
- A metric
- A function on R^n x R^n
- Differentiable

Algorithm: When K is pd, use an SVM

$$h(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i)$$

$$\begin{array}{ll} \min_{\alpha} & \sum_{i=1}^{n} \ell(h(x_i), y_i) \\ \text{s.t.} & \alpha^\top \mathbf{K} \alpha \leq C \end{array}$$

Generalization when K is pd

$$\hat{\mathcal{L}}_S[h] = \frac{1}{n}\sum_{i=1}^{n} \ell(h(x_i), y_i) \qquad \mathcal{L}[h] = \mathbb{E}\left[\ell(h(X), Y)\right]$$

$$\Pr_S\left[\sup_{f \in \mathcal{H}_S} \mathcal{L}[h] - \hat{\mathcal{L}}_S[h] > \epsilon\right]$$

$$\begin{aligned}
\mathcal{H}_S &\equiv \left\{ x \to \sum_{i=1}^{n} \alpha_i K(x, x_i) \mid \alpha^\top \mathbf{K} \alpha \leq C \right\} \\
&\subseteq \left\{ x \to \langle w, \phi(x) \rangle \;\middle|\; \|w\| \leq C \right\} \\
&\equiv \mathcal{H}
\end{aligned}$$

Analyze the generalization of a class of linear separat

Algorithm: When K is not pd

$$h(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i)$$

$$\begin{array}{ll} \min\limits_{\alpha} & \sum_{i=1}^{n} \ell(h(x_i), y_i) \\ \text{s.t.} & \|\alpha\|_1 \leq C \end{array}$$

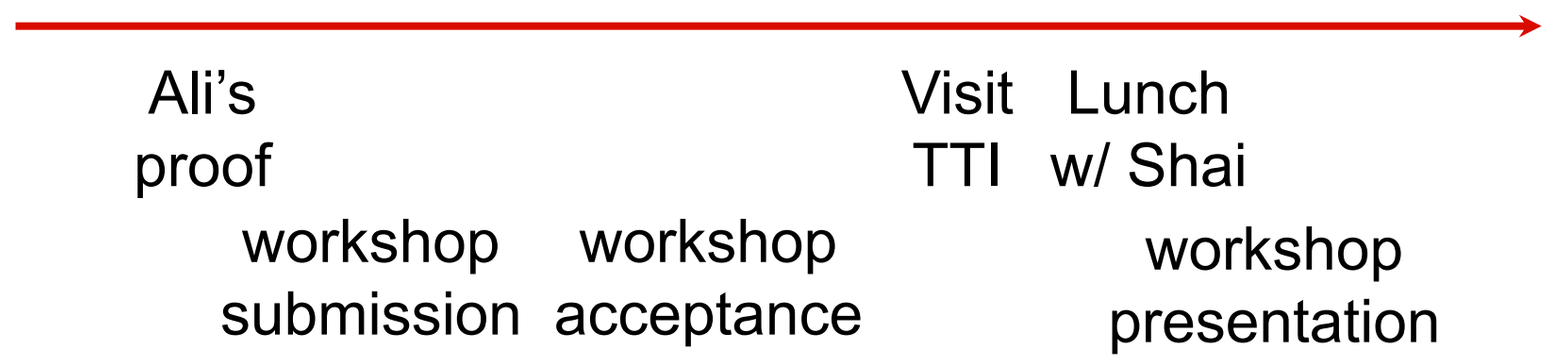
Note that $\alpha^\top \mathbf{K} \alpha \leq \|\alpha\|_1^2 \leq C$ so when $\mathbf{K}$ is pd, this generalizes.

Algorithm: With prototype Selection

$$h(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i)$$

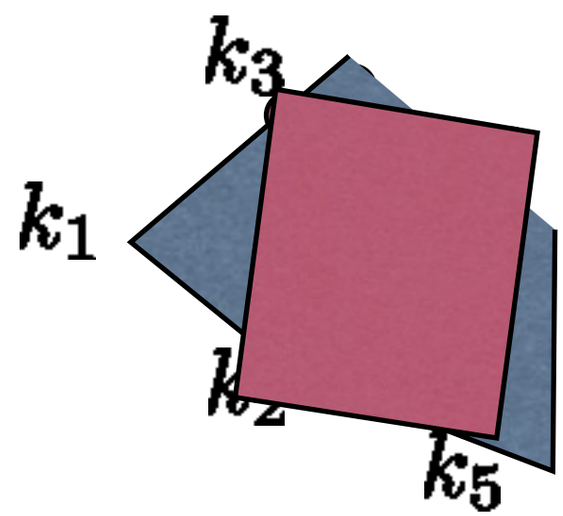
$$\begin{aligned} \min_{\alpha} \quad & \textstyle\sum_{i=1}^{n} \ell(h(x_i), y_i) \\ \text{s.t.} \quad & \|\alpha\|_1 \leq C \\ & \alpha_j = 0, \quad j \in \mathcal{J}(x_1, y_1, \ldots, x_n, y_n) \end{aligned}$$

Generalization Bounds when K is indefinite



Ali's Original Proof Idea

$$\Pr_{S}\left[\sup_{f\in\mathcal{H}_S}\mathcal{L}[h]-\hat{\mathcal{L}}_S[h]>\epsilon\right]\approx\Pr_{S,S'}\left[\sup_{f\in\mathcal{H}_{S'}}\mathcal{L}[h]-\hat{\mathcal{L}}_{S'}[h]>\epsilon\right]$$



Theorem: Let $k_1,\ldots,k_n,k_1',\ldots,k_n'$ be random variables drawn iid from a distribution on a separable Banach space $\mathcal{K}$, with norm bounded above by C. Let $\mathbf{B}_1'$ denote a unit ball in the dual space of $\mathcal{K}$. Let $\mathcal{H}_S$ and $\mathcal{H}_{S'}$ be the absolute convex hulls of $k_1\ldots k_n$ and $k_1'\ldots k_n'$ respectively. Then for any $0<\epsilon<2C$,

$$\Pr\left[\sup_{h\in\mathcal{H}_S}\inf_{h'\in\mathcal{H}_{S'}}\|h-h'\|>\epsilon\right]\leq\mathcal{N}\left(\mathbf{B}_1',\tfrac{\epsilon}{4C}\right)\cdot n\cdot\exp\left(-\frac{n\epsilon}{8C}\right).$$