

---

# Nightmare at Test Time: Robust Learning by Feature Deletion

---

-- Amir Globerson and Sam Roweis

ICML'06

---

# Why Robust Learning?

- Non-stationary feature distribution for training and test data
  - Small Samples/ Imbalanced Class distribution
  - Adversarial classification (Spam filtering)
  - Data with Uncertainty
  - A specific situation: A feature presented at training data but disappear (change to 0) in test data
-

---

# Intuition of Robust Learning

- 3 stocks  $A$ ,  $B$ ,  $C$  with the same risk
  - If you are going to invest \$3000 on stocks.
  - Strategy 1: \$3000  $\rightarrow$   $A$
  - Strategy 2: \$1000  $\rightarrow$   $A$ , \$1000  $\rightarrow$   $B$ , \$1000  $\rightarrow$   $C$
  - Which one to choose?
  - *Do not assign any feature with too much weight. (Regularization term like  $|w|^2$  ??)*
-

---

# Game Theory (Min-Max)

- Consider an adversarial situation:
  - Two Players:
  - P1: Build Classifier
  - P2: Delete features during testing
  
  - What's P1's policy?
  - --Maximize the worst performance
-

---

For each instance  $\mathbf{x}_i$ , the worst case hinge loss is:

$$h^{wc}(\mathbf{w}, y_i \mathbf{x}_i) = \max_{\substack{s.t. \quad \alpha_i \in \{0, 1\} \\ \sum_j \alpha_{ij} = K}} [1 - y_i \mathbf{w} \cdot (\mathbf{x}_i \circ (1 - \boldsymbol{\alpha}_i))]_+$$

For the whole data set,  $\mathbf{w}$  should be

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i h^{wc}(\mathbf{w}, y_i \mathbf{x}_i)$$

---

---

$$h^{wc}(\mathbf{w}, y_i \mathbf{x}_i) = [1 - y_i \mathbf{w}^T \mathbf{x}_i + s_i]_+ ,$$

where we have defined

$$s_i = \max_{\alpha_i \in \{0,1\}, \sum \alpha_{ij} = K} y_i \mathbf{w} \cdot (\mathbf{x}_i \circ \boldsymbol{\alpha}_i)$$

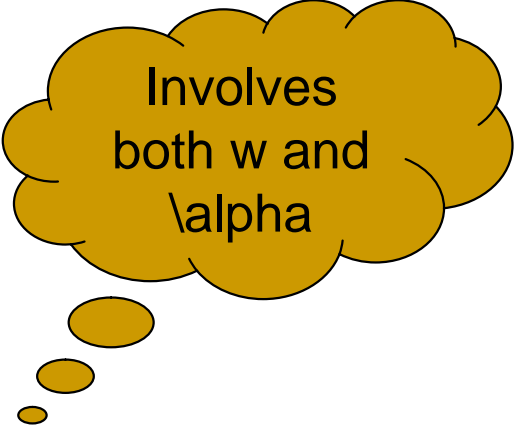
**Solution:**

**choose those features with maximal  $y_i \mathbf{w} \cdot \mathbf{x}_{ij}$**

**The solution won't change if we relax  $\alpha$  to be  $[0, 1]$**

---

# New formulation



Involves both  $w$  and  $\alpha$

$$\begin{aligned} s_i = \quad & \max && y_i (\mathbf{w} \circ \mathbf{x}_i) \cdot \boldsymbol{\alpha}_i \\ & s.t. && 0 \leq \boldsymbol{\alpha}_i \leq 1 \\ & && \sum_j \alpha_{ij} = K \end{aligned}$$

## Dual Form:

$$\begin{aligned} s_i = \quad & \min && K z_i + \sum_j v_{ij} \\ & s.t. && z_i + \mathbf{v}_i \geq (y_i \mathbf{x}_i \circ \mathbf{w}) \\ & && \mathbf{v}_i \geq 0 \end{aligned}$$

---

$$\begin{array}{ll} \min & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i [1 - y_i \mathbf{w}^T \mathbf{x}_i + t_i]_+ \\ \text{s.t.} & t_i \geq K z_i + \sum_j v_{ij} \\ & \mathbf{v}_i \geq \mathbf{0} \\ & z_i + \mathbf{v}_i \geq (y_i \mathbf{x}_i \circ \mathbf{w}) \end{array}$$

---

---

# Disssusion

- New Problem? Spam Filtering?
  - Robust Learning favors keeping all the redundant features, how to run feature selection under robust learning scenario?
-

This document was created with Win2PDF available at <http://www.win2pdf.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.  
This page will not be added after purchasing Win2PDF.