#### SVM-Relation to Logistic Regression

srihari@buffalo.edu

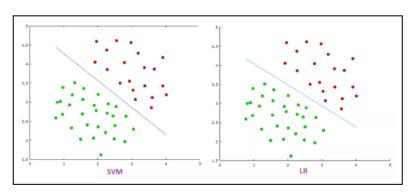
### **SVM Discussion Topics**

- 1. Maximum Margin Classifiers
- 2. SVM Training Methodology
- 3. Overlapping Distributions
- 4. SVM relation to Logistic Regression
- 5. Dealing with Multiple Classes
- 6. SVM and Computational Learning Theory
- 7. Relevance Vector Machines

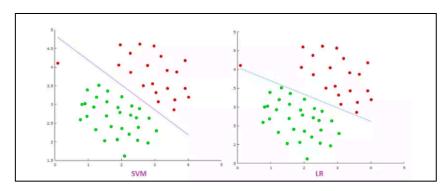
# SVM vs Logistic Regression

- What are similarities and differences between SVM and logistic regression
  - SVM finds widest separating margin
  - Logistic Regression optimizes log likelihood, with probabilities modeled by the sigmoid

Similar results



Sensitivity with an outlier



## SVM extends Logistic Regression

- By using kernel trick: transforming datasets into rich features space
- Complex problems still dealt with in the same "linear" fashion in the lifted hyper space

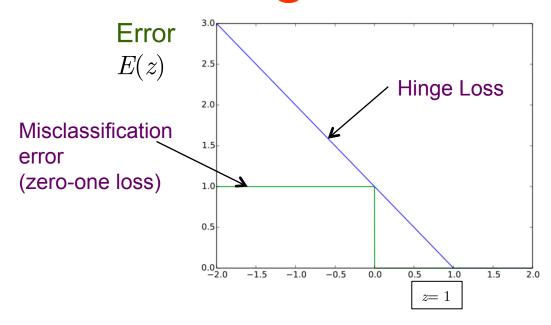
#### **Error Function of SVM**

- A hyperplane is defined by  $y=w^{T}x+b=0$ 
  - For training set:  $\{x_1,...,x_N\}$ ,  $\{t_1,...,t_N\}$ ,  $t_n \in \{-1,1\}$ , n=1,...,N
  - SVM chooses  $\boldsymbol{w}$  such that  $y_n t_n = [\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_n + b] \bullet t_n \geq 1$
- This defines the hinge error
  - For intended output  $t=\pm 1$  and classifier score y, the hinge error of the prediction is

$$E(y) = \max(0, 1-yt)$$

- When t and y have the same sign (ie, y predicts the right class) and  $|y| \ge 1$ , then E(y) = 0
  - But when of opposite sign, E(y) increases linearly with y (and similarly if |y| < 1, even if it has the same sign (correct prediction, but not by enough margin).

## Hinge Error of SVM



Hinge error  $E(z) = \max(0, 1-z)$ 

When t and y have the same sign (meaning y predicts the right class) and  $|yt| \ge 1$ , the hinge error E(yt) = 0

But when they have opposite sign, E(yt) increases linearly with y (one-sided error), and similarly if |yt| < 1, even if it has the same sign (correct prediction, but not by enough margin).

Hinge error penalizes predictions yt < 1, corresponding to the notion of a margin in a SVM

#### Loss Function of SVM

The objective function of SVM can be written as

$$\left| \sum_{n=1}^{N} E_{SV} \left( \boldsymbol{y}_{n} t_{n} \right) + \lambda \left| \left| \boldsymbol{w} \right| \right|^{2} \right|$$

• Where  $E_{
m SV}$  is the hinge error function defined by

$$\left|E_{_{SV}}(y_{_{n}}t_{_{n}})=\left[1-y_{_{n}}t_{_{n}}\right]_{\scriptscriptstyle{+}}\right|$$

– and [.] denotes the positive part

## Loss function of Logistic Regression

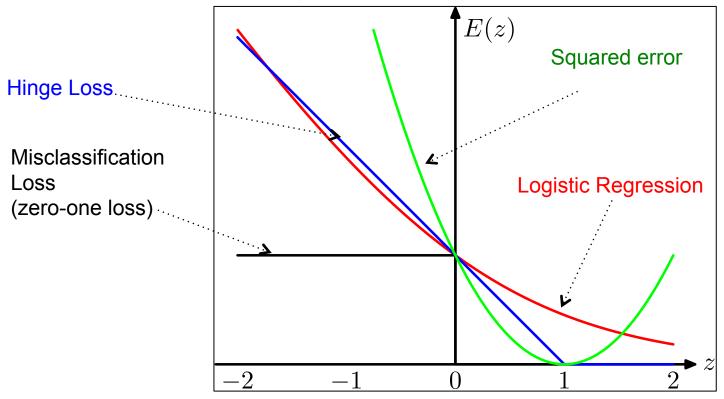
- LR typically uses target variable  $t \in \{0,1\}$ 
  - For comparison with SVM reformulate max likelihood logistic regression using  $t \in \{-1,1\}$ 
    - To do this we note that  $p(t=1|y)=\sigma(y)$ 
      - It follows that  $p(t=-1|y)=1-\sigma(y)$
      - So we can write  $p(t|y) = \sigma(yt)$
  - From negative logarithm of likelihood

$$\boxed{\sum_{n=1}^{N} E_{LR} \Big( \boldsymbol{y}_{n} \boldsymbol{t}_{n} \Big) + \lambda \Big| \Big| \boldsymbol{w} \Big|^{2}}$$

where

$$E_{LR}(yt) = \ln(1 + \exp(-yt))$$

# Summary of Loss Functions



Both logistic error and the hinge loss can be viewed as continuous approximations to the misclassification error

The squared error places increasing emphasis on data points that are correctly classified But that are a long way from the decision boundary on the correct side

# Outlier sensitivity

- Cost function of LR diverges faster than that of SVM
- So LR is more sensitive to outliers than SVM

#### Which classifier to use?

- Try logistic regression first and see how you do with that simpler model
- If logistic regression fails and you have reason to believe your data won't be linearly separable, try an SVM with a nonlinear kernel like a Radial Basis Function (RBF)