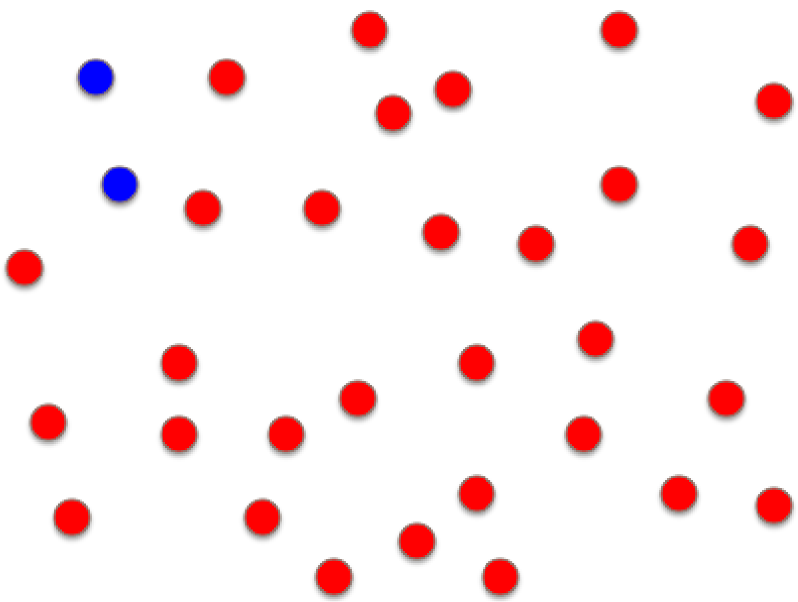
On the Statistical Consistency of Algorithms for Binary Classification under Class Imbalance

Aditya K. Menon[1], Harikrishna Narasimhan[2], Shivani Agarwal[2] and Sanjay Chawla[3]

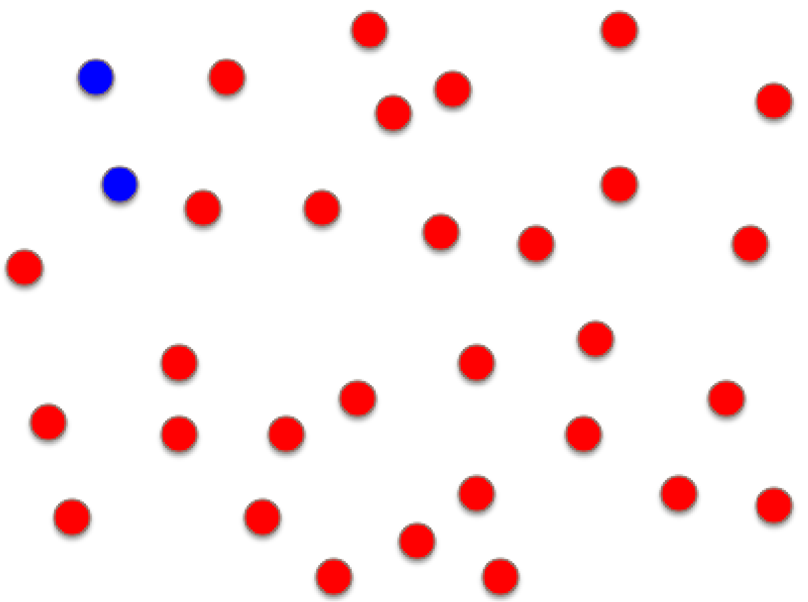
[1]University of California, San Diego
[2]Indian Institute of Science, Bangalore
[3]University of Sydney and NICTA, Sydney

Class Imbalance



- Medical Diagnosis
- Text Retrieval
- Credit Risk Minimization
- Fraud Detection
- ....

Class Imbalance



- Medical Diagnosis
- Text Retrieval
- Credit Risk Minimization
- Fraud Detection
- ....

Standard misclassification error ill-suited!

Class Imbalance

Measure	Definition	References
A-Mean (AM)	$(\text{TPR} + \text{TNR})/2$	Chan & Stolfo (1998); Powers et al. (2005); Gu et al. (2009); KDD Cup 2001 challenge
G-Mean (GM)	$\sqrt{\text{TPR} \cdot \text{TNR}}$	Kubat & Matwin (1997); Daskalaki et al. (2006)
H-Mean (HM)	$2/(\frac{1}{\text{TPR}} + \frac{1}{\text{TNR}})$	Kennedy et al. (2009)
Q-Mean (QM)	$1 - ((\text{FPR})^2 + (\text{FNR})^2)/2$	Lawrence et al. (1998)
F_1	$2/(\frac{1}{\text{Prec}} + \frac{1}{\text{TPR}})$	Lewis & Gale (1994) Gu et al. (2009)
G-TP/PR	$\sqrt{\text{TPR} \cdot \text{Prec}}$	Daskalaki et al. (2006)
AUC-ROC	Area under ROC curve	Ling et al. (1998)
AUC-PR	Area under precision-recall curve	Davis & Goadrich (2006) Liu & Chawla (2011)

Class Imbalance

Measure	Definition	References
A-Mean (AM)	$(\mathrm{TPR} + \mathrm{TNR})/2$	Chan & Stolfo (1998); Powers et al. (2005); Gu et al. (2009); KDD Cup 2001 challenge
G-Mean (GM)	$\sqrt{\mathrm{TPR} \cdot \mathrm{TNR}}$	Kubat & Matwin (1997); Daskalaki et al. (2006)
H-Mean (HM)	$2/(\frac{1}{\mathrm{TPR}} + \frac{1}{\mathrm{TNR}})$	Kennedy et al. (2009)
Q-Mean (QM)	$1 - ((\mathrm{FPR})^2 + (\mathrm{FNR})^2)/2$	Lawrence et al. (1998)
F_1	$2/(\frac{1}{\mathrm{Prec}} + \frac{1}{\mathrm{TPR}})$	Lewis & Gale (1994) Gu et al. (2009)
G-TP/PR	$\sqrt{\mathrm{TPR} \cdot \mathrm{Prec}}$	Daskalaki et al. (2006)
AUC-ROC	Area under ROC curve	Ling et al. (1998)
AUC-PR	Area under precision-recall curve	Davis & Goadrich (2006) Liu & Chawla (2011)

Algorithmic Approaches

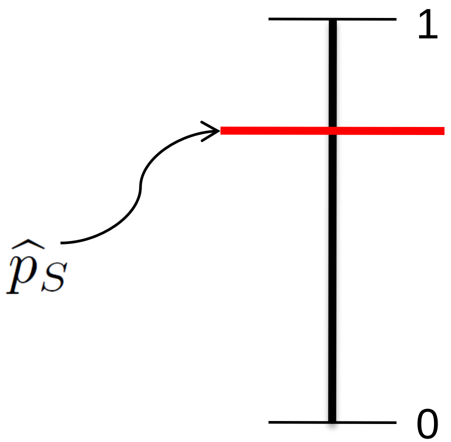
- **Sampling:** (Japkowicz & Stephen, 2002; Chawla et al., 2002, 2003; Van Hulse et al., 2007; He & Garcia, 2009)
  - Over-sample the minority class
  - Under-sample the majority class
  - SMOTE
  - …
- **Plug-in classifier** (Elkan, 2001)
- **Balanced ERM** (Liu & Chawla, 2011; Wallace et al., 2011)

Two Families of Algorithms

Algorithm 1

Plug-in with Empirical Threshold

- Learn a class probability estimator from training data *S*.
- Apply a suitable empirical threshold on the class probability estimate.

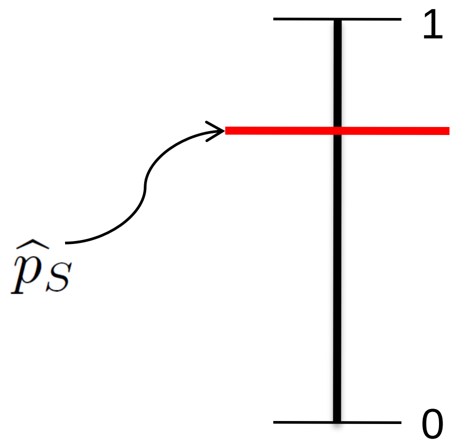


Two Families of Algorithms

Algorithm 1

Plug-in with Empirical Threshold

- Learn a class probability estimator from training data *S*.
- Apply a suitable empirical threshold on the class probability estimate:



Algorithm 2

Empirically Balanced ERM

- Learn a binary classifier by minimizing a balanced surrogate loss.
- Balancing terms estimated from training data.

Main Consistency Results

AM-regret

$$\mathrm{regret}_D^{\mathrm{AM}}[h] = \sup_{h:\mathcal{X}\to\{\pm 1\}} \mathrm{AM}_D[h] \; - \; \mathrm{AM}_D[h]$$

Main Consistency Results

AM-regret

$$\mathrm{regret}_D^{\mathrm{AM}}[h] = \sup_{h:\mathcal{X}\to\{\pm 1\}} \mathrm{AM}_D[h] \; - \; \mathrm{AM}_D[h]$$

AM-consistency

$$\boxed{\mathrm{regret}_D^{\mathrm{AM}}[h_S] \xrightarrow{P} 0}$$

Main Consistency Results

AM-regret

$$\mathrm{regret}_D^{\mathrm{AM}}[h] = \sup_{h:\mathcal{X}\to\{\pm 1\}} \mathrm{AM}_D[h] \;-\; \mathrm{AM}_D[h]$$

AM-consistency

$$\boxed{\mathrm{regret}_D^{\mathrm{AM}}[h_S] \xrightarrow{P} 0}$$

Main Results: Under mild conditions on the underlying distribution and under certain assumptions on the surrogate loss function minimized, Algorithms 1 and 2 are AM-consistent.

Key Ingredients in Proofs

- Balanced losses (Kotlowski et al, 2011)

$$\mathrm{AM}_D[h] = 1 - \mathrm{er}_D^{0\text{-}1,\mathrm{bal}}[h]$$

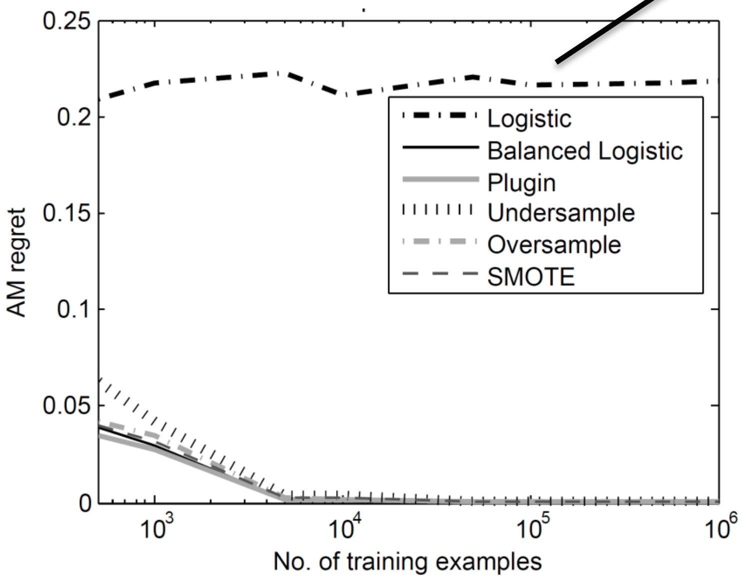
- Decomposition lemma:

$$\mathrm{regret}_D^{0\text{-}1,(\widehat{p}_S)}[h_S] \xrightarrow{\mathrm{P}} 0 \quad \Longrightarrow \quad \mathrm{regret}_D^{\mathrm{AM}}[h_S] \xrightarrow{\mathrm{P}} 0$$

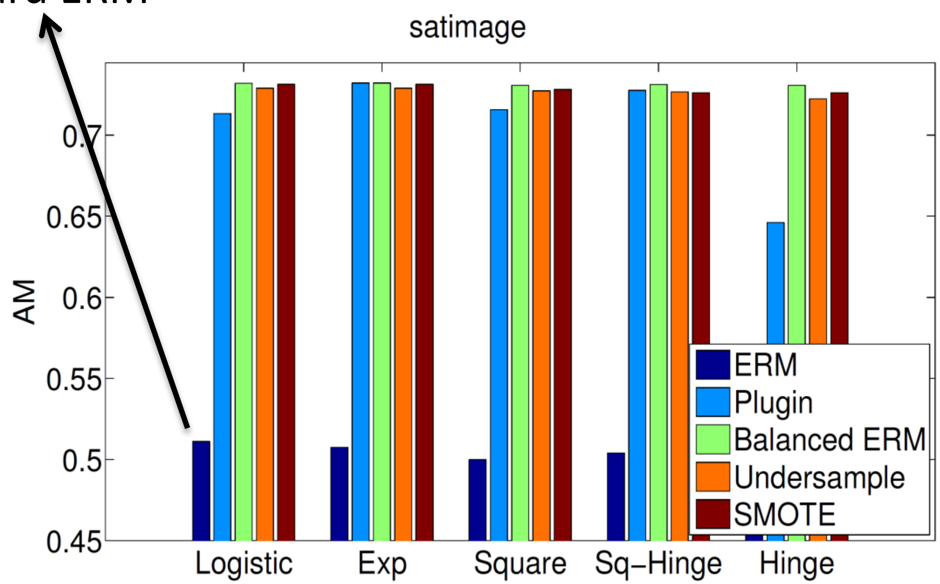
- Surrogate regret bounds for cost-sensitive classification (Scott, 2012)
- Proper and strongly proper losses (Reid and Williamson, 2009, 2010; Agarwal, 2013)
- Surrogate regret bounds for standard binary classification (Zhang, 2004; Bartlett et al, 2006)

Experiments

Standard ERM



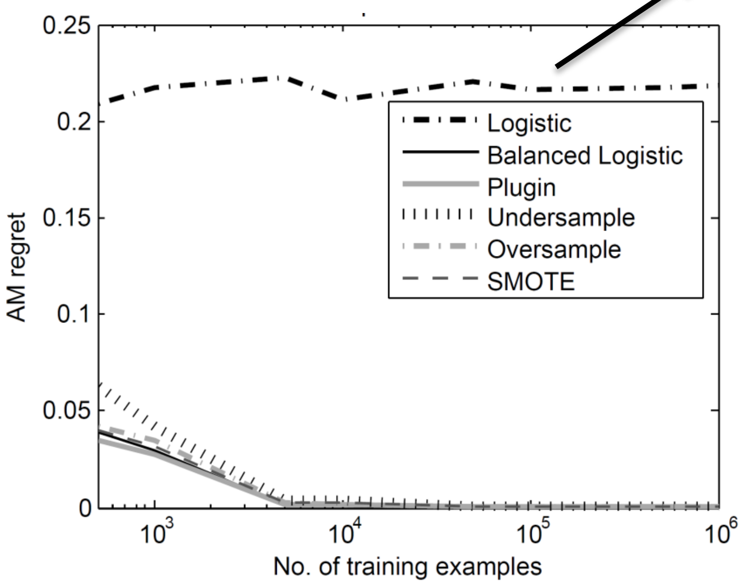
Synthetic data
p = 0.05



Real data
p = 0.097

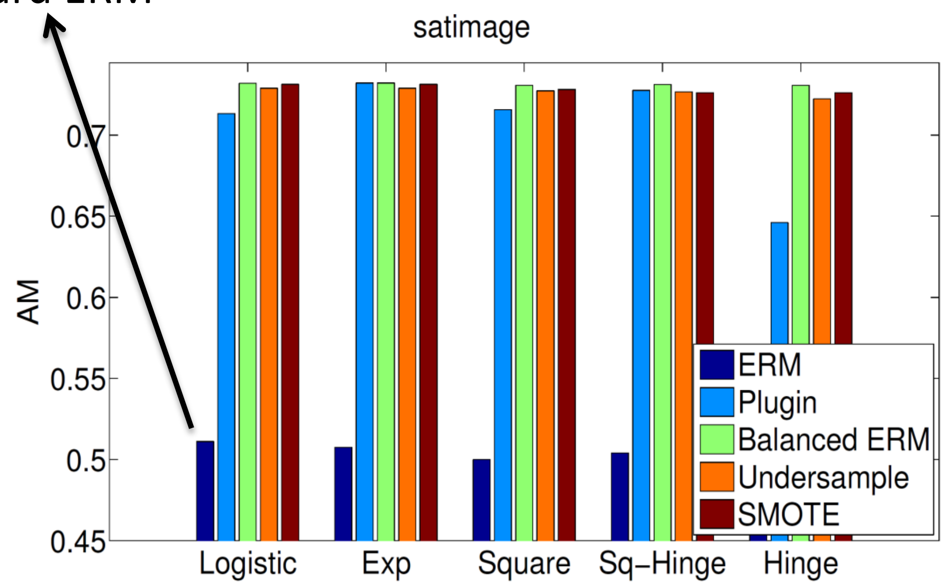
Experiments

Standard ERM



Synthetic data

p = 0.05



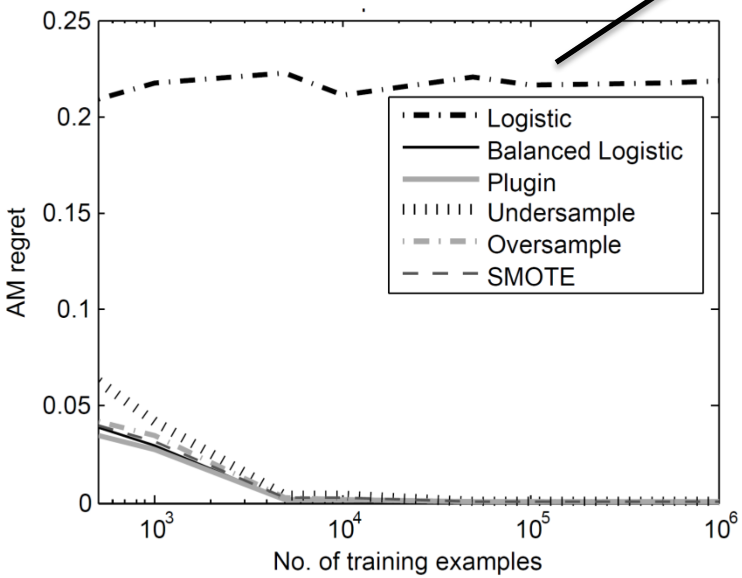
Real data

p = 0.097

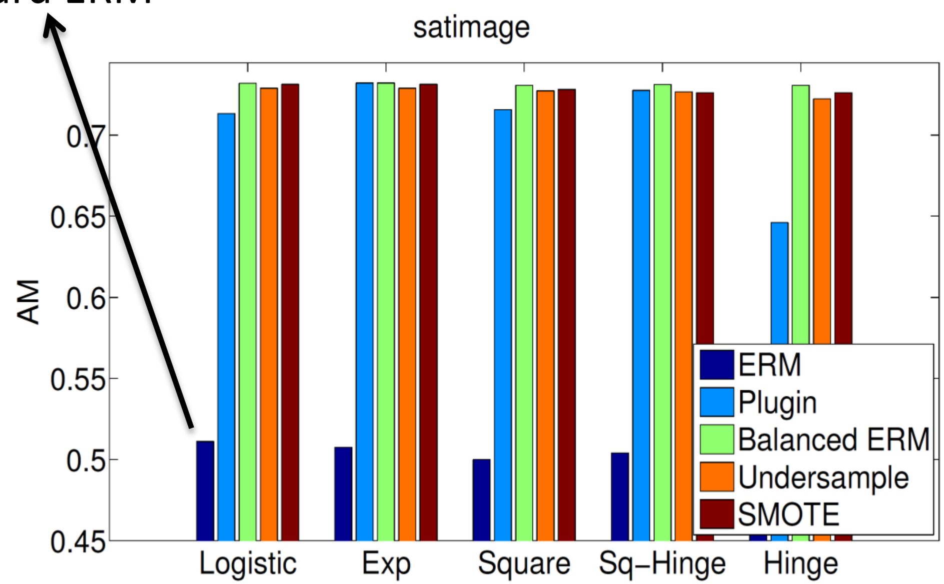
AM performance of Plug-in and Balanced ERM comparable to that of the sampling techniques

Experiments

Standard ERM



Synthetic data
p = 0.05



Real data
p = 0.097

AM performance of Plug-in and Balanced ERM comparable to that of the sampling techniques

Poster 794
Today