Dealing with Imbalanced Data,

and,

Interpretability via Adversarial Regularization

- Rémi Emonet
- Université Jean-Monnet,
  Laboratoire Hubert Curien,
  Saint-Étienne

- Talk at NAVER LABS (Grenoble), 2019-06-13

$whoami

Overview

- Introduction
- Imbalanced classification problems
  - The Problem (and performance measures)
  - Reweight, resampling, etc
  - Correcting k-NN (γ-***NN***)
  - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
  - Pleading for interpretability
  - Adversarial Input↔Parameter Regularization (***AI↔PR***)
  - Previous, current and future approaches
- Discussion

Overview

- Introduction
- Imbalanced classification problems
  - The Problem (and performance measures)
  - Reweight, resampling, etc
  - Correcting k-NN (γ-***NN***)
  - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
  - Pleading for interpretability
  - Adversarial Input↔Parameter Regularization (***AI↔PR***)
  - Previous, current and future approaches
- Discussion

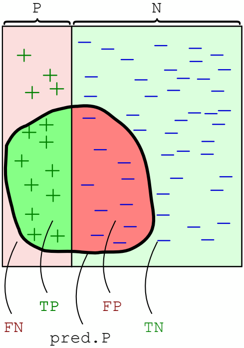
Imbalanced Problems: Examples

- Anomaly detection
  - unsafe situations in videos
  - defect detection in images
  - anormal heart beat in ECG
- Fraud detection
  - fraudulent checks
  - credit card fraud (physical, online)
  - financial fraud (French DGFIP)

Imbalanced Classification Problems

- Binary classification
  - $+$ positive class: minority class, anomaly, rare event, …
  - $-$ negative class: majority class, normality, typical event, …
- Confusion matrix (of a model vs a ground truth)
  - TP: true positive
  - FP: false positive
  - TN: true negative
  - FN: false negative
- Some measures
  - Precision: $prec = \dfrac{TP}{TP + FP}$
  - Recall: $rec = \dfrac{TP}{P} = \dfrac{TP}{TP + FN}$
  - F_β-measure: $F_\beta = (1 + \beta^2)\dfrac{prec \cdot rec}{\beta^2 \cdot prec + rec}$

    *(higher is better)



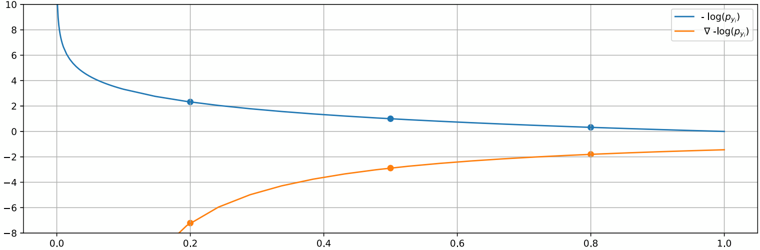
F-measure vs Accuracy ?

$$F_\beta = (1+\beta^2)\frac{prec \cdot rec}{\beta^2 \cdot prec + rec} = \frac{(1+\beta^2)\cdot(P - FN)}{1+\beta^2 P - FN + FP}$$

$$accuracy = \frac{TP+TN}{P+N} = 1 - \frac{FN+FP}{P+N}$$

- Accuracy inadequacy (e.g. $N = 10000, P = 10$)
  - lazy "all$-$" classifier $(TP = 0, TN = N, FP = 0, FN = P)$
  - $accuracy = \frac{0+N}{P+N} = \frac{10000}{10100} = 99\%$
  - $F_\beta = \frac{(1+\beta^2)(P-P)}{1+\beta^2 P - P + 0} = 0$
- F_β-measure challenges
  - discrete (like the accuracy)
  - non-convex (even with continuous surrogates)
  - **non-separable**, i.e. $F_\beta \neq \sum_{(x_i,y_i)\in S} \ldots$

Ok, but I'm doing gradient descent, so …



Disclaimer: multiclass F_β-measures

Overview

- Introduction
- Imbalanced classification problems
  - The Problem (and performance measures)
  - Reweight, resampling, etc
  - Correcting k-NN (γ-***NN***)
  - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
  - Pleading for interpretability
  - Adversarial Input↔Parameter Regularization (***AI↔PR***)
  - Previous, current and future approaches
- Discussion

Counteracting Imbalance

- Undersampling the majority class $-$
- Oversampling class $+$
- Generating fake $+$
- Using a weighted-classifiers learner

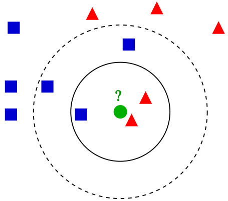
Overview

- Introduction
- Imbalanced classification problems
  - The Problem (and performance measures)
  - Reweight, resampling, etc
  - Correcting k-NN (γ-***NN***)
  - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
  - Pleading for interpretability
  - Adversarial Input↔Parameter Regularization (***AI↔PR***)
  - Previous, current and future approaches
- Discussion

A Corrected Nearest Neighbor Algorithm Maximizing the F-Measure from Imbalanced Data

- **Rémi Viola**, Rémi Emonet , Amaury Habrard, **Guillaume Metzler**, Sébastien Riou, Marc Sebban
- ???

k-NN: k Nearest Neighbor Classification

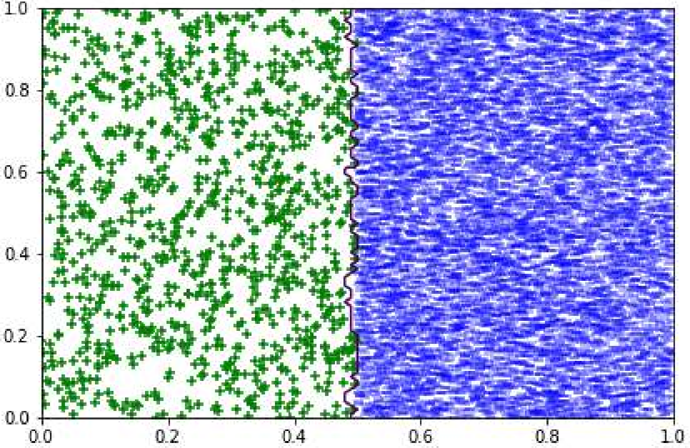


- k-NN
  - to classify a new point
  - find the closest k points (in the training section)
  - use a voting scheme to affect a class
  - efficient algorithms (K-D Tree, Ball Tree)
- Does k-NN still matter?
  - non-linear by design (with similarity to RBF-kernel SVM)
  - no learning, easy to patch a model (add/remove points)
  - Limits of k-NN for imbalanced data?

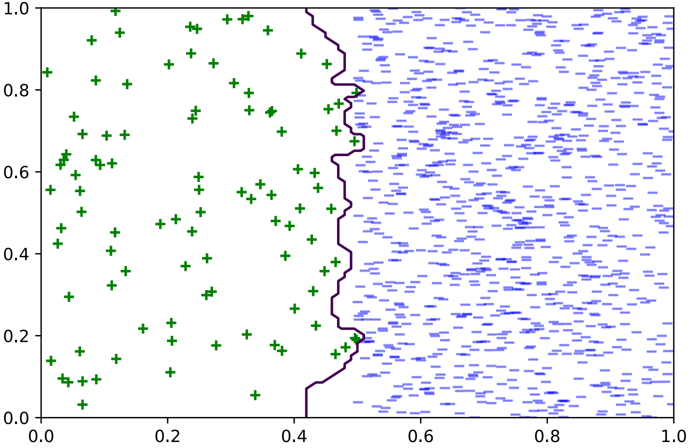
Limits of k-NN for imbalanced data?

1. k-NN behavior in uncertain areas
    - i.e., for some feature vector, the class can be $+$ or $-$
    - i.e., the Bayes Risk is non zero
    - ✔ not so bad (respects imbalance)
2. k-NN behavior around boundaries
    - i.e., what happens if classes are separate but imbalanced
    - ✖ sampling effects cause problems

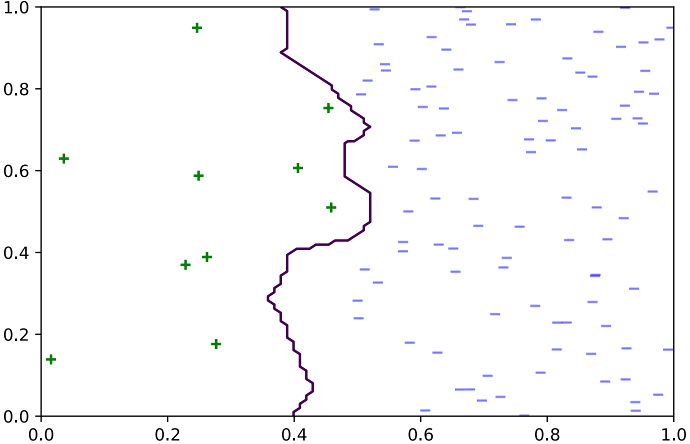
k-NN at a boundary (1000 +)



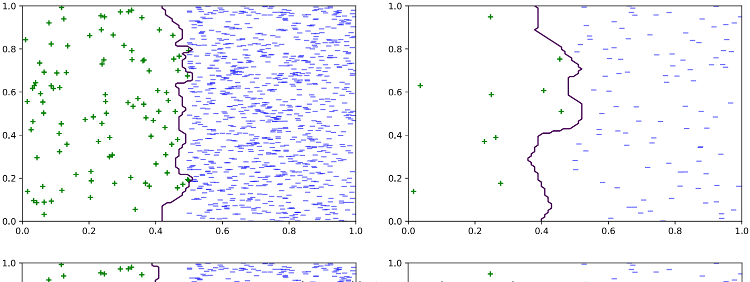
k-NN at a boundary (100 $+$)



k-NN at a boundary (10 $+$)



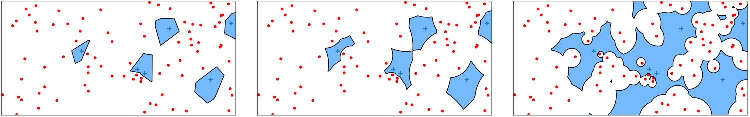
k-NN: increasing k?



A Corrected Nearest Neighbor Algorithm Maximizing the F-Measure from Imbalanced Data

- **Rémi Viola**, Rémi Emonet , Amaury Habrard, **Guillaume Metzler**, Sébastien Riou, Marc Sebban
- ???

γ-NN Idea: push the decision boundary



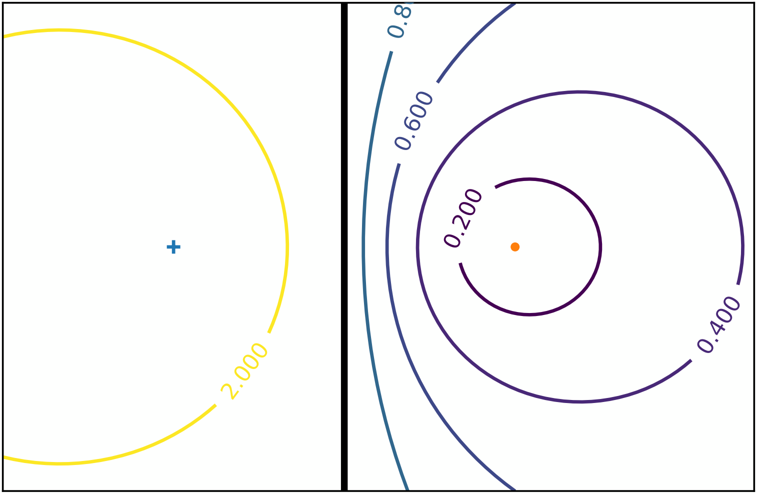
- Goal: correct for problems due to sampling with imbalance
- Genesis: GAN to generate "$+$" around existing ones

  $\Rightarrow$ unstable, failing, complex
- Approach
  - artificially make $+$ closer to new points
  - how? by using a different distance for $+$ and $-$
  - the base distance to $+$ gets multiplied by a parameter γ

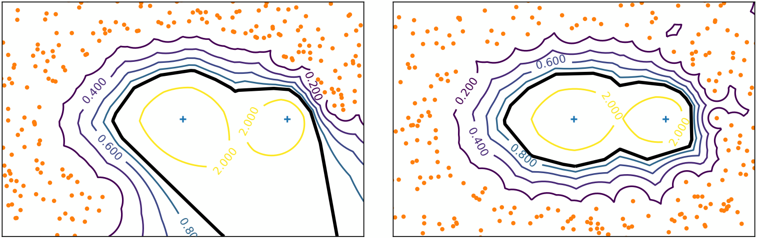
    (intuitively $\gamma \leq 1$ if $+$ is rare)

$$d_\gamma(x, x_i) = \begin{cases} d(x, x_i) & \text{if } x_i \in S_-, \\ \gamma \cdot d(x, x_i) & \text{if } x_i \in S_+. \end{cases}$$

γ-NN: varying γ with two points



γ-NN: varying γ with a few $+$



γ-NN: Algorithm

Algorithm 1: Classification of a new example with γk−NN

Input : a query $\mathbf{x}$ to be classified, a set of labeled samples $S = S_+ \cup S_-$, a number of neighbors k, a positive real value γ, a distance function d

Output: the predicted label of $\mathbf{x}$

$\mathcal{NN}^-, \mathcal{D}^- \leftarrow nn(k, \mathbf{x}, S_-)$ // nearest negative neighbors with their distances

$\mathcal{NN}^+, \mathcal{D}^+ \leftarrow nn(k, \mathbf{x}, S_+)$ // nearest positive neighbors with their distances

$\mathcal{D}^+ \leftarrow \gamma \cdot \mathcal{D}^+$

$\mathcal{NN}_\gamma \leftarrow firstK\left(k, sortedMerge((\mathcal{NN}^-, \mathcal{D}^-), (\mathcal{NN}^+, \mathcal{D}^+))\right)$

$y \leftarrow +$ if $|\mathcal{NN}_\gamma \cap \mathcal{NN}^+| \geq \frac{k}{2}$ else $-$ // majority vote based on $\mathcal{NN}_\gamma$

return y

γ-NN: a way to reweight distributions

- In uncertain regions
- At the boundaries

Results on public datasets (F-measure)

DATASETS	3−NN	DUPk−NN	Wk−NN	CWk−NN	LMNN	γk−NN
BALANCE	0.954 (0.017)	0.954 (0.017)	0.957 (0.017)	0.961 (0.010)	**0.963 (0.012)**	0.954 (0.029)
AUTOMPG	0.808 (0.077)	0.826 (0.033)	0.810 (0.076)	0.815 (0.053)	0.827 (0.054)	**0.831 (0.025)**
IONO	0.752 (0.053)	0.859 (0.021)	0.756 (0.060)	0.799 (0.036)	0.890 (0.039)	**0.925 (0.017)**
PIMA	0.500 (0.056)	0.539 (0.033)	0.479 (0.044)	0.515 (0.037)	0.499 (0.070)	**0.560 (0.024)**
WINE	0.881 (0.072)	0.852 (0.057)	0.881 (0.072)	0.876 (0.080)	**0.950 (0.036)**	0.856 (0.086)
GLASS	0.727 (0.049)	0.733 (0.061)	0.736 (0.052)	0.717 (0.055)	0.725 (0.048)	**0.746 (0.046)**
GERMAN	0.330 (0.030)	0.449 (0.037)	0.326 (0.030)	0.344 (0.029)	0.323 (0.054)	**0.464 (0.029)**
VEHICLE	0.891 (0.044)	0.867 (0.027)	0.891 (0.044)	0.881 (0.021)	**0.958 (0.020)**	0.880 (0.049)
HAYES	0.036 (0.081)	0.183 (0.130)	0.050 (0.112)	0.221 (0.133)	0.036 (0.081)	**0.593 (0.072)**
SEGMENTATION	0.859 (0.028)	0.862 (0.018)	0.877 (0.028)	0.851 (0.022)	**0.885 (0.034)**	0.848 (0.025)
ABALONE8	0.243 (0.037)	0.318 (0.013)	0.241 (0.034)	0.330 (0.015)	0.246 (0.065)	**0.349 (0.018)**
YEAST3	0.634 (0.066)	0.670 (0.034)	0.634 (0.066)	**0.699 (0.015)**	0.667 (0.055)	0.687 (0.033)
PAGEBLOCKS	0.842 (0.020)	0.850 (0.024)	0.849 (0.019)	0.847 (0.029)	**0.856 (0.032)**	0.844 (0.023)
SATIMAGE	0.454 (0.039)	0.457 (0.027)	0.454 (0.039)	0.457 (0.023)	**0.487 (0.026)**	0.430 (0.008)
LIBRAS	**0.806 (0.076)**	0.788 (0.187)	**0.806 (0.076)**	0.789 (0.097)	0.770 (0.027)	0.768 (0.106)
WINE4	0.031 (0.069)	**0.090 (0.086)**	0.031 (0.069)	0.019 (0.042)	0.000 (0.000)	**0.090 (0.036)**
YEAST6	0.503 (0.302)	0.449 (0.112)	0.502 (0.297)	0.338 (0.071)	0.505 (0.231)	**0.553 (0.215)**
ABALONE17	0.057 (0.078)	**0.172 (0.086)**	0.057 (0.078)	0.096 (0.059)	0.000 (0.000)	0.100 (0.038)
ABALONE20	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	**0.067 (0.038)**	0.057 (0.128)	0.052 (0.047)
MEAN	0.543 (0.063)	0.575 (0.053)	0.544 (0.064)	0.559 (0.046)	0.560 (0.053)	**0.607 (0.049)**

Results on DGFiP datasets (F-measure)

DATASETS	3−NN	γk−NN	SMOTE	SMOTE+γk−NN
DGFIP19 2	0,454 (0,007)	0,528 (0,005)	0,505 (0,010)	**0,529** (0,003)
DGFIP9 2	0,173 (0,074)	0,396 (0,018)	0,340 (0,033)	**0,419** (0,029)
DGFIP4 2	0,164 (0,155)	0,373 (0,018)	0,368 (0,057)	**0,377** (0,018)
DGFIP8 1	0,100 (0,045)	**0,299** (0,010)	0,278 (0,043)	**0,299** (0,011)
DGFIP8 2	0,140 (0,078)	0,292 (0,028)	**0,313** (0,048)	0,312 (0,021)
DGFIP9 1	0,088 (0,090)	0,258 (0,036)	0,270 (0,079)	**0,288** (0,026)
DGFIP4 1	0,073 (0,101)	0,231 (0,139)	0,199 (0,129)	**0,278** (0,067)
DGFIP16 1	0,049 (0,074)	0,166 (0,065)	0,180 (0,061)	**0,191** (0,081)
DGFIP16 2	0,210 (0,102)	0,202 (0,056)	0,220 (0,043)	**0,229** (0,026)
DGFIP20 3	0,142 (0,015)	0,210 (0,019)	0,199 (0,015)	**0,212** (0,019)
DGFIP5 3	0,030 (0,012)	0,105 (0,008)	**0,110** (0,109)	0,107 (0,010)
MEAN	0,148 (0,068)	0,278 (0,037)	0,271 (0,057)	**0,295** (0,028)

Overview

- Introduction
- Imbalanced classification problems
  - The Problem (and performance measures)
  - Reweight, resampling, etc
  - Correcting k-NN (γ-***NN***)
  - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
  - Pleading for interpretability
  - Adversarial Input↔Parameter Regularization (***AI↔PR***)
  - Previous, current and future approaches
- Discussion

From Cost-Sensitive Classification to Tight F-measure Bounds

- **Kevin Bascol**, Rémi Emonet, Elisa Fromont, Amaury Habrard, **Guillaume Metzler**, Marc Sebban
- AISTATS2019

Optimizing the F_β-measure?

- Reminder
  - Precision: $prec = \frac{TP}{TP + FP}$
  - Recall: $rec = \frac{TP}{P} = \frac{TP}{TP + FN}$
  - F_β-measure: $F_\beta = (1+\beta^2)\frac{prec \cdot rec}{\beta^2 \cdot prec + rec}$
- **Non-separability**, i.e. $F_\beta \neq \sum_{(x_i, y_i) \in S} \dots$

NB: accuracy is separable, $acc = \sum_{(x_i, y_i) \in S} \frac{1}{m} \delta(y_i - \hat{y}_i)$

⇒ The loss for one point depends on the others
⇒ Impossible to optimize directly
⇒ Impossible to optimize on a subset (minibatch)

Weighted classification for F_β

$$F_\beta = \frac{(1+\beta^2)\cdot(P-FN)}{1+\beta^2 P - FN + FP} = \frac{(1+\beta^2)\cdot(P-e_1)}{1+\beta^2 P - e_1 + e_2}$$

- The F_β-measure is linear fractional *(in $e=(e_1,e_2)=(FN,FP)$)*
  i.e. $F_\beta = \frac{\langle a', e\rangle + b}{\langle c, e\rangle + d} = \frac{A}{B}$
- Relation to weighted classification

$F_\beta \geq t$ (we achieve a good, above t, F_β value)

$$\begin{aligned}
&\Leftrightarrow A \geq t\cdot B \\
&\Leftrightarrow A - t\cdot B \geq 0 \\
&\Leftrightarrow (1+\beta^2)\cdot(P-e_1) - t(1+\beta^2 P - e_1 + e_2) \geq 0 \\
&\Leftrightarrow (-1-\beta^2+t)e_1 - te_2 \geq -P(1+\beta^2) + t(1+\beta^2 P) \\
&\Leftrightarrow (1+\beta^2-t)e_1 + te_2 \leq -P(1+\beta^2) + t(1+\beta^2 P)
\end{aligned}$$

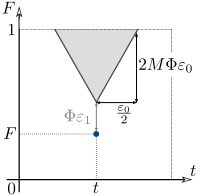
⇒ so, we can minimize the weighted problem
with class weights $a(t) = (1+\beta^2-t, t)$

Inspiration for our work

- **title:** *Optimizing F-measures by cost-sensitive classification*
- **by:** Parambath, S. P., Usunier, N., and Grandvalet, Y.
- **in:** NIPS 2014
- Main result (reformulation)
  - given a value of t that is at most at $\frac{\varepsilon_0}{2}$ from the optimal $t^\star$
  - given a weighted-classification learner
  - then the optimal $F^\star$ is not too far from the observed F
- They conclude that grid search on t is sound
  i.e., we'll get arbitrary close to the optimal F_β-measure
  (and they experiment with 19 t values)

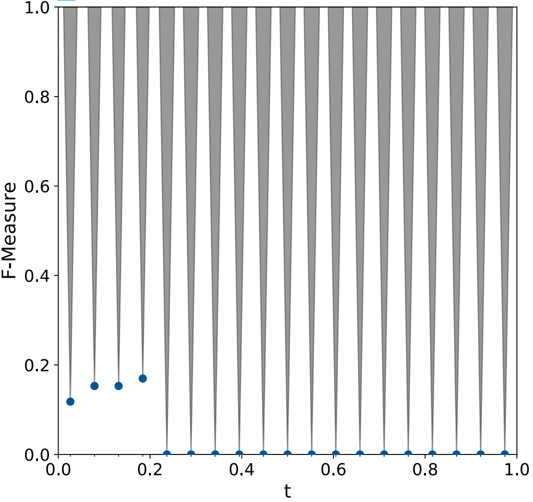
Geometric interpretation of bounds

- A reinterpretation of the bound // (and/or its derivations)
  - given
    - a value of t (and class weights $a(t)$), and
    - a weighted-classifier h (learned with weight $a(t)$)
    - … yielding an error profile e and a correponding $F_\beta(e)$
  - **then***, for any t' (that would yield a classifier e'),
    - *(if t' is close to t' then $F(e')$ is close to $F(e)$)*
    - $F(e')$ is upper bounded by an expression linear in $|t - t'|$



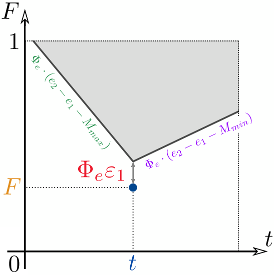
* All bounds are up to ε_1, the sub-optimality of the learned weighted classifier

Visualizing the bound of Parambath et al.



Deriving a tighter bound

- Same objective: drawing cones
- New, tighter, asymmetric bounds
- Given
    - all definitions
    - NB: ε_1 the sub-optimality



Then for all $t' < t$:

$$F(\mathbf{e}') \leq F(\mathbf{e}) + \Phi_{\mathbf{e}}\varepsilon_1 + \Phi_{\mathbf{e}} \cdot (e_2 - e_1 - M_{max})(t' - t),$$

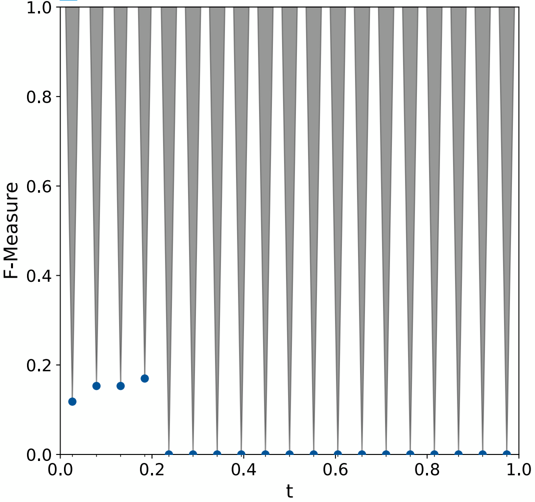
where $M_{max} = \max_{\substack{\mathbf{e}'' \in \mathcal{E}(\mathcal{H}) \\ s.t.\ F(\mathbf{e}'') > F(\mathbf{e})}} (e''_2 - e''_1)$

and, for all $t' > t$:

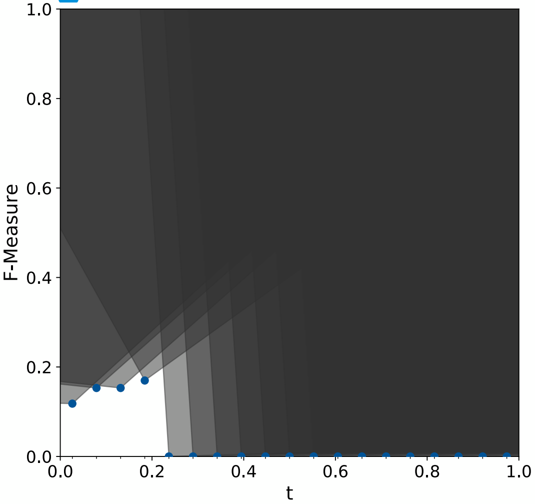
$$F(\mathbf{e}') \leq F(\mathbf{e}) + \Phi_{\mathbf{e}}\varepsilon_1 + \Phi_{\mathbf{e}} \cdot (e_2 - e_1 - M_{min})(t' - t),$$

where $M_{min} = \min_{\substack{\mathbf{e}'' \in \mathcal{E}(\mathcal{H}) \\ s.t.\ F(\mathbf{e}'') > F(\mathbf{e})}} (e''_2 - e''_1).$

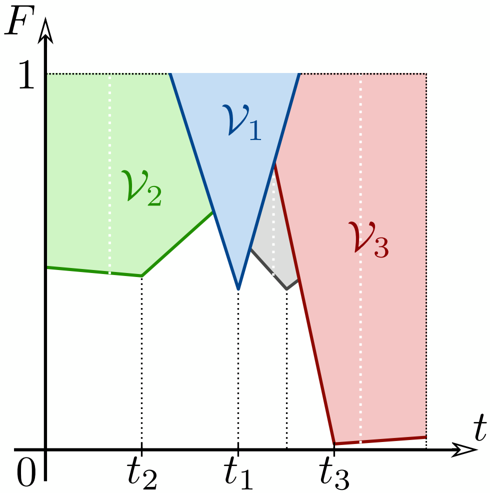
Visualizing the bound of Parambath et al.



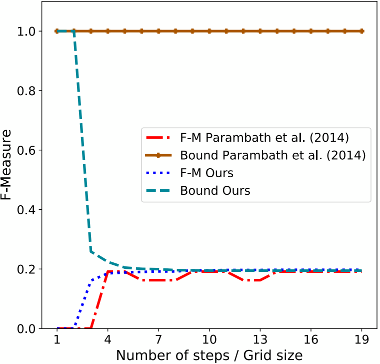
Visualizing our bound

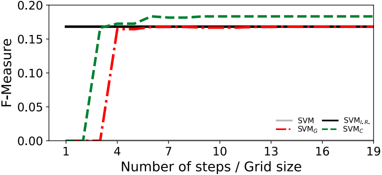


Using the bound to better explore t values



CONE: example results





Non-vacuous bounds, incremental algo., early increase.

CONE: numerical results

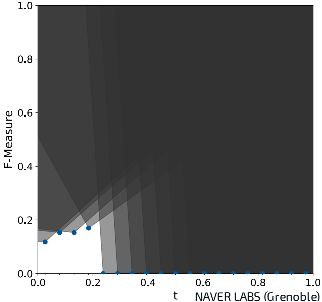
Dataset	SVM	$\text{SVM}_{I.R.}$	SVM_{G}	SVM_{C}	SVM_{C}^{T}	LR^{T}	$\text{LR}^{T}_{I.R.}$	LR^{T}_{G}	LR_{B}
Adult	62.5 (0.2)	64.9 (0.3)	66.4 (0.1)	66.5 (0.1)	66.4 (0.1)	66.5 (0.1)	66.5 (0.1)	66.5 (0.1)	**66.6** (0.1)
Abalone10	0.0 (0.0)	30.9 (1.2)	**32.4** (1.3)	32.2 (0.8)	31.8 (1.9)	30.8 (2.2)	30.7 (1.9)	30.7 (1.9)	31.6 (0.6)
Satimage	0.0 (0.0)	23.4 (4.3)	20.4 (5.3)	20.6 (5.6)	**30.9** (2.0)	21.2 (11.1)	28.6 (1.9)	28.6 (1.9)	21.4 (4.6)
IJCNN	44.5 (0.4)	53.3 (0.4)	61.6 (0.6)	61.6 (0.6)	**62.6** (0.4)	59.4 (0.5)	56.5 (0.3)	56.5 (0.3)	59.2 (0.3)
Abalone12	0.0 (0.0)	16.8 (2.7)	16.8 (4.2)	**18.3** (3.3)	16.3 (3.0)	15.5 (3.1)	17.0 (3.3)	17.0 (3.3)	17.7 (3.7)
Pageblocks	48.1 (5.8)	39.6 (4.7)	66.4 (3.2)	62.8 (3.9)	**67.6** (4.0)	59.2 (8.1)	55.9 (6.4)	55.9 (6.4)	55.7 (5.7)
Yeast	0.0 (0.0)	29.4 (2.9)	38.6 (7.1)	39.0 (7.5)	35.4 (15.6)	37.4 (10.1)	**39.9** (6.5)	27.6 (6.8)	27.6 (6.8)
Wine	0.0 (0.0)	15.6 (5.2)	20.0 (6.4)	22.7 (6.0)	19.3 (7.9)	21.5 (3.7)	**25.2** (4.5)	**25.2** (4.5)	18.3 (7.2)
Average	19.4 (0.8)	34.2 (2.8)	40.3 (3.5)	40.5 (3.5)	**41.3** (4.4)	38.9 (5.2)	40.0 (3.1)	38.5 (3.2)	37.3 (3.8)

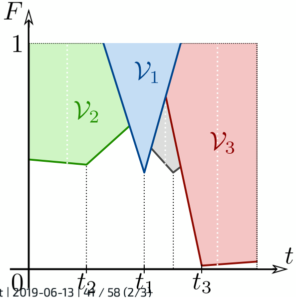
F-Measure for Logistic Regression (LR) [3] and SVM algorithms (averaged over 5 experiments).
"$I.R.$" I.R. based class costs; "G": Reproduction of [1] ; "C": CONE "T" thresholded predictions; "B": Bissection [2]

No huge gain, better F-measure in average.

CONE: Perpsectives

- We
  - provided a geometric interpretation as cones
  - derived new, tighter bounds
  - proposed a search algorithm using the bounds
- Work in progress
  - bounds in generalization
  - lower bounds





Overview

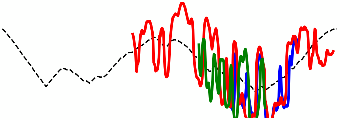
- Introduction
- Imbalanced classification problems
  - The Problem (and performance measures)
  - Reweight, resampling, etc
  - Correcting k-NN (γ-***NN***)
  - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
  - Pleading for interpretability
  - Adversarial Input↔Parameter Regularization (***AI↔PR***)
  - Previous, current and future approaches
- Discussion

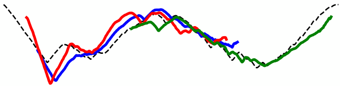
Interpretability

- Why, what

Example of interpretable patterns

- We overlay on the series the pattern that were most relevant to the decision taken by the classifier





- What side (left or right) would be more convincing?

Approaches towards interpretability

- Black box models
- Intepretability by design
  - constraints from the model structure

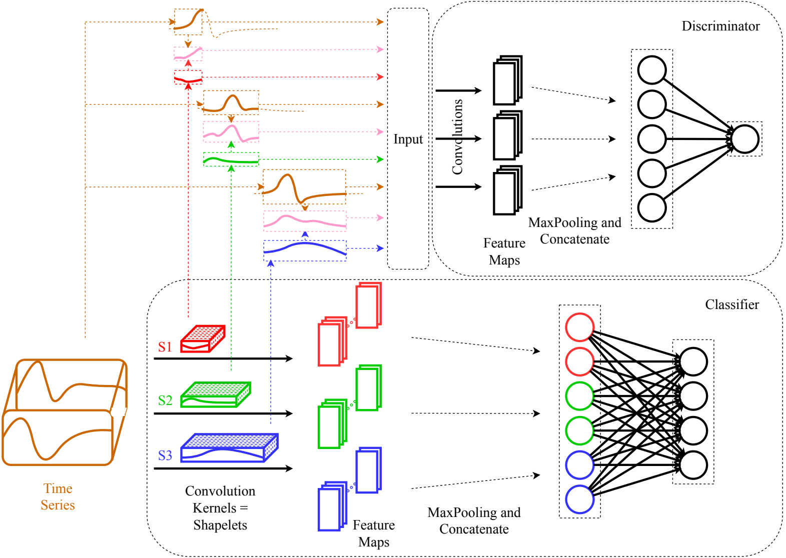
Overview

- Introduction
- Imbalanced classification problems
  - The Problem (and performance measures)
  - Reweight, resampling, etc
  - Correcting k-NN (γ-***NN***)
  - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
  - Pleading for interpretability
  - Adversarial Input↔Parameter Regularization (***AI↔PR***)
  - Previous, current and future approaches
- Discussion

Learning Interpretable Shapelets for Time Series Classification through Adversarial Regularization

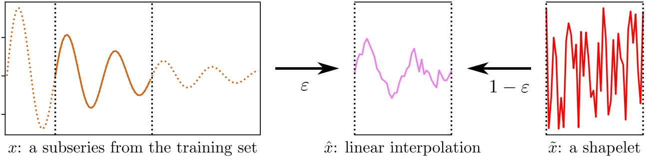
- **Yichang Wang**, Rémi Emonet, Elisa Fromont, Simon Malinowski, **Etienne Menager, Loic Mosser**, Romain Tavenard
- ??? + arxiv2019

Adversarial Input↔Parameter Regularization (*AI↔PR*)



Optimal Transport, Duality and Adversarial Learning

- $L_d(\theta_d) = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_S}\left[D(\tilde{x})\right] - \mathbb{E}_{x\sim\mathbb{P}_x}\left[D(x)\right] + \lambda\,\mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}\left[(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)^2\right]$
- Gradient penalty
- Interpolation to cover the space



3 Losses for 3 Phases

- Classifier training
  $L_C(\theta_c) = ...$
- Discriminator Training, using fixed shapelets
  $L_d(\theta_d) = \mathbb{E}_{\tilde{x}\sim\mathbb{P}_S}\left[D(\tilde{x})\right] - \mathbb{E}_{x\sim\mathbb{P}_x}\left[D(x)\right] + \lambda\,\mathbb{E}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}\left[(||\nabla_{\hat{x}}D(\hat{x})||_2 - 1)^2\right]$
- Shapelet regularization, using the fixed discriminator
  $L_r(\theta_s) = -\,\mathbb{E}_{\tilde{x}\sim\mathbb{P}_S}[D(\tilde{x})]$

Algorithm

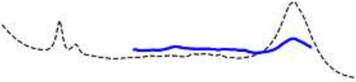
Algorithm 1: Learning Interpretable Shapelet

Require: number of shapelets n_S
Require: random initialization for the classifier/discriminator/shapelets $\theta_c, \theta_d, \theta_s \subset \theta_c$
Require: gradient penalty coefficient λ
Require: number of epochs n_{epochs}, mini-batch size m
Require: number of classifier/discriminator/regularization mini-batches per epoch n_c, n_d, n_r
Require: optimizer (Adam) hyperparameters α, β_1, β_2

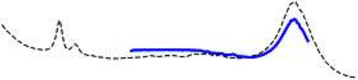
for $i = 1, \ldots, n_{epochs}$ **do**
    for $t = 1, \ldots, n_c$ **do**
        for $j = 1, \ldots, m$ **do**
            Sample a pair (Z_j, y_j) from the training set

Discriminator work across iterations

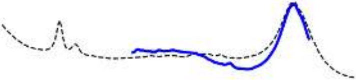
(b) **Shapelet at epoch** 20



(c) **Shapelet at epoch** 200

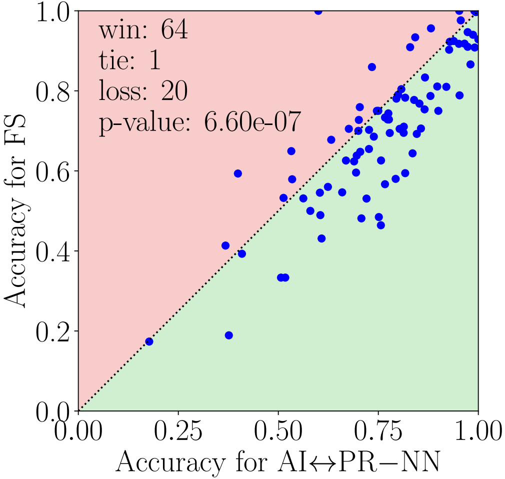


(e) **Shapelet at epoch** 800

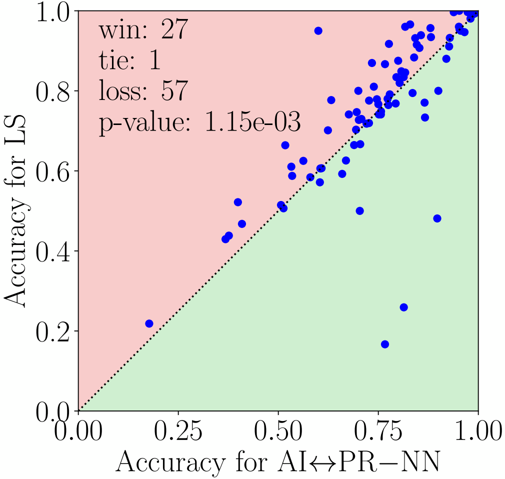


(f) **Shapelet at epoch** 8000

Results on 85 datasets (vs Fast Shapelets)



Results *(vs Learning Shapelets, non-interpretable)*



Overview

- Introduction
- Imbalanced classification problems
  - The Problem (and performance measures)
  - Reweight, resampling, etc
  - Correcting k-NN (γ-***NN***)
  - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
  - Pleading for interpretability
  - Adversarial Input↔Parameter Regularization (***AI↔PR***)
  - Previous, current and future approaches
- Discussion

More...

- Perspectives around ***AI***↔***PR***
  - multivariate
  - group-lasso used well
  - deeper regularization
- Other interpretable families of methods
  - probablisistic models
  - auto-encoders with regularizations and constraints

  - neural EM etc

Overview

- Introduction
- Imbalanced classification problems
    - The Problem (and performance measures)
    - Reweight, resampling, etc
    - Correcting k-NN (γ-***NN***)
    - Focusing on the F-Measure optimization (***CONE***)
- Interpretability
    - Pleading for interpretability
    - Adversarial Input↔Parameter Regularization (***AI↔PR***)
    - Previous, current and future approaches
- Discussion

Thank you! Questions?